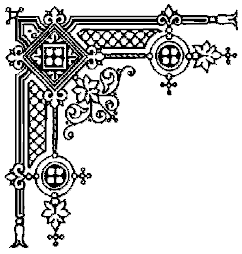
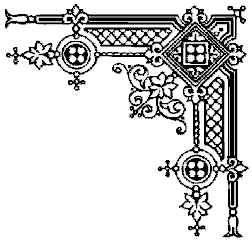



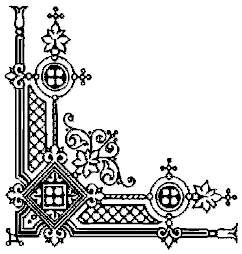
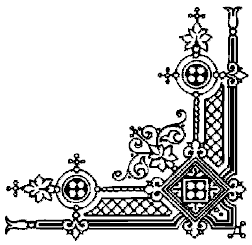
Visual Animal Biometrics

Automatic Detection and Individual Identification by Coat Pattern

Tilo Burghardt

A dissertation submitted to the University of Bristol
in accordance with the requirements of the degree of Doctor of Philosophy
in the Faculty of Engineering, Department of Computer Science.

2008

ABSTRACT

The field of computer vision has repeatedly been recognised as an intellectual frontier whose boundaries of applicability are yet to be stipulated. This thesis explores one novel application: *visual animal biometrics*.

The work demonstrates that vision can achieve an *automatic identification of animals* filmed in their natural habitat. The thesis proposes and evaluates algorithms for *detecting a species* and – in the case that the animals carry *Turing patterns* – for *recognising individuals* in visual material comprising various poses, changing lighting and clutter. Lions, plains zebras and African penguins serve as sample species to showcase the different capabilities and limitations of the approach.

First, an algorithmic framework for species recognition is discussed. In particular, it is shown that the appearance context of a number of *reference points* on coat patterns contains a species-specific component, which can be utilised for achieving a robust detection of the investigated animal specimens. The proposed model employs *boosted point-surround classifiers* as local appearance descriptors.

Second, it is illustrated how *pose-normalised texture maps* of Turing-patterned coat regions can be extracted based on the extracted reference points. The maps are shown to contain *individually-specific features* which – using an extension of *shape contexts* – can be represented as deformation-robust sets of histograms. Finally, it is discussed how *distance measures* can be used for comparing these sets with population databases to retrieve animal identities.

In order to provide a practical proof of concept, a *prototype system* was tested in a colony of African penguins and in a preliminary study on a photo collection of plains zebras. The results indicate a system performance that allows for disambiguating individuals with a level of confidence sufficient for tasks of population monitoring. This outcome marks a promising first step towards an automated, truly non-invasive observation of wild animal populations, which would benefit field-based biology and assist the conservation of species in decline.

ACKNOWLEDGEMENTS

I would like to express my gratitude to those numerous people who have continuously supported and encouraged me during the period of my life that I dedicated to this work.

First of all I would like to thank my advisors Barry Thomas and Neill Campbell for their invaluable guidance and help both on a scientific as well as a personal level.

Second, I would like to thank the physicist and penguin enthusiast Peter Barham for his inspiration and feedback, especially during field sessions and various ardent discussions. I thank him also for his initiative in helping to raise funds for taking ahead this work to be applied with a prospect of conservational relevance.

Furthermore, I would like to thank Janko Ćalić and Richard Sherley for practical support, encouraging conversations and valuable feedback. A special thanks also goes to all my other inspirators, friends and helping hands at the University of Bristol (foremost Innes Cuthill, Majid Mirmehdi, Nigel Franks, Sion Hanunna, David Gibson, Oli Cooper, Ben Daubney and Lisa Gralewski).

I would like to acknowledge the various collaborators in the United Kingdom (Granada Media, Matrix Data), South Africa (Les Underhill, Rob Crawford, Mario Leshoro), France (Sophie Grange, Patrick Duncan) and Germany (Anja Wasilewski) for providing access to their image collections and/or study populations. In addition, I would like to mention that I am grateful for the funding received from 3CResearch and the Earthwatch Institute.

Above all, I would like to say thank you to the people who have been backing me all the way: foremost my mother Inge, my fatherly friend Gustav, and all of my family and friends who carried me through the ups and downs of compiling this thesis.

DECLARATION

I declare that the work in this dissertation was carried out in accordance with the Regulations of the University of Bristol. The work is original, except where indicated by special reference in the text, and no part of the dissertation has been submitted for any other academic award. Any views expressed in the dissertation are those of the author.

Date: *16/06/2008*

Signature: *Tilo Burghardt*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

'I suggested that a system of [...] substances [...] reacting together and diffusing [...] may later develop a pattern or structure.' [201]

(Alan M. Turing, 1912 - 1954)

1.1 Preface: Evolution, Camouflage and Visual Uniqueness

Driven by an ongoing evolutionary adjustment, life forms on earth have been adapting to the environmental conditions of their habitats for millions of years. As a result, a multitude of 'patterns of life' have emerged. Some of the patterns are of a visual nature, which form textures on the surfaces of organisms. In the animal kingdom, such prominent visual markings are commonly known as *coat patterns.* They often appear on major body parts as colourations of either fur, feathers, scales, skin or cuticle. These patterns have been repeatedly recognised as 'smart' visual designs [175] that can boost the likelihood of a species' survival [47]. Examples include 'eyespot' patterns on butterfly wings, coloured rings on snakes, contrasting stripes on zebras and dots on cheetahs. Figure 1.1 depicts a small selection of species whose surface colouration has visually adapted to their environment.

For a multitude of animals their markings provide camouflage from predators [192], rendering an individual to some extent indistinguishable from its natural environment or its group of conspecifics [194]. A vast variety of specialised forms of such deception methods have evolved. Most of them draw on imitation techniques, be that cryptic resemblance [54, 139] (where the pattern is composed out of random colours of the surrounding habitat) or disruptive colouration [47] (where the animal's visual outline is broken and disguised by contrasting peripheral markings). Other adaptation strategies provide a visual resemblance to an altogether different species, known as mimicry [14, 143], or draw upon the mere

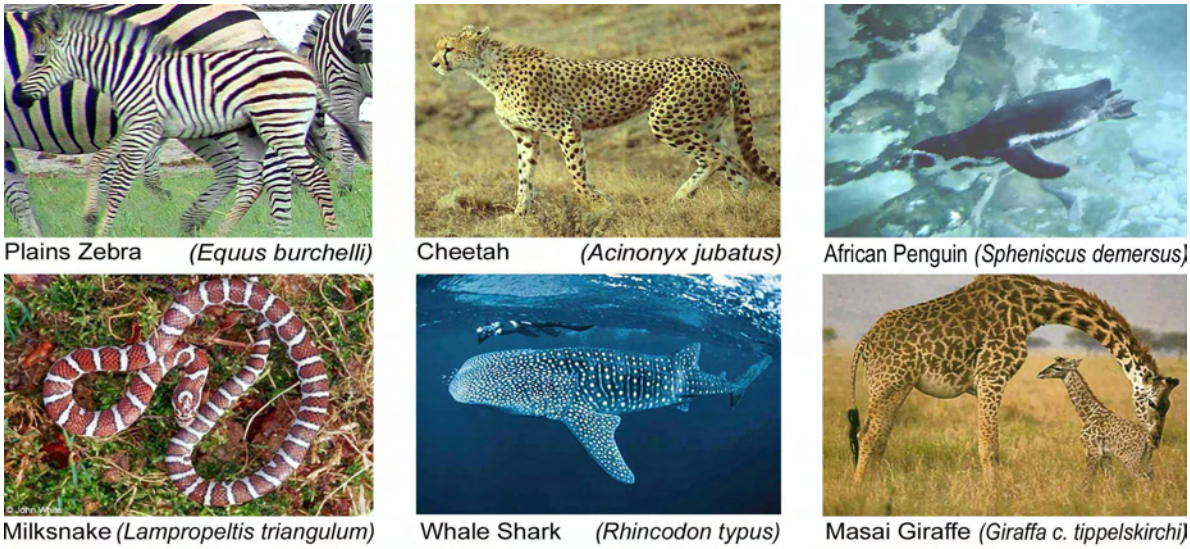


Figure 1.1: **Selection of Coat Patterned Species.** The images show 6 very different species which have visually adapted to their habitat by developing surface patterns that provide effective camouflage within their environment or group of conspecifics. [image sources: I02, I03, I09, I07, I05, I04]

pretence of dangerous properties, known as forms of aposematism [55, 74].

Surface patterns that implement these imitation concepts have long been suspected[1] to result from entropy-reducing, widely self-organising formation processes. One frequently found pattern group, the class of so-called 'Reaction-Diffusion' patterns, has been modelled in mathematical detail and is, to this day, one of the best researched and understood auto-generative systems[2]. The underlying theory was introduced as early as 1952 by *Turing* [201], who thereby pioneered the discipline of biomathematics. Belousov (as reported by *Tyson* [204]), later *Prigogine, Lefevre* [160] and *Castets et al.* [29] conducted practical experiments on chemical reactions backing Turing's theoretical predictions: they showed that chemical systems can autonomously reach stable states of non-trivial spatial patterns that exist *far away* from the thermodynamical equilibrium of overall homogeneity.

Interweaving these findings in 1988, the British mathematician *Murray* [145] successfully explained the structure of a number of animal patterns by hypothetical Turing systems, drawing once more on the co-occurrence of reaction and diffusion in chemical substances[3].

[1] The physicist *Schrödinger* indicated the importance of entropy-reducing processes for the generation of biological patterns in broad principle already in his 1943 lecture series 'What is life?' [183].

[2] The reactants, that is the actual chemicals that drive Reaction-Diffusion in animal skin/cuticle, remain widely unknown, although a few, widely theoretical models have been proposed, e.g. [78, 112, 135, 136, 137]. Note that the reactants do not necessarily represent the colour pigments of the surface (such as melanin). Instead, they are suggested to be mediating chemicals of the colouration process.

[3] *Turing* [201] employs the term 'morphogenes' (*morphe*=form and *genea*=generation) for the chemicals involved in the process of pattern formation, stressing the 'creational potential' of the reactants.

Most interestingly, a source of individual 'randomness' is present in the very early stages of pattern formation: minimal quantitative differences in embryonic parameters – such as body shape or initial distribution of chemical densities – exist and *are specific* to a single individual. During pattern formation these perturbations undergo a significant, self-induced amplification, a phenomenon more commonly known as the *butterfly-effect*[4].

It is this process that leads to a qualitatively strong intra-species diversity of structural details in the fully developed coat patterns of a multitude of animals. Figure 1.2 illustrates the extent of these variations in two sample species.

Figure 1.2: **Individuality of Coat Patterns in Plains Zebras and African Penguins.** The images show plains zebras (left) and African penguins (right), two examples of species that develop highly individual markings while following a species-wide, visual theme. Note the unique distributions of line bifurcations on the zebras and of spots on the penguins. Individual features are concentrated around the upper front leg and the chest area, respectively. [image sources: I02, I01]

In general, pattern details differ significantly from individual to individual while following a wider theme typical for a species where the pattern variance is concentrated in a few areas of the body surface, e.g. scapular stripes on zebras or the chest spots on African penguins. These regions *alone* carry feature combinations that are often unique and complex enough to accurately characterise individuals within entire populations.

This thesis exploits the naturally evolved, visual properties of coat patterns. By using *visual animal biometrics*, that is applying vision and biometric techniques to digital photo and video data of animals, it proposes and evaluates algorithms for the automatic retrieval of species membership and individual animal identity.

Before the central research theme and the associated problems are discussed in detail, the following section briefly summarises the interdisciplinary relevance of the subject in order to situate the work within its broader scientific and applicational context.

[4]The meteorologist *Lorenz* [129] introduced the term 'butterfly-effect', referring to the phenomenon witnessed in a number of complex, yet deterministic systems (e.g. the earth atmosphere): minute changes of the system state (e.g. induced by the wing movements of a butterfly) can cause a significant, global pattern alteration (e.g. a large scale weather event) in later stages of the system's evolution.

1.2 Interdisciplinary Relevance: Monitoring Wildlife

Field-based biological studies rely heavily on data about the movements and whereabouts of individual animals. At present when researchers study large animal populations based on sighting information, they commonly adopt an approach of sampling and analysis known as *capture-mark–recapture*[5] [18]. The method and its derivatives are widely used for the estimation of population vitality parameters such as movement, survival and growth from usually incomplete observation data; but it can also be applied to approximate actual population sizes or to aid behavioural studies [45, 216]. As the name suggests, the method relies on the availability of relatively large collections of data about *repeated sightings or captures* of the same individuals or groups.

For the collection of sighting data in penguin populations in particular, the birds are in most cases caught and fitted with artificial identifiers that can be read every time the animal is observed or recaptured. Researchers usually employ visual colour or number tags [152], paint, passive integrated transponders (PITs) [82] or external radio- or satellite transmitters [53] as unique markers. These techniques have been invaluable for the understanding of population dynamics; yet they are invasive. All involve capturing and handling the animals to fit the devices.

Penguins that carry tags or other appliances may be detrimentally affected [11, 15, 45, 46]. Markers might loosen, get lost or wear out [34], affecting study results and possibly harming members of the species under observation [2, 35, 176]. Metallic flipper bands[6], for example, suspected to cause higher energy costs of swimming [40, 41], increased foraging-trip duration [53], substantially lower survival in the first year after banding [71], flipper damage from partially opened bands [34], and decreased reproductive success [71, 77].

In addition, all artificial marking methods mentioned above comprise significant technical or economical drawbacks: the reading of visual markers usually involves manual monitoring on large scale while radio and satellite tags are very costly, priced up to $10,000 per device. Subcutaneous transponders, on the other hand, are hidden from visual observation and typically operate only in decimeter proximity to a reader station or hand-held device.

[5] The method is also referred to as ‘CMR’, ‘capture-recapture’, ‘sight-resight’ or ‘band recovery’. Several specific techniques for further statistical analyses exist including the (most commonly used) *Peterson-Lincoln Index* [156], the *Jolly-Seber Index* [156] and the *Cormack-Jolly-Seber Model* [121].

[6] Metallic flipper bands have been used frequently for tagging penguins. Being both heavy and sharp they can inflict injuries as witnessed by myself at multiple occasions during the field sessions of this project.

Generally, the recording of dense population measurements proves to be of growing importance since an increasing human interference with vital ecosystems of the planet generates a pressing need for intelligent conservational and ecological policies. Effective political decisions rely on well supported studies that accurately capture population developments.

Biological and environmental researchers, therefore, desire non-invasive, easy to use, and relatively inexpensive means to identify species as well as individual animals, means that operate fast and robustly under field conditions and on large populations.

1.3 Problem Outline: How to Fingerprint Animals Visually?

The recent evolution in sensor and video technology has made available low-cost hardware that can provide live streams of high-resolution images revealing detailed visual features of filmed objects of interest.

It can thus be hypothesised that the species-specific, yet individually unique visual appearance of animal coats coupled with the availability of imagery of sufficient resolution may provide all the raw data necessary for an automatic identification of individual animals *without* a need for synthetic markers of any kind. This poses a computer vision challenge robustly to extract the few bytes of animal identity information from potentially massive volumes of high-resolution image and video data. Given this general scope, a number of problem classes arise that reflect the limitations of currently available techniques and, thereby, determine the research directions for the project undertaken:

Fingerprinting Coat Patterns. A compact and generic representation and comparison scheme for the unique component of coat-patterns has not been proposed in the literature yet; species-specific templates are rare and, if existent at all, they are optimised for totally or partially manual identification of a single species. The formulation of a robust visual methodology for the individual-specific representation and comparison of Turing-like coat-patterns is a primary condition for handling the biometric identity data inherent to animals.

Efficient and Effective Species Detection in Natural Environments. Natural habitats constitute highly cluttered and uncontrolled environments. Under these conditions animal patterns are difficult to separate from background components and undergo significant alteration whilst being registered. As a consequence, the appearance of the same object is found to be different at different times of measurement, introducing high intra-class vari-

ation. The alteration is partially due to changes in lighting, partial occlusion, variation in viewpoint parameters as depicted in Figure 1.3(a) and (b).

Figure 1.3: **Acquisition-dependent Variability of Coat Patterns.** The images show individual animals with changing (projected) appearance in stills taken from video sequences. The variation is mainly due to **a)** alteration in scale, occlusion and lighting conditions, **b)** variable azimuth/pose and **c)** organic non-linear deformation. Note the distortion of the biometrically unique patterns, that is zebra stripes and penguin spots. [image sources I01, I10]

Most human biometric applications cope with variations by introducing artificially controlled sensor environments such as standard lighting or fixed viewing conditions. However, filming in natural habitats cannot avoid altering influences.

Consequently, recognition techniques that are invariant – or at least robust – towards high intra-class variability of patterns are of special relevance to this work. The subject in general has been a long-standing research topic in both the computer vision and the biometrics communities. Nevertheless, the optimisation and validation of existing techniques in the domain of organic objects has mainly been concerned with the human subject. Automated visual animal recognition has not been the focus of biometric studies as yet.

Registration of Variable 3D Patterns. In many species the unique coat pattern of interest covers main body parts. While this increases general visibility and aids registration it also prominently reveals the three-dimensional nature of an animal's surface: coat patterns are not planar, but spatially distributed over 2D surfaces embedded in 3D space, causing perspective distortions and partial (self)occlusions depending on the animal's pose.

This scenario contrasts with conditions of most (small scale) human biometric entities, such as iris patterns [48] or fingerprints [101], which are generally assumed to be planar textures. Consequently, human biometric entities are traditionally registered, modelled and processed in only two spatial dimensions. Only recently facial detection systems, presented for instance in [56, 150, 198], have started to deal with the changes of facial pose employing rigid or active 3D models in conjunction with viewpoint-robust 2D appearance descriptors.

Non-Linear Deformation. Most organic bodies are flexible entities that change their surface shape together with the coat patterns imprinted. Thus, in addition to acquisition-induced and rigid-structural alteration which may be approximated by linear models, the surfaces of organisms undergo significant, non-linear deformation during motion.

As a consequence, any Euclidean, similar, affine or perspective spatial constraint will fail to describe deforming coat patterns. As a matter of fact, most animal species fall into the class of non-linearly deformable, self-occluded (thus singularly mapped) objects as illustrated by comparison in Figure 1.4.

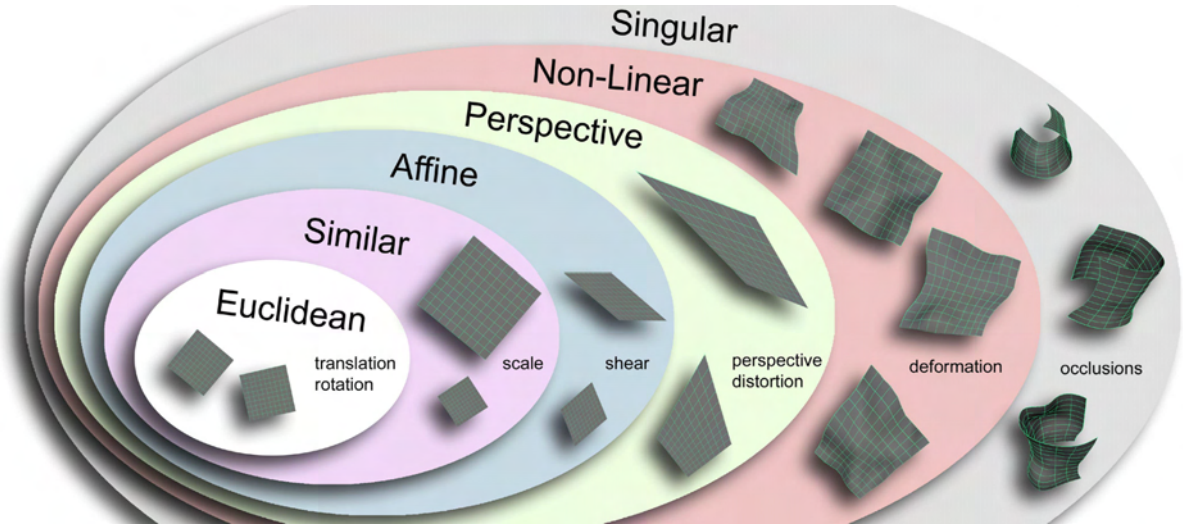


Figure 1.4: **Hierarchy of Image Transforms.** The chart illustrates different forms of geometrical transforms organised in a hierarchical system. The modifying operations associated are specified.

The severe effects of such distortions are, for instance, depicted in Figure 1.3(c) showing the flexibly changing body shape of an African penguin walking.

Tackling the issue of non-linear deformation is an emerging research topic in computer vision. Whilst general-purpose vision systems (in the narrow sense) still struggle to deal robustly with the broad task of deformable object recognition, application-specific models, on the other hand, often perform sufficiently for practical applications [99, 170].

The biometrics community has also started to create models for non-rigid pattern deformation designed to aid biometric applications for human identification, namely 2D fingerprint [169, 170] and 3D face scan systems [131] where the degree of non-linear pattern variance is comparatively low.

Non-Cooperative Scenario. Sufficient robustness of biometric systems often relies on the opportunity to create repeated system inputs incase patterns can not be recognised the first time. This strategy demands some level of cooperative behaviour from the individuals examined during the measurement process. Since wild animals will not cooperate, practical problems arise regarding the design and set-up of a biometric animal identification system so as to maximise the chance of successful visual capture and identification.

1.4 Aims and Objectives: A Real-world Proof of Concept

The primary goal of this work is to provide a proof of concept, showing that it is theoretically sound and practically feasible automatically to identify a Turing-patterned species and its individual members from imagery taken in natural habitats.

This task involves tackling the problems mentioned above in a systematic way, focussing on recognition in cluttered and highly variable environments. Therefore, special attention must be paid 1) to *efficient means* of species detection despite the high degree of variance present, 2) to the *robust representation and comparison* of coat patterns, as well as 3) to a *practical evaluation* of all methods proposed. In the following section these main research objectives are categorised and outlined in further detail.

Species Recognition. Firstly, the thesis aims to examine the practical applicability of statistical recognition techniques to the detection of complex, deformable, and individually variable species in natural scenes. In particular, model-based, statistical techniques to locate and recognise a patterned species in visual media are to be implemented and evaluated.

This covers tackling the vision-specific problems of 1) efficient and robust *appearance description* and *detection* of coat patterns in various poses against cluttered backgrounds, 2) the organisation of primitives/features into *flexible models* allowing for organic deformability of their spatial configuration.

Biometric Identification. Secondly, the system must yield recognition information down to the granularity of an individual. This involves the construction of schemes 1) to extract and generically represent the *unique visual features* of Turing-patterned animal coats, 2) robustly to partition these representations into sets describing the same individual despite non-linear pattern distortions, and 3) to evaluate and interpret the results with regard to the task of monitoring animals. The latter covers the exploration of theoretical boundaries of the approach as well as a critical discussion of observed system characteristics.

Application. Finally, the work needs to address the practical aspect of the recognition task by testing a prototype system in real-world scenarios, evaluating the methods under environmental conditions in a large scale animal colony bearing large numbers of individuals. This involves gathering practical experience with regard to the future development of a generic, widely applicable enabling technology for aiding field-based biological research.

1.5 The Approach in a Nutshell: Locate-Pose-Identify

In this thesis it is proposed to solve the task of individual animal identification by employing a coarse-to-fine-to-coarse strategy, probing image regions of interest for meaningful feature combinations, which are progressively recomposed into species descriptors, pose models, maps of surface textures and, finally, individual animal identifiers.

In particular, species-specific *key reference points* on coat patterns of a species are initially identified. They are used to infer the location and pose of *surface models* wherefrom pose-corrected *2D coat texture maps* are extracted. Specific *pattern features* therein are finally transformed into individually characteristic *1D profiles* that represent the *identities* of the animals filmed. Note that this extraction process forms a progressive recognition pipeline, which can be divided into three cohesive subtasks:

Locating. The initial detection of species-universal coat features *localises* key surface positions of a species in images. So the attention is focussed onto a few surface regions most characteristic for the species of interest.

Posing. The detected landmarks semantically correspond to counterparts on a surface model, which is *posed* into the scene. Back-projection from the pose-normalised model yields a corrected view of the uniqueness-bearing animal texture of interest which is passed on to the biometrical system component.

Identifying. In a last step, a set of individually distinctive feature locations are identified in the Turing-like coat texture. A histogram-based comparison of the feature set with a population database finally reveals the *animal identity* desired.

The paradigm 'locate-pose-identify' provides the conceptual backbone for the identification methodology proposed. Figure 1.5 visualises how the approach is reflected in the structure of the thesis.

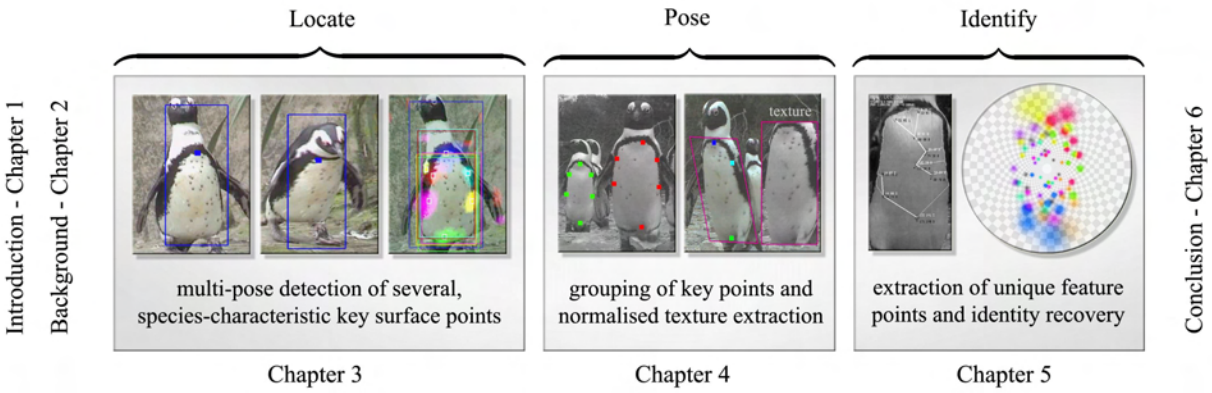


Figure 1.5: **Structure of the Thesis.** The chart illustrates the proposed recognition pipeline designed to achieve individual animal identification from visual material. It can be seen that the thesis is organised along a pipeline of components, sectioning the core content into 3 main chapters.

1.6 Thesis Structure

The thesis is organised into six chapters. The core chapters 3, 4 and 5 describe the research and work conducted along an increasingly abstract information flow as explained in Figure 1.5. The other chapters place the work into its scientific and applicational context:

Chapter 2 reviews relevant previous work in the research areas of interest. It briefly introduces the history, foundations and various recent approaches in the fields of human and, specifically, non-human biometrics together with relevant techniques of entity matching and recognition. Advantages and shortcomings of existing approaches are discussed with respect to the task at hand. Special attention is payed to techniques that deal with highly variable objects. In particular, the statistical detection framework by Viola and Jones [212] is reviewed in depth. Flexible spatial models and the matching of distorted landmark data are also surveyed. During a concluding section, the chapter gives a brief introduction to the theory of Reaction-Diffusion since properties of this generative pattern model are later exploited for representing a species and its individuals more effectively and efficiently.

Chapter 3 introduces a model for an appearance-based detection of species-characteristic key points on the animal coat. The chapter describes a detection method based on the Viola-Jones framework that recognises sets of such key points on animal coats in visually cluttered scenes and in the case of high intra-class variability. Several specific extensions to the established statistical detection technique are proposed and evaluated on wildlife content filmed in natural environments. The chapter also discusses an optimisation to achieve close-to-realtime runtime performance given megapixel image resolution. Finally, the chapter investigates multi-detector models to enhance the coverage of pose space.

Chapter 4 illustrates how the component-like key point representation can be exploited to extract surface textures of biometric interest. A spatially flexible model for spatial representation and detection termed *feature prediction trees* is used to associate of key point instances to animal instances. The resulting correspondence sets are then exploited for fitting geometrical surface models that allow for the extraction of *pose-corrected texture maps*.

Chapter 5 approaches the biometric task of associating the unique part of the normalised animal texture with an individual identity. The chapter illustrates a technique to extract specific features of the unique component of Turing-patterns and to organise this information into compact templates that can represent the observed individual. A method is outlined robustly to compare templates despite local distortions of the original patterns. Subsequently, the resulting distance metric is interpreted with the help of a population database to retrieve a specific identity of the individual. The chapter concludes by presenting practical identification results based on experiments involving real animal populations.

Chapter 6 summarises the presented work and discusses its virtues, limitations and applicational prospects. Finally, future research directions that could enhance and extend the methods and techniques proposed are indicated. ■

Chapter 2

BACKGROUND

'The sciences do not try to explain, they hardly even try to interpret, they mainly make models [...] that work.' [148]

(John von Neumann, 1903 - 1957)

2.1 Chapter Overview

This chapter gives an introduction to the discipline of biometrics and reviews previous work in vision considered relevant to the theoretical structure and to the practical design of the monocular animal identification system proposed.

Visual wildlife biometrics is a relatively new approach to animal identification. The subject appears to be situated between several vision-related disciplines; computer vision, biometrics and machine learning. During the last decade, however, these subjects have become partially merged showing increasing overlap in their concepts, methods, and techniques employed *for modelling* the real world.

Acknowledging those developments, this review is organised along the lines of modelling strategies relevant to the proposed 'locate-pose-identify' paradigm. Apart from reviewing the general design of biometric systems, the fields surveyed cover approaches and ideas to object detection, structural recognition, and techniques for matching landmark patterns:

- biometric systems for humans and animals, their design and performance (Section 2.2)
- sparse landmark matching and deformation-robust comparison (Section 2.3)
- prototypical, statistical and structural models for object recognition (Section 2.4)
- biomathematical Turing model for textures of coat patterns (Section 2.5)

2.2 Biometric Systems

2.2.1 Basic Concept: Measuring Living Beings

There are three means of authentication, that is approaches to verifying an individual's identity: 1) by some secret data, e.g. a password, 2) by possession of a device, e.g. keys or tags and 3) by what the individual actually is, e.g. its anatomy, physiology and behaviour [154]. The scientific domain investigating the last is known as '*biometrics*', a term derived from the Greek words *bios* (life) and *metron* (measure). This very etymology captures the core concept of the approach: biometric identification systems are designed to analyse *measurements* taken from *living beings* for the purpose of classifying their population, that is either authenticating (1-to-1 match) or identifying (1-to-n match) individuals or groups.

Clearly, the concept of directly measuring an organism – instead of relying on synthetic markers – renders the approach highly suitable for scenarios where little interference with the subjects is permitted, e.g. passive biometrics such as secret surveillance, public observation or, in the case at hand, wildlife monitoring aiming for minimal disturbance.

2.2.2 A Brief Historical Note: From Clay Prints to Population Databases

Biometric identification has had a place in human life for more than 4000 years. The approach has been predominantly applied to the problem of profiling and retrieving the identities of humans. Dating back as far as 2000 BC, the ancient Babylonians accompanied contracts with fingerprints pressed in clay to prevent forgery [25, 42]. Figure 2.1(a) depicts one of the earliest findings of this kind. Subsequently, skin impressions of either hands or feet were used as signatures by early cultures in Asia [122] and North America [25].

Although the uniqueness of skin ridges is mentioned in scripts by renaissance scholars[1], it was only at the onset of the industrial age that European scientists started investigating biometrics on a quantitative, truly scientific basis. The first physiological measurements on humans, that is on their skin ridges, ear-shape, head-shape etc., were systematically categorised and analysed[2] in the mid 19^{th} century.

[1] Written during the 1680's, scripts by *N. Grew* [43], *G. Bidloo* [24] and *M. Malpighius* [133] describe unique entities of human physiology mainly focussing on the observation that human skin carries individual features.

[2] Europeans who investigated biometric entities during the industrial period include *J. E. Purkinje* [162], *W. Herschel*, *A. Bertillon* (as described by *Garfinkel* [76]), *F. Galton* [73] and *E. Henry*. They mainly studied the physical uniqueness inherent to humans for anthropological, economic and, only later, for forensic purposes.

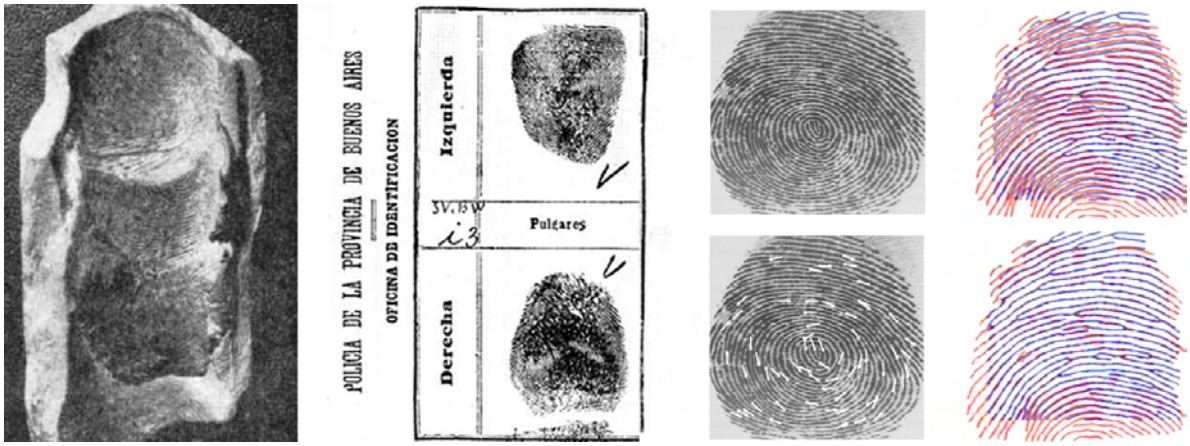


Figure 2.1: ***Representation of Fingerprints throughout the History of Biometrics.*** The images illustrate the historical advances of biometrical techniques in representing fingerprints. **(a)** clay print from mesopotamia dated about 2000 BC; **(b)** ink prints taken by the Buenos Aires police in 1891; **(c)** digital fingerprint image with extracted minutae (ridge endings and bifurcations) and ridge directions superimposed – the state of the art in 1996; **(d)** fingerprints aligned using 2D thin plate splines controlled by minutae matching (top) and ridge curve matching (bottom) conducted in 2006. [images taken from *Cummins* [42], *Rodriguez* [168], *Jain* [102] and *Ross et al.* [170]]

Ever since the introduction of precise sensing devices (e.g. cameras), modern forensic analysis (e.g. population catalogues [168]), and increasing computerisation have gradually removed the need for human inspection. They have made possible the engineering of largely autonomous, active biometric systems that are now applied in a number of security-relevant domains. Most of them, however, rely significantly on the user's collaboration. Table 2.1 gives an overview of performance benchmarks of the most widely applied, currently available human biometric systems for comparison with the animal identification system proposed.

The level of robustness, diversity and economical impact of biometric techniques is still steadily and rapidly increasing. In fact, biometric techniques are now being applied to human populations on a large scale: in January 2006 the United States Federal Bureau of Investigation (FBI) officially stored about 47 million fingerprints in its IAFIS[3] database and entity-specific biometric standards for humans are under development within ISO[4] and INCITS[5] [207]. The United States of America as well as the European Union are currently

[3]The Integrated Automated Fingerprint Identification System (IAFIS) is the main U.S. fingerprint and criminal history system linking several subsystems under a common standard. It is maintained by the FBI. Further information can be found at the US National Science and Technology Council (NSTC) [206].

[4]The International Standards Organisation (ISO) is the major international body for general standardisation covering more than 150 national member organisations and nearly 3000 technical bodies.

[5]The Inter-National Committee for Information Technology Standards (INCITS) is a major body of

Entity (method)	**EER [%]**	**FAR/FRR [%]**	**FER [%]**	**TS [kB]**	**SSD [m]**
DNA (STR)	$\approx 10^{-7}$ [195]	close to 0 [205]	≈ 0 [195]	> 50	contact
retinal vessels (IR)	$\approx 10^{-5}$ [205]	$\approx 0/\approx 0.4$ [173]	≈ 0 [173]	≈ 0.05	≈ 0.075
iris (IR)	$\approx 7.6 \cdot 10^{-4}$ [205]	$\ll 10^{-3}/\approx 6$ [103]	≈ 7 [103]	≈ 0.3	≈ 0.3
fingerprint (MB)	≈ 0.2 [205]	$\approx 0.1/\approx 0.25$ [217]	≈ 4 [217]	≈ 0.3	contact
facial features	n/a	$\approx 10/\approx 4$ [155]	n/a	≈ 0.05	≈ 20
voice pattern	n/a	$\approx 10/\approx 15$ [173]	1-30 [173]	≈ 0.02	≈ 3
penguin spots	n/a	$\approx 10^{-1}/\approx 8$	≈ 80	≈ 0.1	≈ 3 - 7

Table 2.1: **Current Performance Characteristics in Biometrics.** The table summarises the authentication performance details of several biometric techniques. A representative subset of the most widely spread techniques in human forensics is listed and sorted by identification accuracy. Note the enormous variance in characteristics amongst the methods. At the bottom part, details of the proposed animal identification system are given for comparison. (Key: FAR...false acceptance rate, FRR...false rejection rate, EER...equal error rate (rate where FAR=FRR), FER...failure to enrol rate, TS...template size, SSD...sensor-subject distance, STR...string comparison methods, IR...infrared capture, MB...minutae based; See the individual references for more details.)

starting to introduce biometric passports for their citizens, creating controversies over the practicality as well as the philosophical and ethical justification of biometric registration procedures in general. Such ventures on the international stage dramatically illustrate an ongoing 'biometric revolution' that leaves an ever more prominent footprint on the procedures that underpin life in modern societies.

Meanwhile, the research community has started to leap ahead from the classical forms of visual biometrics, e.g. fingerprint [102, 170], face [140, 203] or iris recognition [48, 197], into new territory. Recent biometrical research starts focussing on passive and ubiquitous biometrics [119], on multi-modal biometric fusion [114], on behavioural fingerprinting [75], and on a number of unconventional identification techniques using signatures of gait [44], hair pattern [66], ear-shape [120], and even odor [113]. Yet the research efforts concentrate almost exclusively on the human subject. While DNA-based methods have become an essential tool for the identification and classification of non-human species in multiple natural sciences [195, 205], the application of *visual biometrics* is still broadly limited to humans.

2.2.3 Biometric Entities and Individually Characteristic Information

In general, two main categories of systems can be differentiated depending on the descriptiveness and granularity of the classification performed. If only broad classes of a population are to be identified by (commonly scalar) measurements, e.g. body weight, eye-colour,

standardisation in the field of Information and Communications Technologies. It is accredited by, and operates under rules approved by the American National Standards Institute (ANSI).

height etc., the methods are referred to as *soft* or *weak biometrics* [206].

In contrast, procedures that aim to identify individuals with high confidence levels are known as *strong* biometrics [114]. These methods commonly rely on statistics from multi-dimensional aspects of physiological or behavioural characteristics that differ (significantly) within the population of interest. This work exclusively focusses on strong biometrics since only they provide reliable identity data – rather than subclass membership information – as needed for individual-based applications in biological and conservational sciences.

Strong biometrics require the presence of a high-dimensional configuration space spanned by some of an individual's measurable properties. Categories of such properties are known as *'biometric entities'*; they are commonly differentiated by their form and by their location on an organism's body as well as by the type of information used.

Naturally, an organism carries multiple such entities. Figure 2.2 depicts a selection of biometric entities used for human identification.

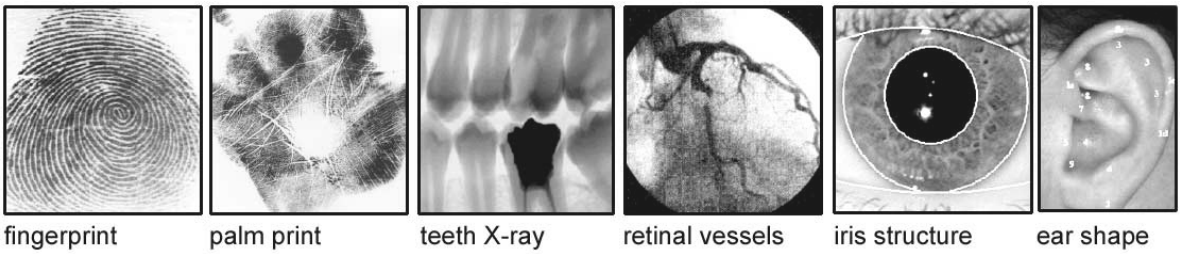


Figure 2.2: ***Selection of Human Biometric Entities.*** The images illustrate some unique human biometric entities that can serve as a source of individuality data. Illustrated are the unique structure of (from left to right) ridge lines in finger and palm prints, shapes of X-rayed teeth, blood vessel shape in the retina, iris patterns and ear shapes. [images taken from *Jain et al.* [103]]

A number of human biometric entities can also be found (and theoretically used) in mammals, e.g. iris patterns, skin-ridge prints, voice etc.

However, measuring them to the necessary detail proves difficult in field scenarios, if not infeasible. In contrast, coat patterns, while not present in humans, provide an entity of full body size that is (even intended to be) visible over distance, providing excellent measurability. Table 2.2 compares coat patterns to classical biometric entities with respect to the properties relevant for identification.

Characteristic	Intuition	D	R	I	F	Fa	V	CP (this thesis)
uniqueness	distinctive information content	↑	↑	↑	↑	↓	↓	↑
permanence	extent of temporal change	↑	↑	↑	↑	↓	↓	↑
universality	population coverage	↑	↑	↑	↕	↑	↕	↕
measurability	simplicity of extraction	↓	↓	↕	↕	↑	↕	↑
comparability	simplicity of comparison	↓	↕	↕	↕	↓	↓	↕
invasiveness	interference with subjects	↑	↑	↑	↕	↓	↓	↓
performance	accuracy, speed, robustness	↑	↑	↑	↕	↓	↓	↕
acceptability	level of society support	↓	↓	↓	↕	↑	↑	×
circumvention	ease of active cheat	↓	↓	↓	↕	↑	↑	×

Table 2.2: **Characteristics of Biometric Entities.** The quality of a biometric entity can be characterised by five essential and four system-dependent measures listed in the table above. Key: ↑...high; ↕...medium; ↓...low; ×...n/a; D...DNA; R...retina; I...iris; F...fingerprint; Fa...face; V...voice; CP...coat patterns (penguin spot patterns and scapular zebra stripes in particular); [extended comparison based on an earlier suggestion by *Jain et al.* [100]]

The characteristic properties of a set of biometric entities are termed a *'biometric profile'*. The profile is used to represent the individual in an information system or population database. The substitution of an identity by a profile is, however, based on the assumption that an individual's identity is *intrinsic* to the information locked in its biometric entities [124, 222]. Hence, in order to qualify as a biometric entity, a feature set must satisfy a number of properties. *Jain* [100] presented a set of properties that approximate this condition. His suggestions are listed and extended below (they also form the basis of Table 2.2). I suggest categorising the properties into three fundamental groups of accountable factors:

Physiological Factors. First, the feature set has to contain components that are both *singular* in their physical/behavioural structure (uniqueness) and *constant* in their appearance over time (permanence). In addition, a feature set should be *all-inclusive* in the sense that – ideally – each and every member of the population carries it (universality). Naturally, the process of identifying potential entities in a species is guided by these criteria.

Technological Factors. Second, and this is an engineering requirement, biometric entities must be 'collectable', suggests Jain. Here, it is proposed to split his criteria into *measurability* and *comparability* since these two features directly relate to the components of feature extraction and pattern matching found in the design of actual identification systems.

Interactive Factors. In terms of interaction requirements, *invasiveness* (not mentioned by *Jain* [100]) is relevant to wildlife applications. Passive visual methods are non-intrusive by their nature and, therefore, highly applicable. Social *acceptability* and active *circumvention*, on the other hand, are mainly important in human security applications.

In later parts of this work it will be argued that a multitude of coat-patterns carry all the physiological properties necessary for biometrical identification. In particular, it will be shown that, for a number of species, coat patterns are a universal feature and remain stable for life after an initial lay-down period. It will be demonstrated that coat-patterns contain both a species-specific as well as an individually distinctive pattern component. By isolating and exploiting each of the two data categories, it will be shown that measurability as well as comparability of coat patterns can be achieved, that is species members can be registered in their habitat and individually identified against conspecifics, respectively.

Assuming that the five fundamental conditions mentioned – namely universality, uniqueness, permanence, measurability and comparability – are met, biometric systems can operate without the use of synthetic markers. Accordingly, a biometric profile cannot be separated from an individual in a natural way – if not occluded, they are omnipresent. Coat patterns can, therefore, be observed with a relatively low level of cooperation of the individual investigated (passive biometrics). Thus, for the purpose of animal identification, coat pattern biometrics can potentially excel over other identification techniques since they are applicable in wildlife scenarios 1) where cooperation is unlikely, 2) where devices are difficult to fit, and 3) where interaction with or disturbance of endangered species is to be avoided.

2.2.4 Design of Biometric Systems: Components and Operational Modes

Having discussed the properties of biometric entities, the focus is now shifted towards the methodology of extracting the uniqueness-bearing information inherent to them. Modern biometric designs achieve identification by a number of system components that perform conceptually different tasks along one, widely common recognition pipeline. In accordance with previously suggested models, by for instance *Jain et al.* [103] or *Hüseyin* [98], Figure 2.3(a) depicts the broad architecture of biometric systems in form of a flow chart. The model comprises the stages of acquisition, extraction, storage, matching and application. The extraction component is resolved further where green bars indicate how this ties in with the 'locate-pose-identify' paradigm proposed. It can be seen that the registration (object detection) and the normalisation (pose identification) components are identified as tasks in their own right. This is, so to say, a 'vision view' on biometric systems, a view that recognises the complexity of locating and posing the biometric entity in the first place.

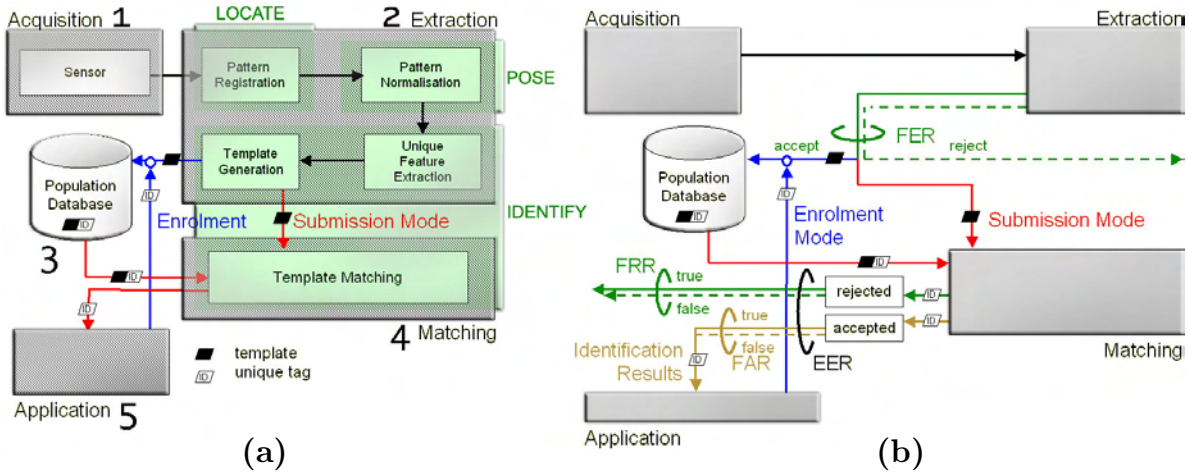


Figure 2.3: ***Components and Performance Measures of Biometric Systems.*** **(a)** The flow chart depicts the general components and their interrelation in a biometric system. Note a difference between the information flux of the enrolment and the submission mode. **(b)** Schematic measurement points of performance characteristics. (Key: FAR...false acceptance rate, FRR...false rejection rate, EER...equal error rate (rate where FAR=FRR), FER...failure to enrol rate)

Each of the five components fulfills a different task:

Acquisition. The initial measurements on individuals are taken by one or multiple sensors[6] referred to as *acquisition devices*. This work deals exclusively with optical input from the visual spectrum[7].

Extraction. Separating the descriptive components of the biometric entities from redundant data in the images is taken care of by a *registration* procedure applied to locate the entity within the measurement. During this process parameters are generated that describe the context of the detection such as pattern size, lighting, viewpoint, pose or the pattern visibility. These parameters are then used to *normalise* the extracted pattern, decreasing the intra-class variance caused by different acquisition conditions. The raw biometric patterns, i.e. images of relevant body parts, are generally large (usually several 100 kB) while only a small subset of this data is individually discriminative (a few bytes). Therefore, uniqueness-bearing *features* are extracted in order to decouple the subject-specific data component. The individually discriminative features can then be combined to form a *biometric template* providing a data structure for an efficient handling of biometric information.

[6]Sensors may include cameras operating at different electromagnetic wavelengths, pressure sensors, chromatographs, scales, accelerators, microphones, distance sensors, chemical sensors or even more sophisticated techniques such as DNA sampling or tomographic techniques.

[7]There is no physiological or technical reason that prevents the capture of biometric entities on the coats of penguins or zebra in the IR/UV spectrum for instance in order to enable night observation.

Storage. During an *enrolment phase*, templates are saved in a population database together with a unique identifier, e.g. a name tag or some integer etc. In praxis, the size of the database can grow rapidly to several million entries, even above actual population sizes since profiles might be stored over several generations. In order to keep matching times down, a cascaded database design that emphasises the concepts of rigorous indexation, parallel processing and working sets is usually employed.

Matching. During *submission*, that is the actual runtime phase, the system performs identification by establishing an association between a measurement and entries in the population database. A matching scheme, usually a distance measure, is employed to compare an input with the existing profiles either exhaustively or stepwise by means of indexation.

Application. Finally, the matching results (or the newly formed templates in the case of enrolment) are presented to the application together with metadata on the entity's quality and the anticipated certainty of the produced result.

Summarising the model, the system as a whole is built around a core recognition pipeline of vision algorithms (extraction and matching) that extract and compare a biometric entity presented to a sensor (acquisition device) against stored profiles (database). According to the match calculated, the system finally reacts to the input presented (application).

2.2.5 Performance Characteristics: Measuring the Quality of Operation

Essentially, a visual biometric system constitutes a classification framework that maps from images to identities. Its performance is, therefore, traditionally quantified by characteristics that describe the statistics of a classifier. Assuming authentication, i.e. matching against a single profile in the database, biometric systems are required to categorise an individual as either genuine (accept) or as an impostor (reject).

For a set of test images presented to the system, this binary classification scheme translates into two error measures, that is the *false acceptance rate* (FAR) and the *false rejection rate* (FRR). Complementary, the *genuine acceptance rate* (GAR) is $1-\text{FRR}$ and the *genuine rejection rate* (GRR) is $1-\text{FAR}$. By adjusting the parameters of an identification system, one can typically trade an increase in GAR for an increase in FAR. Hence, introducing a particular point in this trade-off, the *equal error rate* (EER) – defined as the cross-over value where FAR and FRR coincide – is often used to compare the performance of different systems.

However, in this work the entire spectrum of GAR-FAR-pairs, that is the *'receiver operating characteristic'* (ROC), will be used for performance description. This approach provides more complex a measure, accounting for different system behaviours at different GAR-FAR-trade-offs (see [219] for details). Note that one point on the ROC curve is chosen as the *working area*, pinpointing the system to operate at one specific trade-off ratio.

2.2.6 Enrolment Limitations: The Albino Problem

Not all individuals of a population are typically able to register or *'enrol'* with a biometric system. Insufficient quality[8] of biometric entities (lack of universality) as well as inappropriate capturing conditions (lighting, pose, occlusion, motion blur etc.) can cause a *failure of enrolment* (FoE). In the animal kingdom a lack of entity quality is often caused by injuries (predation marks, loss of body parts), genetic conditions (albinos) or natural variance (lack of chest spots in African penguins). For illustration, Figure 2.4 visualises the distribution of spot counts over a penguin subpopulation (note the log scale of the graph).

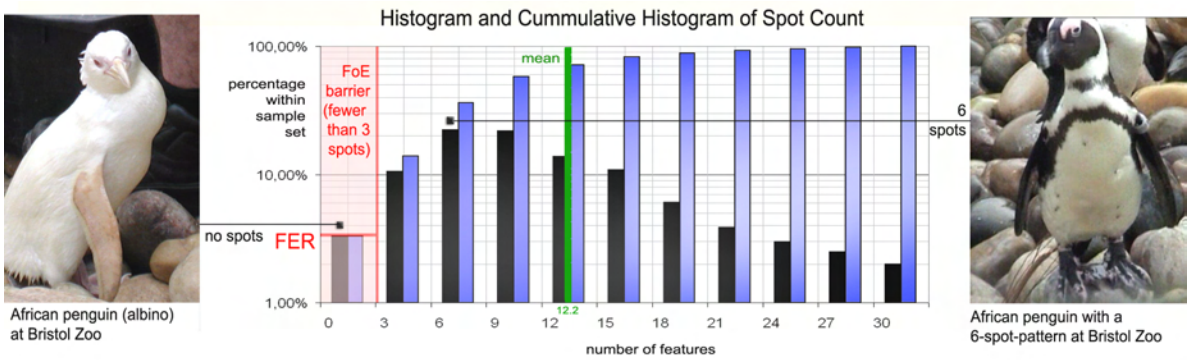


Figure 2.4: ***Histogram of Spot Counts in African Penguins.*** Manual screening of 400 African penguins revealed the above histogram of spot counts. On average, there were 12.2 spots on a penguin's chest, while most penguins carried 7 spots (bin 6-8). It can be seen that the entity is widely universal to the sample population, yet about 3% of animals fall below the defined failure of enrolment (FoE) barrier of a minimum of 3 spots where less than 1% did not carry any chest spots. Apart from regularly patterned penguins with no spots, rare albinos (shown on the left) fall into this category. They miss all dark colouration of feathers, beak and skin. [image sources: I12, I01]

The proportion of the population that fails to complete the enrolment is specified by the *failure to enrol rate* (FER) indicated by red marks in the histogram above. It reflects the combined universality (given by the distribution of spot counts) and measurability (represented by a manually introduced *failure of enrolment* (FoE) barrier which is defined in order to reject entities below a certain quality measure).

[8] For instance, a subset of the human population have fingerprints of poor quality (bricklayers, steel workers) or miss the entity entirely (accidents, disabilities).

2.2.7 Visual Animal Biometrics

Having discussed the fundamental structure and properties of biometric systems, the few applications and (application attempts) that employ biometrics for the identification of animals will be reviewed now. During the last few decades researchers have started to investigate the use of photographic evidence for the collection and biometric recovery of animal identities. Early identification schemes of this sort pre-date modern computer-based techniques and rely on manual feature counts as, for example, work by *Gill* [79] on red-spotted newts or *Loafman* [128] on using salamander scale patterns as biometric entities.

More detailed manual inspection methods involve the creation of a photographic catalogue of individuals taken by a hand-held camera or a sketch collection (e.g. *Doody* [50], *Peterson* [153], *Hagström* [88], *Scott* [184]). Commonly, each time a new sighting is obtained by a human observer the catalogue is searched for a match by eye (e.g. *Friday et al.* [69]; *Gowans and Whitehead* [80]; *Castro and Rosa* [30]).

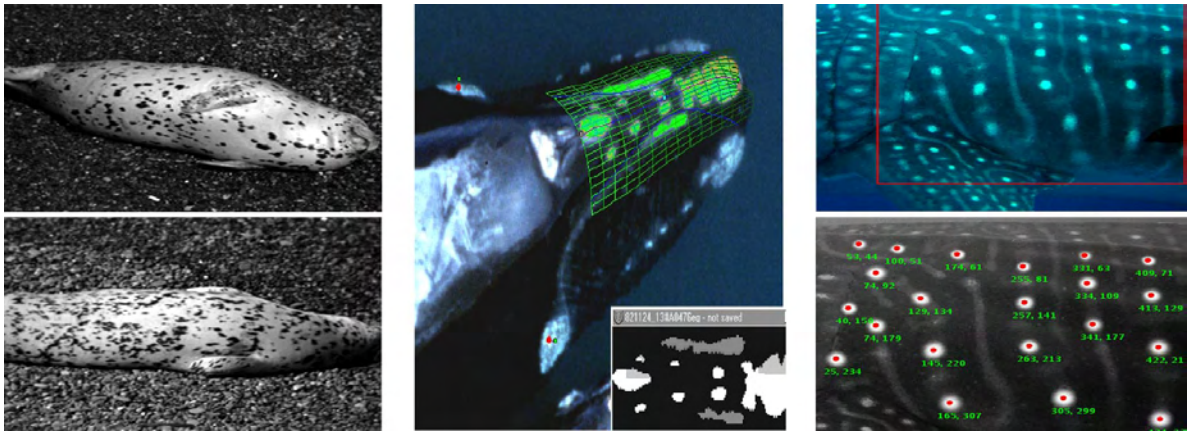
The effort for manually cataloging and identifying large numbers of animals is enormous and can take months or years of work. Despite this fact, less than a dozen computer-aided, that is *partially automated*, prototypical animal identification systems exist as of today. In all cases, the animals are detected and registered[9] by the user where the pose is either manually normalised or corrected based on human inputs. Most of the research conducted is focused on large marine animals, i.e. whales, sharks and seals, since skin deformations are minimal and smooth in nature. Table 2.3 provides an overview of the systems published.

Publication	**Species**	**Feature Types**	**Labelling**	**Matching**
Hiby & Lovell [93], Kelly[106]	seal, cheetah, right whale	dense body texture	2D spline fitting	patch-based distance metric
Vincent et al. [210]	grey seal	dense side pattern	area selection	FG/BG ratio
Ravela & Gamble [166]	salamander	dense top pattern	segmentation	PCA-like model
Tienhoven et al. [208]	shark	sparse side spots	Euclidean	point pairing
Speed et al. [190]	whale shark	sparse side spots	Euclidean	BIC on pair data
Arzoumanian et al. [6]	whale shark	sparse side spots	Euclidean	Groth's method
Ranguelova et al. [164]	humpback	dense fin contour	segmentation	spatial histograms
Forcada et al. [63]	monk seals	fluke silhouette	area selection	wavelet analysis
Foster et al. [66]	plains zebra	vectorised stripes	segmentation	vector set similarity

Table 2.3: **Overview of Computer-aided Vision Systems for Wildlife Biometrics.** (Key: FG/BG...fore-/background, PCA...principal component analysis, BIC...Bayesian Info Criteria [3])

[9] Although a common task in general vision, only a few, rather basic attempts have been made so far to address the species detection and normalisation issue in biometrical applications dealing with animals, e.g. selecting patches of a certain body region on dolphin images by *Araabi et al.* [5].

Dense Texture Comparison. *Hiby and Lovell* [93] suggest fitting a spline over the image by manual control-point adjustment in order to achieve pose normalisation as shown in Figure 2.5(b) for right whales *(Eubalaena glacialis)*. Subsequently, texture patches are compared using a similarity measure that is not disclosed in their publications. The system finally outputs sets of likely matches that are presented to the user for a final, manual screening. *Kelly* [106], for example, utilised the approach for the individual identification of cheetahs *(Acinonyx jubatus)* from scanned photographs.



(a) **(b)** **(c)**

Figure 2.5: ***Examples of Computer-aided Animal Identification Systems.*** Detection, registration and normalisation are provided manually. **(a)** *Vincent et al.* [210] label the body area and calculate spotted/unspotted ratios for a broad classification. **(b)** A system by *Hiby and Lovell* [93] relies on manually fitting a spline (green) for normalisation before performing a simple, texture-based patch comparison. **(c)** After manually normalising a patch of interest and labelling the landmarks, *Arzoumanian et al.* [6] employ Groth's algorithm for achieving a more robust, sparse spot matching.

The Hiby-Lovell system effectively uses the human ability to fit surfaces accurately in order to normalise for pose variations. It is effective and has proven itself in a number of cases. However, the technique has two main drawbacks: 1) intensive manual labelling, and 2) pixel-based, dense comparison which is highly susceptible to lighting changes and image artefacts.

Addressing the latter problem by trading specificity for robustness (essentially turning the approach of dense comparison into a soft biometric technique), *Vincent et al.* [210] suggest using binary information, that is ratios of spotted and non-spotted areas on grey seals *(Halichoerus grypus)* visualised in Figure 2.5(a), broadly to classify populations in order to aid later manual identification. In contrast, *Forcada et al.* [63] approach the lighting issue by focussing on 1D structures only (fluke silhouettes) that can be reconstructed more robustly, exploiting the limitation in dimensionality. The authors apply the method to identify monk seals *(Monachus monachus)* by comparing wavelet coefficients obtained from

the fluke contour. *Ranguelova et al.* [164] propose a similar 1D technique for humpback whales *(Megaptera novaeangliae)* engaging grid-based histograms for matching fluke contours. A reduction to 1D structures during pre-processing (e.g. skeletonisation by *Foster et al.* [66]) without the use of high-level models is only applicable to near perfect imagery. Otherwise, it runs the risk of misinterpreting data before the actual matching procedure.

However, using *sparse, localised features* embedded in the full two dimensions of the animal surface, i.e. local features defined by shape, colour or orientation, can provide a great wealth of information yet, in case of disruptive colouration [47], being prominent enough to avoid weak recognisability. Sparse 2D features will be exploited in this work: similar to minutae detection in fingerprints [102], the singularities of Turing patterns, that is spots, bifurcations and line endings (considered within a species-dependent band of spatial frequencies) will be used as generic primitives for the construction of biometric profiles. Only a few attempts employing rather species-specific, manually aided forms of sparse profiling of animals exist.

Sparse Landmark Comparison. *Van Tienhoven et al.* [208] exploit manually edited imagery of spot-like flank markings to achieve a computer-aided identification of individual raggedtooth sharks (*Carcharias taurus*). Hand-labelled silhouette landmarks (the base of three fins) are used as a reference system for Euclidean normalisation. A cloud of 12-40 hand-labelled spot markings represented within this geometry are compared to a database via an exhaustive search where each point is paired with the closest point in the stored point set and the average Euclidean distance between pairs provides the ranking index. Users are finally presented with the ranked list and make the identification decision.

Technical drawbacks of this largely manual identification system are the ambivalent nature of the biometric entity, i.e. spots gradually blend in with background, and also the simplicity of the matching metric that models the shark as rigid and entities as certain. It does not account for missing or distorted data.

Speed et al. [190] apply the method to whale shark *(Rhincodon typus)* identification, extending the approach by replacing ranking indices with a combination of Akaike's Bayesian Information Criteria [3] and evidence ratios [27]. However, neither the rigid nature of the modelling methodology nor the non-flexible pairing mechanism for features are tackled.

Arzoumanian et al. [6] present a more advanced system employing *Groth's algorithm* [84] as a robust point cloud matcher for whale shark patterns (visualised in Figure 2.5(c)) where,

again, the animal registration is performed manually. The matching technique, developed in the 1980's for the robust identification of astronomic constellations, engages voting based on a binary similarity measure between *all* the triangles that can be constructed over the landmark set. Similarity between two triangles A and B is then confirmed on the basis of a constraint on 1) side-length ratios R between the longest side r_1 and the shortest side r_2 as well as on 2) the cosine C at angle γ between r_1 and r_2, that is:

(GROTH'S SIMILARITY CRITERIA)

$$((R_A - R_B)^2 \ll t_A^2 + t_B^2) \bigwedge ((C_A - C_B)^2 \ll s_A^2 + s_B^2) \tag{2.1}$$

$$\text{where } \; t = 2R^2F, \quad s = 2\sin^2\gamma F + 3cos^2\gamma F^2 \; \text{ and } \; F = \epsilon\left(\tfrac{1}{r_1^2} + \tfrac{C}{r_1 r_2} + \tfrac{1}{r_2^2}\right).$$

The parameter ϵ is a user-defined measure of the anticipated positional uncertainty. Triangles are matched based on the above measure, best matches are used for pairing triangles. Since the number of triangles produced is large – there are $n!/6(n-3)!$ triangles over n points – basic Gaussian statistics are used to find a common magnification factor and a relative orientation between triangle instances. This operation normalises the patterns for rotation and scale. Depending on the distance to the means of the two measures, voting weights are created for each triangle pair, focussing on the matches closest to the most likely scale and rotation. The weighted sum over all votes constitutes a final score which is employed for representing the certainty of identity between the two entities measured.

Despite the manual species identification and landmark labelling, the technique copes well with some missing and distorted data by employing one of the most robust and sophisticated matching scores used in animal biometrics today.

However, none of the systems published combines autonomous animal detection, registration and pose normalisation with sparse, robust matching methods. This is where my work ties in, exploring the possibilities and limitations of fully automated identification.

2.3 Landmark Matching

2.3.1 Rigid Alignment: Matching by Transformation

The previous survey of systems for animal identification has made clear that landmark matching crucially underpins the process of estimating the similarity between sparse biometric entities – they essentially provide distance models in multidimensional pattern spaces. In the following, a brief review of sparse matching schemes is presented to put in context

the histogram-based matching approach proposed in this work:

Generally, a comparison of two 2D landmark sets $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_i \in \mathbb{R}^2, ..., \mathbf{x}_m\}$ and $\mathbf{Y} = \{\mathbf{y}_1, ..., \mathbf{y}_j \in \mathbb{R}^2, ..., \mathbf{y}_n\}$ can be achieved by 1) mapping both patterns into a common reference system (or domain), i.e. performing a *normalising transform*, together with 2) defining a distance measure in this space, i.e. enabling a *ranking* or *clustering* of patterns.

In the most simple case, assuming a *Euclidean similarity transform* between the sets, normalisation schemes have to account for translation, scale and rotation. One standard procedure for this sort of matching is Procrustes analysis which can be traced back to early work by *Kendall* [107, 108] on shape spaces and shape manifolds.

He suggests recovering the translation between the sets $\mathbf{X}$ and $\mathbf{Y}$ by turning the centre of mass $\bar{\mathbf{x}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i$ into the origin of new reference systems, essentially mapping $\mathbf{x}_i \rightarrow \mathbf{x}_i - \bar{\mathbf{x}}$ (the same is applied to the $\mathbf{y}_j$). Subsequently, scale is normalised exploiting the mean sample distance to the origin, that is $\mathbf{x}_i \rightarrow \mathbf{x}_i \left(\sum_{k=1}^{m}(\mathbf{x}_k - \bar{\mathbf{x}})^2\right)^{(1/2)}$. Sorting the points by angle, fixing $\mathbf{X}$ and then minimising the pattern variance by exhaustively rotating $\mathbf{Y}$ around the origin over angles γ, finally yields the best rotational fit between the patterns:

$$\begin{aligned} &\text{(PROCRUSTES FIT)} \\ &\arg\min_{\gamma} \textstyle\sum_{i=1}^{n} |(y_{i,1}\cos\gamma - y_{i,2}\sin\gamma, y_{i,2}\sin\gamma + y_{i,1}\cos\gamma) - \mathbf{x}_i| \end{aligned} \tag{2.2}$$

where $\mathbf{y}_i = (y_{i,1}, y_{i,2})$ is paired with $\mathbf{x}_i$ (based on an angular sort over index γ), and $|\mathbf{x}|$ denotes the Euclidean vector norm.

Other derivatives of the method employ string distances such as the edit-distance [159]. The necessity for the application of an angular sort is, however, intrinsic to the approach. Bearing in mind that misdetections of features as well as distortions may occur during the registration of sparse features, this angular sort is likely to fail in biometric analyses, resulting in a comparison of effectively unrelated landmarks. Furthermore, the generation of the reference system is, in general, not robust to outliers, that is misdetections alter the normalisation terms based on overall measures such as $\bar{\mathbf{x}}$.

In order to counter these shortcomings, *Luo and Hancock* [132] suggest employing the expectation-maximisation (EM) algorithm for a stepwise optimisation. Other advanced schemes for a Euclidean alignment of rigid structures either employ iterations, e.g. the Iterated Closest Point (ICP) technique by *Besl et al.* [23], or define different estimators for recovering transform parameters such as the centroid bounding criteria by *Griffin et al.* [83].

All the approaches mentioned attempt to find a rigid mapping between the patterns $\mathbf{X}$

and $\mathbf{Y}$ where the matching score – the most important result for biometric applications – is generally calculated from the variances that could not be corrected for by the transform discovered. Avoiding such explicit definitions of matching scores, a completely different class of approaches aims at clustering the pose space, engaging parametric voting schemes instead of aligning models in image space.

2.3.2 Parametric Clustering: Matching by Partitioning

Geometric hashing, as proposed by *Wolfson and Rigoutsos* [220], employs the paradigm of *voting* in an especially time-efficient, yet memory-intensive framework. The technique provides a means for matching a pattern $\mathbf{Y}$ against a hash table of models $\mathbf{X}_1, \mathbf{X}_2$ etc. For building the table from models $\mathbf{X}_i$, the technique recruits subsets $\mathbf{V}_i \subseteq \mathbf{X}_i$ of landmarks wherefrom all possible systems of basis vectors are generated. In the case of affine transforms, point triples yield a basis $\mathbf{B}_{\mathbf{x}_i} = \{\mathbf{b}_i\}$. This basis is then used for encoding all remaining points in $\mathbf{X}_i$. Wolfson suggests forming a hash key from these resulting coordinates and saving the model index i as the hash entry to this key.

An unknown pattern $\mathbf{Y}$ can then be compared by assembling a suitably sized set (the larger the set, the more accurate the prediction) of basis vectors $\mathbf{B}_\mathbf{y}$ – each used as a hash key. Building a histogram over all model indices i found in the database according to these keys reveals a winning model as a peak in the histogram. Defining appropriate techniques for robustly identifying peaks can, however, pose a challenge. The technique is applicable strictly in the case of rigid content since deformations will affect the generation of hash keys, causing the voting process to fail. Several extensions, for instance contributions by *Rigoutsos* [167] and *Tsai* [199], try tackling the issue by fuzzifying the key generation.

Other voting-based schemes employ more abstract parameter configurations instead of low-level location data. The generalised Hough transform [9], for instance, is characterised by an accumulation of pose evidences, followed by a clustering step that selects hypotheses with strong support. In biometrics, the scheme has been applied to minutae matching in fingerprints by *Ratha et al.* [165] accounting for Euclidean transforms in the patterns.

The major problem of Hough's technique for matching biometric entities lies in defining a suitable parameter space – i.e. the Hough space – for more applicable, yet more complex transforms that cover all hypotheses in a deformable domain, but having acceptable requirements in terms of storage and computational resources.

2.3.3 Histogramming and Similarity Indices: Matching by Distance

Matching schemes that process patterns element by element such as Procrustes analysis [107] are, as indicated, very sensitive to perturbations in the completeness and ordering of elements in $\mathbf{X}$ or $\mathbf{Y}$. Provided the likelihood of pairings between elements in $\mathbf{X}$ and $\mathbf{Y}$ is known, the similarity problem can be formalised using a quadratic-form distance which employs an assignment matrix $\mathbf{A}$ for ranking the probability of element pairings. Table 2.4 summarises this approach together with other distance measures that tackle the problem of missing and unreliable data.

If no cross-assignment data is available, cross-index dissimilarity measures that operate over *element sets* (such as the Match-distance or the Kolmogorov-Smirnov distance) have still been found to perform, in general, more robustly than element-based techniques [193].

Meta-distances pursue another strategy, maximising hypotheses over all possible combinations of pairing-based distances between different elements of the sets to produce a single, order-independent scalar measure. The Hausdorff distance and the earth mover's distance[10] are powerful examples of this category of distance measures.

Element-based Measures	
Minkowski-form L_r [193]	$d_{L_r}(\mathbf{x},\mathbf{y}) = \left(\sum_i \lvert x_i - y_i\rvert^r\right)^{(1/r)}$
Kullback-Leibler (KL) [116]	$d_{KL}(\mathbf{x},\mathbf{y}) = \sum_i x_i \log \frac{x_i}{y_i}$
χ^2-distance	$d_{\chi^2}(\mathbf{x},\mathbf{y}) = \sum_i \frac{(x_i - m_i)^2}{m_i}$
Jeffrey divergence	$d_J(\mathbf{x},\mathbf{y}) = \sum_i (x_i \log \frac{x_i}{m_i} + y_i \log \frac{y_i}{m_i})$
Cross-index Measures	
Quadratic-form (Mahalanobis)	$d_Q(\mathbf{X},\mathbf{Y}) = \sqrt{(\bar{\mathbf{X}} - \bar{\mathbf{Y}})^T \mathbf{A} (\bar{\mathbf{X}} - \bar{\mathbf{Y}})}$
Match-distance	$d_M(\mathbf{X},\mathbf{Y}) = \sum_i \lvert \sum_{j<i} \mathbf{x}_j - \sum_{j<i} \mathbf{y}_j \rvert$
Kolmogorov-Smirnov (KS)	$d_{KS}(\mathbf{X},\mathbf{Y}) = \max_i \lvert \sum_{j<i} \mathbf{x}_j - \sum_{j<i} \mathbf{y}_j \rvert$
Meta Distance Measures	
Hausdorff distance	$d_H(\mathbf{X},\mathbf{Y}) = \max\{\sup_{\mathbf{X}} \inf_{\mathbf{Y}} d(\mathbf{x},\mathbf{y}), \sup_{\mathbf{Y}} \inf_{\mathbf{X}} d(\mathbf{x},\mathbf{y})\}$
Earth mover's distance (EMD) [172]	$d_{EMD}(\mathbf{X},\mathbf{Y}) = \frac{\sum_{\mathbf{X}} \sum_{\mathbf{Y}} d(\mathbf{x},\mathbf{y}) f(\mathbf{x},\mathbf{y})}{\sum_{\mathbf{Y}} f(\mathbf{x},\mathbf{y})}$

Table 2.4: **Selection of Similarity Measures.** The table shows a selection of both element-based and cross-index distance measures used in point pattern matching where $\bar{\mathbf{X}}$ represents the matrix with all $\mathbf{x} \in \mathbf{X}$ as column-entries, $\mathbf{A}$ is a a-priori known mixture matrix, $d(\mathbf{x},\mathbf{y})$ is some element-based metric, $f(\mathbf{x},\mathbf{y})$ is some minimal flow and $m_i = \frac{x_i + y_i}{2}$ is used as a convenient shorthand.

Shape contexts, originally introduced for character recognition by *Belongie et al.* [20], provide a distance measure based on spatial histograms that incorporate a local coherence assumption. The latter property is highly applicable to coat patterns since underlying

[10] The Earth-Mover's distance is used in this work and discussed in detail later in Chapter 6.

skin dynamics show a correlation between landmark distance and the severeness of their relative distortion. Belongie's technique partitions the scale-normalised image space with respect to a selected reference landmark (the origin) by building a polar histogram $H = (h_1, h_2, ..., h_j, ...)$ over other landmark positions. The histogram shows increasing bin size further away from the reference. Generating an entire set $\{H_i\}$ of such histograms, i.e. the shape context, where each landmark $\mathbf{x}_i$ is used as the reference once, introduces robustness to missing/unreliable data by providing invariance from one specific reference.

Two histogram sets $\{H_i\}$ and $\{K_j\}$ – of possibly different landmark cardinality – are matched based on an assignment matrix $\mathbf{M} = [m_{ij}]$ where, according to Belongie, the entries m_{ij} represent the χ^2-distances between H_i and K_j. The Hungarian method by *Kuhn* [115] is then employed to solve the assignment problem of the matrix $\mathbf{M}$, essentially associating histograms H_i and K_j in an undirected, 0-1-regular bipartite graph[11]. Since each histogram H_i is associated with a reference landmark $\mathbf{x}_i$, the assignment also yields pairings of landmarks $(\mathbf{x}_i, \mathbf{y}_j)$ between the patterns $\mathbf{X}$ and $\mathbf{Y}$. Belongie suggests using the sum of the associated matching costs as a distance metric in pattern space where a penalty term can be assigned to unpaired landmarks.

Its robustness to distortions, the feature of outlier detection and the possibility of comparing point sets of different landmark cardinality render the measure applicable to the task of matching auto-detected landmarks in Turing structures. The technique is of a statistical nature and performs especially well for large landmark sets.

However, as it will be shown in Chapter 6, Belongie's technique can be extended both 1) to deal with small sets of typically 5 to 30 landmarks only (as found in the sample species investigated), and 2) to incorporate a-priori information regarding anticipated pattern distortion. This extended technique is formulated and applied in this work to match coat patterns of African penguins.

2.3.4 Other Landmark-based Approaches: Voting, Warping and Co.

The distance between landmark patterns can also be recovered by voting procedures that accumulate evidence from shape/similarity invariants of triangle sets. The algorithm by *Groth* [84], discussed earlier, as well as the Delaunay-triangulation scheme by *Cox et al.* [39] fall into this algorithm category. Despite the fact that voting enhances the robustness with

[11] A graph is 0-1-regular if and only if *all* vertices have 1 or less edges. A graph is bipartite if and only if there exist two partitions of vertices such that no edge connects vertices of the same partition.

regard to missing landmarks, measures that exploit triangular similarity are generally limited to similar transforms.

If a nominal criteria of the landmark set can be established, patterns may be normalised using image *warping techniques*[12]. For instance, *Senior et al.* [185] use equal spacing between ridge curves in fingerprints to normalise patterns. *Ross et al.* [170] assume a smooth nature of the distortion field and model occurring deformation by employing thin plate splines anchored at ridge curves.

The *SoftAssign* algorithm by *Rangarajan et al.* [163] pursues an iterative optimisation in order to achieve a bipartite matching. Their technique resembles the alterative nature of the expectation-maximisation (EM) algorithm [132], implementing a progressive, sequential interplay between 1) moving landmarks of one pattern closer to their suspected counterpart(s) in the other, and 2) resetting association weights between landmark pairs based on their distance. However, the comparison task is not solved by association alone. As in warping, a mapping score must be defined over the space of transforms for quantising the similarity between patterns.

Stressing the link between measurement and deformation, *distortion avoidance*, as suggested by *Dorai et al.* [51], aims at selecting a minimally deformed signal during acquisition. The strategy relies on both the availability of multiple inputs and a means of rating the severeness of the distortion by a mapping score. Assuming the availability of dense time-series data, flow models [33] have, for instance, been used to determine the extent of deformation. However, multiple inputs rely on some degree of cooperation often denied by wild animals.

[12]The main difficulty of warping lies in finding an appropriate definition of either a 'nominal representation' or 'warping score' that quantifies the severeness of transforms between two patterns given missing data. In most cases, warping techniques require either regular or semantically annotated feature maps. They are, in broad contrast to shape contexts, not easily applicable to sets of landmark data without information on the landmark pairing.

2.3.5 Summarising Note on Matching Techniques

The discovery of a distortion function on the one hand or the assignment of pairings (defining a sparse homography[13]) on the other hand form alternative avenues towards landmark matching. Both processes are related: a sparse pairing *confines* a dense mapping while a dense distortion function exhaustively *defines* the transform.

Thus, a given transform always implies the pairing while a pairing does not imply the distortion function. (However, homogeneity assumptions and physical distortion properties have been used to link sparse and dense mappings.) In conclusion of this section, Figure 2.6 summarises the concepts reviewed and provides a schematic visualisation of the concepts discussed.

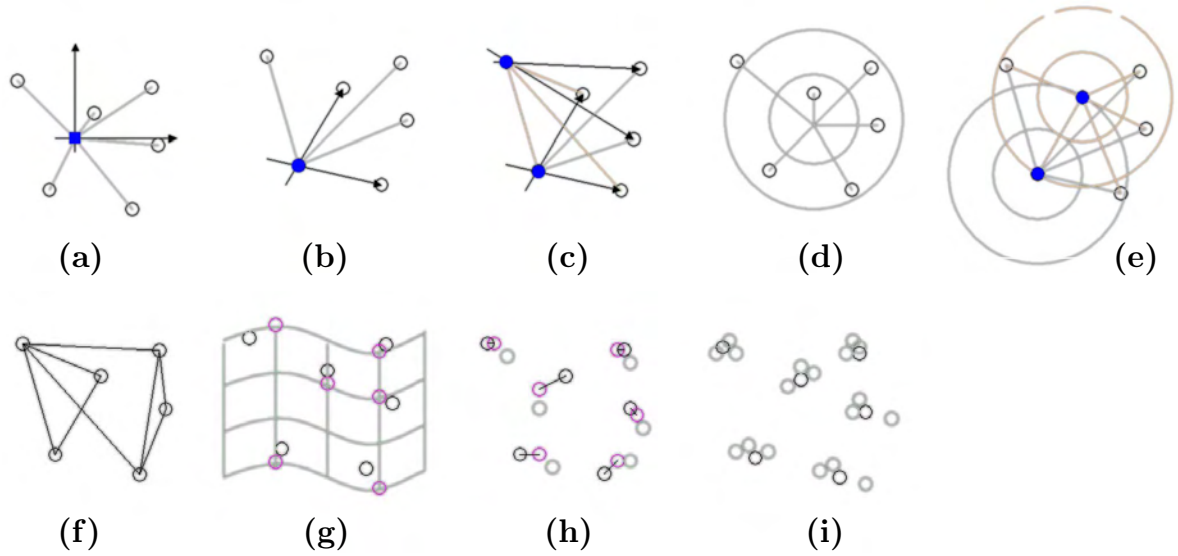


Figure 2.6: ***Selection of Geometrical Concepts for Normalisation and Matching.***

(a) statistically estimated reference system (e.g. Procrustes analysis);
(b) designated landmarks as reference system;
(c) multiple subsets of landmarks as reference systems (e.g. Geometric Hashing);
(d) statistically aligned spatial histogram (e.g. Procrustes analysis + Histogramming);
(e) multiple, landmark centred spatial histograms (e.g. Shape Context);
(f) similarity of triangle sets (e.g. Groth's algorithm);
(g) reference-based interpolation (e.g. Spline Models);
(h) iterative landmark pairing (e.g. SoftAssign);
(i) manual sample selection from multiple measurements (e.g. distortion avoidance);

[13] A homography describes a bijective mapping of point sets between two domains. Thus, it is defined as a relation between two landmark configurations, such that any given landmark in one configuration corresponds to one and only one point in the other, and vice versa.

2.4 Recognition of Complex Objects

2.4.1 Mind the Semantic Gap: From Pixels to Objects

In order to apply the matching procedures discussed, biometric entities, e.g. coat patterns, need to be spatially extracted from cluttered images. Accurate entity segmentation can avoid 1) admitting background components to biometric patterns (inter-class disambiguation), and 2) recruiting features over several individuals (intra-class disambiguation). Therefore, a robust recognition of the species and its body parts of interest is required, overcoming the apparent *'semantic gap'* between the pixels and the meaning grouping them.

Object recognition aims at bridging this gap by *categorising* entire image regions $\mathbf{I}_j$ into one semantic class ω_i (e.g. a body part, clutter etc.) member of a superset $\mathbf{\Omega}$ of classes of interest, which form a semantic space. Defining $\mathbf{\Psi}$ as the set of all existing regions $\mathbf{I}_j$, i.e. an image space, an equivalence relation $\phi \subseteq \mathbf{\Psi} \times \mathbf{\Psi}$ can be used to represent class membership of regions where, in an ideal case, the resulting equivalence classes $[\mathbf{I}_j]_\phi \sim \omega_i \in \mathbf{\Omega}$, i.e. the object categories, form a factor set $\mathbf{\Psi}/\phi$. Note that this rather basic model view explains the semantics $\mathbf{\Omega}$ as an unambiguous, $\mathbf{\Psi}$-complete[14] set of disjunct subsets $[\mathbf{I}_j]_\phi \subseteq \mathbf{\Psi}$.

As for the majority of vision tasks, the mapping ϕ is ambiguous for the identification of species classes ω_i. This is mainly due to hidden data (e.g. self-occlusion) and incomplete class descriptions (e.g. use of sparse population samples for learning ϕ). In addition, the actual cardinality of ϕ is vast. Thus, in practical terms, one retreats to *approximate* ϕ on the basis of information that specifies the relationship between the image space $\mathbf{\Psi}$ and the semantic space $\mathbf{\Omega}$, be that sample data or a-priori knowledge in form of scientific models.

It follows a brief survey and a problem-specific critique of relevant recognition strategies is presented, approaches that all seek (by some form of model) to estimate the relation ϕ.

2.4.2 Templates and Local Descriptors: Defining Prototypes

Template matching [101] represents a class ω_i by a single prototype – the template. Distance metrics – reviewed in the previous section – are employed to produce a matching score between this prototype and a novel observation. Templates have been built from 1) patches of the image function $\mathbf{I}$ for use in convolution/cross-correlation, 2) from principal components of regions of $\mathbf{I}$ such as eigensignatures of faces (eigenfaces) used by *Turk et al.* [203],

[14] The constraint ensures that there is an interpretation available for every possible region: for every $\mathbf{I}_j$ in $\mathbf{\Psi}$ there is an associated label in $\mathbf{\Omega}$ such that every image has a label and $\bigcup_j [\mathbf{I}_j]_\phi \equiv \mathbf{\Psi}$ holds.

3) from shape data, e.g. moments, segment signatures, contours, chain codes, for instance, reviewed by *Basri* [12], 4) from local colour histograms, or 5) from structural descriptors such as kernel-based texture measures, e.g. Gabor-filters [72] or wavelets [94].

As a consequence of the *local regularity* of coat patterns, structural descriptors are of special interest for the task at hand. Haar-like features, that is macro-arrangements of localised Haar-wavelets, will be employed in this work to provide primary features. It will be shown that Haar-like features resemble locally stable elements of Turing patterns and can be used to construct effective and efficient descriptors.

Generally, prototypes are either constructed from a single measurement, known as one-shot learning [57], or they are assembled by combining larger sets of samples. The latter method is somewhat more sophisticated since it captures intra-class variabilities. In any case, template matching is limited to modelling an object class based on a single, prototypical point in pattern space. Clearly, extensions are necessary for representing geometrically complex, class-dependent distributions. This can, for instance, be achieved 1) by modelling relations between multiple templates (structural approach), 2) by discovering domain-specific embeddings that simplify the class distributions, e.g. kernel transforms, or 3) by tailoring the feature set that spans the pattern space (statistical approach).

Current realtime systems often exploit the stability (and simplicity!) of the *local object structure* by building spatially confined templates over *local* neighbourhoods. *Wide baseline matching*, e.g. k-nearest neighbour discovery in a large pattern space, is commonly utilised to associate measurements to known prototypes stored in either a database, a hash-table [220] or a kd-tree [19]. This approach is, for instance, used by *Lowe* [130] in his framework of scale-invariant feature transforms (SIFT), by *Rothganger et al.* [171] in a similar, also affine-invariant patch detector, and by *Lepetit et al.* [123] in their technique of randomised trees.

SIFT identifies extrema in scale-space as interest points [180] and uses the local distribution of affine-corrected orientations as a characteristic data vector. In contrast to class descriptors, SIFT is, similar to image jets [65] and local cross-correlation measures [161], constructed as a *local image measure* that represents data from a single one-shot-measurement only – in this sense SIFT constitutes a true prototype. Both core concepts of SIFT, that is the use of scale-space extrema and orientation descriptions, require the underlying image to remain structurally constant, e.g. no introduction/removal of image elements, since the

values used are not weighted in terms of their class-dependent relevance.

In coat patterns, only a few, often complex feature combinations prove truly species-specific. To capture these combinations, features are required to be selected, represented and combined. Thus, techniques are needed that account for visual trends in the population of a species, focussing on characteristic pattern properties and property combinations.

2.4.3 Statistical Object Recognition: Defining Decision Boundaries

The goal of the *statistical recognition approach* [101] is to isolate/combine those image features that allow pattern vectors belonging to different object categories (the ω_i's) to occupy disjoint, yet compact regions in pattern space. Establishing decision boundaries between the class-representative clusters then allows for a predictive classification on novel data. Adopting a Bayesian viewpoint, the risk $\mathcal{R}$ of mismatching a class ω_i given an image $\mathbf{I}$ is:

$$\begin{aligned}&\text{(CLASS-CONDITIONAL RISK)}\\&\mathcal{R}(\omega_i|\mathbf{I}) = \textstyle\sum_j L(\omega_i,\omega_j)\mathcal{P}(\omega_j|\mathbf{I})\end{aligned} \tag{2.3}$$

where $\mathcal{P}(\omega_j|\mathbf{I})$ is the probability of assigning the class ω_j given $\mathbf{I}$ and $L(\omega_i,\omega_j)$ is the loss incurred by deciding that $\mathbf{I}$ is associated to class ω_i when the true class is ω_j. Minimising this loss function brings forth the decision boundaries desired. For instance, given the identity relation as the loss function, risk minimisation yields the *maximum a-posteriori* (MAP) decision rule [3] assigning $\mathbf{I}$ to ω_i if and only if $\forall_{j\neq i} : \mathcal{P}(\omega_i|\mathbf{I}) > \mathcal{P}(\omega_j|\mathbf{I})$.

In species detection – as in most practical cases – posterior densities $\mathcal{P}(\omega_i|\mathbf{I})$ are unknown and have to be estimated based on class-specific likelihoods $\mathcal{P}(\{c_k(\mathbf{I})\}|\omega_i)$ of image features $c_k(\mathbf{I})$. The features are commonly approximated to be statistically independent from one another. The posterior can then be formulated as a Bayesian inference [17]:

$$\begin{aligned}&\text{(POSTERIOR CONSTRUCTION FROM INDEPENDENT EVIDENCE)}\\&\mathcal{P}(\omega_i|\mathbf{I}) = \tfrac{\mathcal{P}(\{c_k(\mathbf{I})\}|\omega_i)\mathcal{P}(\omega_i)}{\mathcal{P}(\{c_k(\mathbf{I})\})} \approx \mathcal{P}(\omega_i)\textstyle\prod_k \tfrac{\mathcal{P}(c_k(\mathbf{I})|\omega_i)}{\mathcal{P}(c_k(\mathbf{I}))}\end{aligned} \tag{2.4}$$

essentially combining the class-dependent feature likelihoods $\mathcal{P}(c_k(\mathbf{I})|\omega_i)$ as independent events. For the problem at hand in particular, both the form of the posterior and the likelihoods are initially unknown: how are features distributed over the population and under deformation? Moreover, considering the ordered structure of Turing patterns, local features are anticipated to be statistically *dependent*. Thus, in order to avoid any initial assumptions, a non-parametric learning method is used to discover species-specific patterns.

2.4.4 In-depth Focus: The Extended Viola-Jones Detector

In 2001, *Viola and Jones* [212] suggested first to learn (from large sets of annotated sample imagery) and then scale-invariantly to detect (for the first time in realtime!) members of a complex object class. This thesis utilises and extends their work for species identification. The following section provides a review of this statistical recognition framework.

The basic idea of the approach is to model a class of visual objects by characteristic sets of Haar-like features arranged in boosted, linear combinations of classification and regression trees (CART's) [219]. A cascade of such decision trees is employed for describing the object-specific subspace within the pattern domain [126]. Figure 2.7 puts this approach into its context within a taxonomy of other statistical recognition strategies.

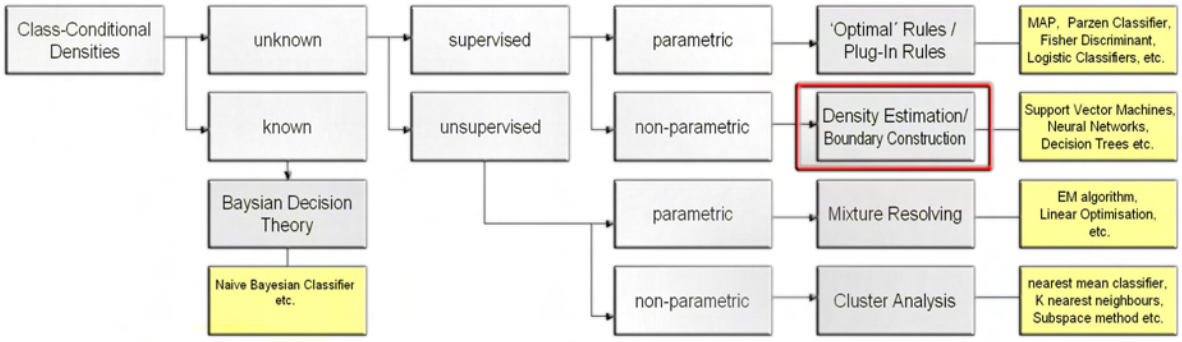


Figure 2.7: ***Taxonomy of Statistical Recognition Strategies.*** A selection of different approaches to statistical pattern recognition (grey boxes) are categorised based on their properties (white boxes). A set of examples (yellow boxes) is associated to each of the concepts. The red mark indicates the position of methods used in this work. (taxonomy inspired by *Jain* [101])

The Viola-Jones detector performs classification based on partitioning the pattern space spanned by so-called Haar-like features. Each Haar-like feature Ψ_K is defined (and implemented) as an accumulation of weighted block features $\mathbf{S_k}$:

$$\begin{aligned} &\text{(HAAR-LIKE FEATURE)} \\ &\Psi_K(\mathbf{x}) = \textstyle\sum_{\mathbf{k}\in K} w_{\mathbf{k}} \mathbf{S_k}(\mathbf{x}) \end{aligned} \tag{2.5}$$

where $K = (\mathbf{k}_1, \mathbf{k}_2, ..., \mathbf{k}_p)$ is a list containing a number of p type-vectors $\mathbf{k}$ each of which defines the shape[15] of one of the block images used. Essentially, Haar-like features encode block-like contrast arrangements at selective spatial frequencies and positions (see

[15] 2D block features $\mathbf{S_k} = \mathbf{S}_{\mathbf{s},\mathbf{t},b}$ are also referred to as car-box images [212] or as binary 'boxlets' [188]. They can be defined as a scaled (by vector $\mathbf{s}$), shifted (by vector $\mathbf{t}$) and sign altered (by $b \in \{1, -1\}$) versions of the unit square block $\mathbf{S}^{\square}$ of value 1, that is $\mathbf{S}_{(\mathbf{s},\mathbf{t},b)}(\mathbf{x}) = \mathbf{S_k}(\mathbf{x}) = b\mathbf{S}^{\square}\left(\mathbf{x}\cdot\mathbf{s}^T - \mathbf{t}\right)$. Each block image then bears a single, rectangular block of constant entries of $\mathbf{1}$ or $-\mathbf{1}$ on a background of $\mathbf{0}$. Weights are calculated as $w_{\mathbf{k}} = w_{(\mathbf{s},\mathbf{t},b)} = w_{((s_1,..,s_j)^T,\mathbf{t},b)} = \frac{1}{\prod_{i=1}^{i=j}\mathbf{s}_i}$.

Figure 2.8). They resemble both Haar functions as well as elementary building blocks of Turing patterns. This resemblance in pattern structure provides, as it will be demonstrated later, a means for an efficient coding of regions of a species' camouflage pattern.

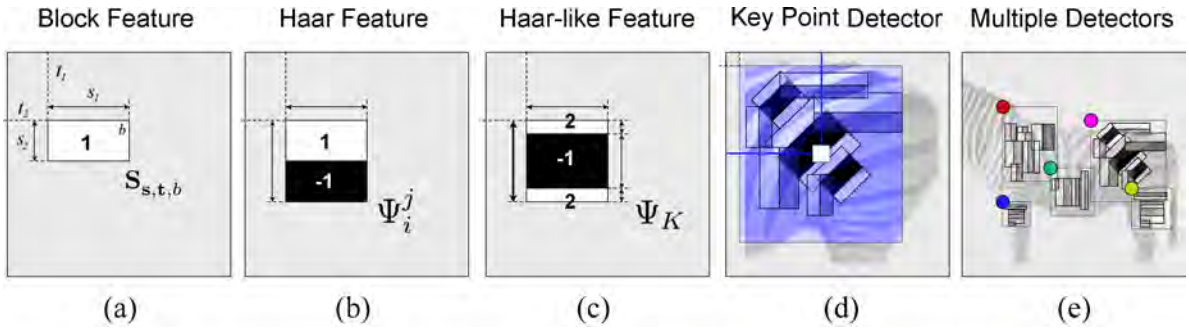


Figure 2.8: ***Structure of Haar-like Features.*** Basic block-features as given in **(a)** are combined to form Haar-wavelets in **(b)** and more complex macro-arrangements, that is Haar-like features in **(c)** where numbers show the weights w associated to block regions. Illustration **(d)** shows several of these features arranged in a template describing the structure of a neighbourhood window characterising the distinctive component of a zebra pattern. **(e)** Multiple descriptors model an animal.

Given an image $\mathbf{I}$, the likelihood of the presence of a single feature kernel Ψ_K can be measured by convolution of the kernel with the image yielding a coefficient c_K:

$$\begin{aligned} &\text{(FEATURE RESPONSE)} \\ &c_K(\mathbf{I}) = \Psi_K * \mathbf{I} = \textstyle\sum_{\mathbf{k}\in K} w_{\mathbf{k}}(\mathbf{S}_{\mathbf{k}} * \mathbf{I}) \end{aligned} \tag{2.6}$$

where $*$ constitutes convolution. The approach is sufficient since, fixing the image location of application, convolutions can be interpreted as dot product operations. Therefore, as elaborated in *Forsyth and Ponce* [65], the response c_K is largest if vector $\mathbf{I}$ and vector Ψ_K are close to parallel, that is they resemble each other.

Viola and Jones suggest significantly enhancing the speed of the convolution operation – a major precondition for realtime operation – by exploiting the general integration rule[16] for linear convolution described by *Heckbert* [91] as:

$$\begin{aligned} &\text{(INTEGRATION RULE OF CONVOLUTION)} \\ &(\mathbf{S}_{\mathbf{k}} * \mathbf{I})^{[n]} = \mathbf{S}_{\mathbf{k}}^{[q]} * \mathbf{I}^{[p]} \quad \text{given} \quad n = p + q \end{aligned} \tag{2.7}$$

where $\mathbf{I}^{[n]}$ represents an n-times image integration. Setting $p = 2$ and $q = -2$ causes *any* block image $\mathbf{S}_{\mathbf{k}}$ to collapse yielding a second derivative $\mathbf{S}''_{\mathbf{k}}$ with only four non-zero entries

[16] The necessary condition for the rule to hold is that both functions provide finite support. By assuming all values outside the image are zero, this is trivially true for discrete images of limited size.

as illustrated in Figure 2.9(b-d). Independent of the image size, an addition/subtraction of the four non-zero entries in the integral image $\mathbf{II}$ then yields the convolution of $\mathbf{S_k}$ with $\mathbf{I}$:

(FAST BLOCK IMAGE CONVOLUTION)

$$\begin{aligned}\mathbf{I} * \mathbf{S_k} = \mathbf{II}(\mathbf{t}_1 - 1, \mathbf{t}_2 - 1) + \mathbf{II}(\mathbf{s}_1 + \mathbf{t}_1 - 1, \mathbf{s}_2 + \mathbf{t}_2 - 1) \\ - \mathbf{II}(\mathbf{s}_1 + \mathbf{t}_1 - 1, \mathbf{t}_2 - 1) - \mathbf{II}(\mathbf{t}_1 - 1, \mathbf{s}_2 + \mathbf{t}_2 - 1)\end{aligned} \tag{2.8}$$

where $\mathbf{k} = ((\mathbf{s}_1, \mathbf{s}_2), (\mathbf{t}_1, \mathbf{t}_2), b)$ holds the scale and translation parameters of the block image. Figure 2.9(g-h) visualises the calculation where colours used in the equation refer to the ones in the illustrations.

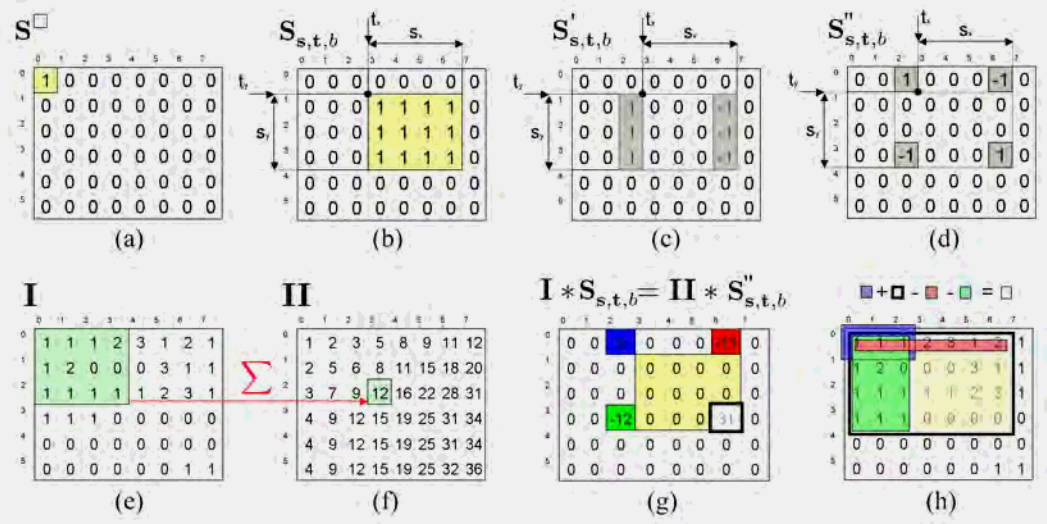


Figure 2.9: ***Efficient Convolution using Block Images.*** **(a)** the unit square; **(b)** a block-image (yellow) is a shifted and scaled (possibly sign-altered) version of the unit square; **(c)** application of horizontal differentiation leads to two 1D structures; **(d)** further application of vertical differentiation yields four impulses; **(e)** an arbitrary image; **(f)** integral image as block sum over all top-left pixels; **(g)** convolution of block-image with arbitrary image equals convolution of integral image with four impulses; **(h)** schematic visualisation of the summation process on the arbitrary image;

The integral image $\mathbf{II}$ can be calculated rapidly by an iterative scheme of linear complexity:

(IMAGE INTEGRATION)

$$\begin{aligned}\mathbf{II}(-1, y) &= 0; \qquad & \mathbf{II}(x, y) &= \mathbf{II}(x - 1, y) + A(x, y); \\ A(x, -1) &= 0; \qquad & A(x, y) &= A(x, y - 1) + \mathbf{I}(x, y).\end{aligned} \tag{2.9}$$

It will be shown later that this speed-up proves especially efficient for the high-resolution imagery necessary to capture sufficient details for individual coat pattern identification.

In contrast to Haar-wavelets, Haar-like features form an over-complete descriptor set. I

argue, over an area of $x \times y$ pixels, there exist a binomial number of non-redundant[17] features:

(CARDINALITY OF COMPLETE HAAR-LIKE FEATURE SET)

$$|\{\Psi_K\}| = \begin{pmatrix} l \\ m \end{pmatrix} \quad \text{where} \quad l = |\{K\}| = 2\sum_{i=1}^{x}\sum_{j=1}^{y}(x-i+1)(y-j+1). \tag{2.10}$$

The parameter l reflects the number of block images $\mathbf{S_k}$ in the area and m is the number of block images recruited to form features, that is the cardinality of lists K. Figure 2.10 illustrates the exponential growth of the Haar-like feature pool plotted against image size.

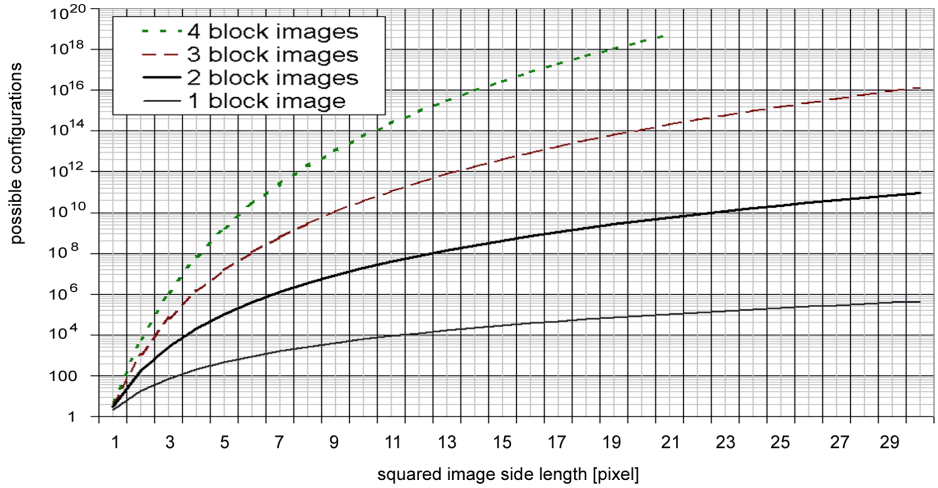


Figure 2.10: **Cardinality of Unpruned Haar-like Feature Pool.** The graph illustrates the number of Haar-like features that can be composed out of m block images, for $m = 1, ..., 4$, constructed over differently sized square images according to Eq. (2.10) given $x = y$. Since each feature constitutes a dimension, the spanned pattern space is accordingly vast. Hence, a dimensionality reduction or a limitation of the feature space to some reasonable subset is crucial to enable practical use.

Clearly, pruning the feature pool in a domain-motivated way is an imperative for balancing the system's generality against its parsimony [127]. Table 2.5 compares the selection of feature classes used in this work to pools engaged in other systems/publications.

Note that spot-like, line-like and edge-like features are employed since these feature types are found in Turing patterns. Non-adjacent, widely spread compositions of block images, as suggested by *Zhang et al.* [223], are of limited relevance since, as it will be shown later, the variability in coat patterns increases significantly with the distance between related features. In order to estimate the feature-dependent distribution $\mathcal{P}(\omega|c_1(\mathbf{I}), .., c_i(\mathbf{I}), ...)$ of object likelihoods in the image, Viola and Jones suggest learning the association between coefficients c_i and semantic classes ω from sets of training images $\mathbf{I}^{pos}$ and $\mathbf{I}^{neg}$ where y_j is used as the characteristic function for images $\mathbf{I}_j$ that is 1 for positives and -1 for negatives.

[17] Features that use multiple block images over the same region $(\mathbf{s}, \mathbf{t})$ and sign b are considered redundant.

	Edges				Diagonals		Lines								Spots		Other
							width 3				width 4						
Haar Transform	X	X			X												
Oren et al.	X	X			X												
Mohan et al.	X	X			X												
Bahlmann et al.	X	X			X												
Viola & Jones	X	X			X		X	X									
Shakhnarovich et al.	X	X			X		X	X									
Lienhart & Maydt	X	X	X	X			X	X	X	X	X	X	X	X	X	X	
Kölsch & Turk	X	X					X	X	X	X							X
Menezes et al.	X	X	X	X			X	X	X	X	X	X	X	X	X	X	
Lie, Blake et al.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Barczak et al.	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
this work	X	X	X	X			X	X	X	X	X	X	X	X	X		

Table 2.5: **Pruned Feature Pools.** The table summarises different layouts of Haar-like classifier pools as they were applied to different detection problems. The last row shows the feature pool used in this work. [The data shown refers to a selection of publications detailed in the references: *Haar* [87], *Oren et al.* [149], *Mohan et al.* [142], *Bahlmann et al.* [8], *Viola and Jones* [212], *Lienhart et al.* [126], *Kölsch and Turk* [111], *Menezes et al.* [138], *Barczak et al.* [10]]

An elementary classifier h_i – representing a single feature c_i – is then built by exhaustive search for the best descriptor over features c_i and parameters (p, q) in form of a threshold classifier:

(ELEMENTARY THRESHOLD CLASSIFIER)

$$h_i(\mathbf{I}_j) = \begin{cases} 1 & pc_i(\mathbf{I}_j) < pq \\ -1 & \text{elsewise} \end{cases} \tag{2.11}$$

where $q \in \mathbb{R}$ constitutes the classification threshold and $p \in \{1, -1\}$ is a binary sign variable. To account for dependencies between Haar-like features – as frequently occurring in Turing patterns – these single-feature classifiers are combined in binary classification and regression trees (CART's[18]) [219]. Figure 2.11 visualises the concept on two trees trained to describe the appearance of frontally captured African penguins.

The leafs of a CART are assigned with values $\mathbf{h}$, that is the difference of the measured posterior densities of samples with respect to arrangements of the n threshold classifiers:

(FEATURE-DEPENDENT CLASS DIFFERENCE)

$$\mathbf{h}(\mathbf{I}_j) = \mathcal{P}(y_j = 1 | h_1(\mathbf{I}_j), .., h_i(\mathbf{I}_j), .., h_n(\mathbf{I}_j)) - \mathcal{P}(y_j = -1 | h_1(\mathbf{I}_j), .., h_i(\mathbf{I}_j), .., h_n(\mathbf{I}_j)) \tag{2.12}$$

Since small CART's (of limited depth) are used, the classifiers incorporate limited data on only a few features. Intuitively, they constitute 'rules of thumb' that provide quick, but relatively weak disambiguation. Therefore, Viola and Jones propose combining multiple learners $\mathbf{h}_t$ into one superior classifier H, using a framework first suggested by *Kearns and*

[18] *Lienhart et al.* [126] originally used boosted CART's in conjunction with Haar-like features for capturing the class-specific component of human faces.

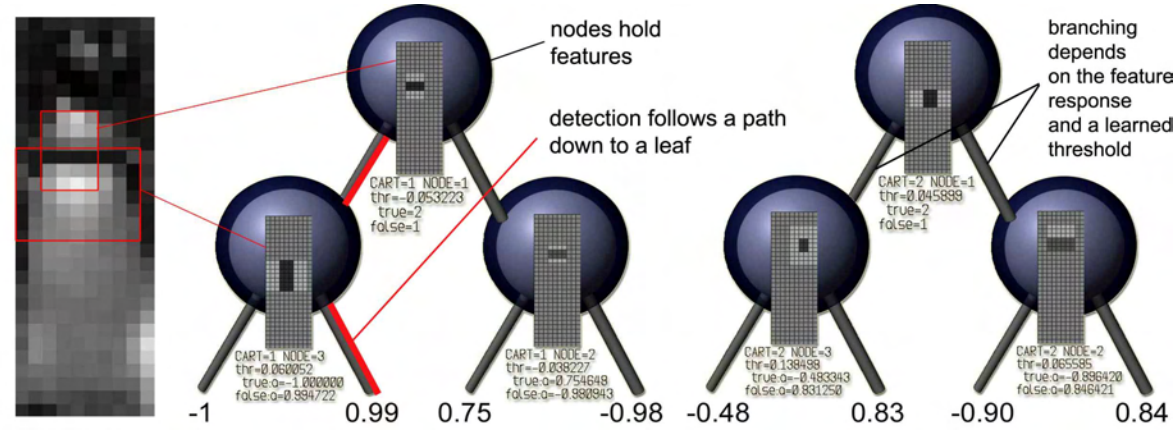


Figure 2.11: **Classification and Regression Trees.** For illustration, two CART classifiers are depicted which are found most effective in disambiguating frontal African penguins from habitat clutter. The leaves of the tree contain the values $\mathbf{h} \in [-1, 1]$ which reflect a tree's possible classification verdicts. CART's are non-linear classifiers. Note that different, XOR-related forms of the chest stripe are encoded (left tree, root and right leaf) covering different manifestations of feature width.

Vazirani [105] in 1988 and later formalised by *Schapire* [178] and *Freund* [67] in a concept termed *boosting*. AdaBoost [68], the specific form employed, iteratively seeks linearly to combine tree outputs $\mathbf{h}_t$ into a single, precise rule H for predicting the class membership:

(STRONG, BOOSTED CLASSIFIER)

$$H(\mathbf{I}_j) = \begin{cases} 1 & \sum_t \mathbf{h}_t(\mathbf{I}_j) \geq \beta \\ 0 & \text{elsewise.} \end{cases} \tag{2.13}$$

where β is a data-dependent bias [126]. A 'gentle' version of AdaBoost which constructs this classifier is outlined in Algorithm 2.1. Note the iterative nature of the classifier construction: the algorithm pursues a stepwise assembly of H by progressively selecting well performing weak classifiers $\mathbf{h}_t$ that *complement each other* [70]. The procedure has been shown to minimise an exponential error functional Err:

(ADABOOST ERROR FUNCTIONAL)

$$Err = \sum_j e^{-y_j H(\mathbf{I}_j)} \tag{2.14}$$

An account for the correlation between features is especially relevant in highly ordered structures such as Turing patterns. The algorithm implements the strategy using a weight w_j for each sample $\mathbf{I}_j$ that reflects previous classification success. After each round of selecting a classifier $\mathbf{h}_t$, all weights w_j are decreased where the chosen weak rule $\mathbf{h}_t$ classifies the sample $\mathbf{I}_j$ correctly. While this strategy focuses on learning 'difficult' samples and eliminates overcounting due to dependencies, it is relatively sensitive to false labellings.

input: cascade depth M, training set $\{(\mathbf{I}_1,\mathrm{y}_1),...,(\mathrm{I}_j,\mathrm{y}_j),...,(\mathrm{I}_n,\mathrm{y}_n)\}$ where $\mathbf{I}_j \in \mathbb{R}^n$, $\mathrm{y}_j \in \{-1,1\}$
output: classifier H
variables: weights $w_j \in \mathbb{R}$

---

step 1. *initialise weights*
$w_j = 1/n$

step 2. *construct weak classifiers* $\mathbf{h}_t$
repeat for $t = 1,...,M$
  fit CART tree $\mathbf{h}_t$ via weighted least squares, minimising $\sum_j w_j(\mathbf{h}_t(\mathbf{I}_j) - y_j)^2$
  adjust weights: $w_j = w_j e^{\mathbf{h}_t(-y_j)}$
  normalise weights: $w_j = \frac{w_j}{\sum_j w_j}$

step 3. *assemble strong classifier*
$H = \begin{cases} 1 & \sum_t \mathbf{h}_t \geq \beta \\ 0 & \text{elsewise} \end{cases}$

Algorithm 2.1: **Gentle AdaBoost.** Pseudo-code of the algorithm used where β is calculated according to work by *Lienhart et al.* [126].

Amongst other sub-algorithms listed in Table 2.6, Gentle AdaBoost [70] is chosen for the task at hand to balance this sensitivity by bounding the increase of sample weights per learning step. It will be shown later that, for learning coat patterns, Gentle Adaboost outperforms binary valued and logistically constructed derivatives of AdaBoost.

Sub-algorithm	$\mathbf{h}_t$	**Special Feature**
AdaBoost.M1 [178]	binary	greedy combination of weak rules, originally formulated over sets
Discrete AdaBoost [67]	binary	standard approach using binary weak classifiers
AsymBoost [211]	binary	asymmetrical weighting function applied to samples
LogitBoost [67]	binary	logistical weighting function applied to samples
Real AdaBoost [179]	real	introduction of real-valued weak classifiers
Gentle AdaBoost [70]	real	linear weighting function balances outlier importance
FloatBoost [223]	real	rejection of ineffective learning steps
Modest AdaBoost [209]	real	favour generalisation over training success

Table 2.6: ***Derivatives of AdaBoost.*** The table lists commonly used derivatives of the AdaBoost algorithm together with their characteristic feature.

In order to avoid an expensive evaluation of all weak classifiers $\mathbf{h}_j$ at once, Viola and Jones finally suggest the decision process be arranged in an '*attentional cascade*', that is a sequence $H_1, H_2, ..., H_i, ..., H_n$ of boosted classifiers where each stage consists of a small number of CART classifiers $\mathbf{h}_j$ as schematically visualised in Figure 2.12(a).

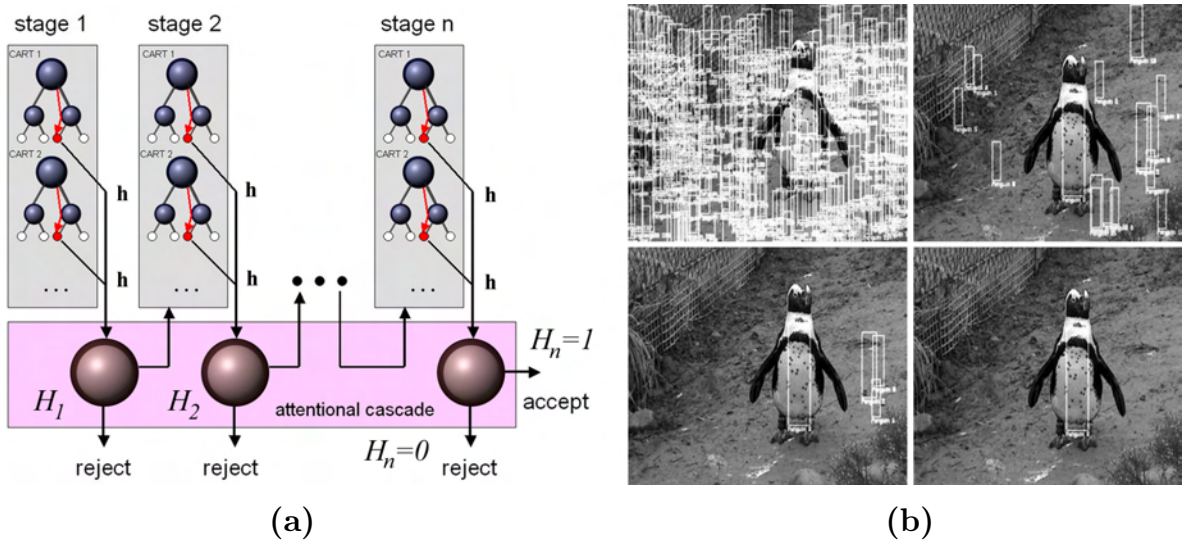


Figure 2.12: ***Structure and Speed-up Effect of Attentional Cascades.*** Tree classifiers are bundled into stages and evaluated in a sequence where an early rejection of detection windows allows fast evaluation with only a minor increase in false rejections. **(a)** schematic overview of an attentional cascade; **(b)** visualisation of the significance of early pruning. Rectangles indicate the areas that pass the entire cascade after 1, 3, 5 and 9 stages of the cascade.

For each location $\mathbf{x}$ and scale s, the overall decision $H_\omega(\mathbf{x}, s)$ with regard to an object class ω is progressively evaluated, passing the sample from stage to stage. The strategy focusses computational resources on the most promising image regions and, by trading some false negatives for speed, it allows for a very fast decision procedure.

Clearly, for the task at hand the strength of the technique lies in combining this capability of fast decision (during online operation) with a greedy, yet efficient search in a very large feature space (during offline training). The extensive search during training shifts further computational effort away from the actual detection, allowing for realtime processing.

In the literature, AdaBoost's impressive generalisation performance on novel data has been confirmed experimentally in a multitude of domains [8, 111, 211, 212, 213]. In this thesis, it will be shown by experiment that AdaBoost can be used to generalise successfully over large animal populations, accounting for natural variation and a substantial extent of deformation and view point change.

Previously, the technique has been applied to detecting various other object classes in artificial environments. Successful applications include systems to detect human faces [211, 212], hands [111], pedestrians [213] and street signs [8].

However, the Viola-Jones framework in the reviewed form exhibits a number of significant drawbacks for the application at hand, including poor localisation in space, vast

labelling efforts and high sensitivity to wide-ranging deformations, be that a substantial change in view aspect (e.g. frontal to profile) or intra-object deformation (e.g. change of body proportions). In Chapter 3 and Chapter 4, approaches will be discussed on how to overcome some of these challenges. They will include ways to customise and extend the framework for extracting point locations (instead of patches) on animal coats, aiming for better localisation and performance. Multiple points will then be used to reconstruct approximately the object area of interest, overcoming the limitation of mere 'bounding box' detections. A model for capturing the shape of such multi-component configurations relies on methods of structural representation which will be reviewed in the section following.

2.4.5 Structural Object Recognition: Defining Spatial Relations

So far, this review has focussed on modelling single entities. However, class-dependent information is also locked in the configuration of an entity's sub- and superstructure. *Structural recognition* adopts this hierarchical perspective where an object class, e.g. an animal species, is viewed as being composed of components, e.g. body parts or regions, which are themselves built from yet simpler subelements [101].

Several suggestions have been made towards the modelling of relationships between different levels of a structural hierarchy:

2D Graphs. Partially or fully connected graphs are traditionally used to model relationships (the graph's edges) between components (the graph's vertices). The concept of 'pictorial structures', published by *Fischler and Elschlager* [62] in the 1970's, is widely considered the origin of graph-based object representation. The technique employs graphs to represent rigid *connectivity properties* between object-components in skeleton-like representations. Recent work by *Felsenszwalb and Huttenlocher* [60] has led to a renaissance of the topic realising its potential as a formal, statistical framework for fitting graphs with vertices $v_i \in V$ to image locations $l_i \in L$ coinciding with corresponding semantics. Pictorial structures L^* are fit by minimising an *a-posteriori function* in pose space:

$$\begin{aligned} &\text{(FITTING PICTORIAL STRUCTURES)} \\ &L^* = \arg\min_{l_i \in L} \left(\sum_{l_i \in L} m_i(l_i) + \sum_{(v_j, v_k) \in V} d_{jk}(l_j, l_k) \right) \end{aligned} \tag{2.15}$$

where m_i estimates the likelihood-based cost of locating feature v_i at location l_i, and $d_{jk}(l_j, l_k)$ represents some deformation cost for pairs of vertices.

Duc et al. [52] and *Smeraldi and Bigun* [189] apply a similar concept of graph representation to identifying facial landmarks, modelling their spatial arrangement in a graph-based architecture where likelihoods of primitives are represented by Gabor coefficients [72]. Focussing on architectural aspects, *Lades et al.* [118] describe a similar, also graph-based system for relating structural components.

The topology and rigid geometrical arrangement of object components can often be represented by a (usually star-shaped) skeleton. In this case, dynamic programming [4] provides an very efficient way of evaluating Eq. (2.15). However, describing the specifics of complex deformations – as occurring in organic surfaces where vertex positions are locked in specific, often elastic interdependency – requires additional model structures.

Active Appearance Models. The revolutionising concept of 'active appearance models', introduced by *Cootes et al.* [36] in 1998, attempts solving the task by iteratively fitting a flexible, landmark-based shape generated from sample sets as a PCA-based description of the variances from a mean shape. A major drawback of the technique is, however, the need for an appropriate shape initialisation.

Tesselations. Recent works by *Felsenszwalb* [58, 59] suggest triangulating the vertex network of graphs to achieve a regularisation. Minimising the deformation cost over the simplices (triangles) of the tesselation is then used for fitting the graph model to the observation data by structured search in pose space.

Splines. Modelling *organic deformations* demands, however, more domain-specific constraints. Given a sparse pairing ($\mathbf{x}_i = [x_i, y_i]^T \in \mathbb{R}^2, \mathbf{v}_i \in \mathbb{R}^2$) between model points and observations, *splines* provide a framework for modelling deformations[19] by representing a *physically motivated*, dense transform that maps the given pairs $(\mathbf{x}_i, \mathbf{v}_i)$ accurately onto each other. Minimising the bending energy $\iint_{\mathbb{R}^2} (\frac{\partial^2 f}{\partial x^2} + \frac{2\partial^2 f}{\partial x \partial y} + \frac{\partial^2 f}{\partial y^2})\, \partial x \partial y$ of a surface, for instance, yields the model of thin plate splines [49] that simulates elastically bent metal. The quality of fit for a pairing $(\mathbf{x}_i, \mathbf{v}_i)$ is simply given by the metallic bending energy of the optimally aligned spline surface found. In the biometric domain *Chen et al.* [32] use thin plate splines for describing pressure-induced fingerprint deformations employing stored and

[19] The very term 'spline' is derived from a long, narrow and thin strip of wood originally used for vessel building in shipyards. In turn, the mathematical entities model 'bending behaviour' under forced deformation.

measured minutae as reference pairs $(\mathbf{x}_i, \mathbf{v}_i)$. While surface properties can be explained well by splines, self-occlusions and perspective deformations demand a true 3D representation.

3D Models. 3D object models can be used to predict perspective transforms as well as singularities such as self-occlusions even for cases where the common assumption of homography fails. The use of rigid 3D models, reviewed in detail by *Basri* [12] and *Basri and Moses* [13], in conjunction with a camera model is common practice for modelling the changing relation of object landmarks with regard to pose. However, some work has also looked into approximating changing/deformable objects by rigid geometries.

Everingham and Zisserman [56] suggest using projections of a rigid, textured 3D hyper-ellipsoid to generate 2D views of heads for the training of pose-indexed trees of patch classifiers. An strategically similar approach is taken by *Ozuysal et al.* [150] who model objects by approximating the geometry of features also as sitting on a hyper-ellipsoid.

Both approaches synthesise (or 'harvest', as Ozuysal puts it) sets of dense, 2D texture data with the help of sparse 3D geometry models (*generative step*) and then use large sets of reproduced 2D data for the training of object classifiers (*discriminative step*).

Physically Inspired Models. Several attempts in the graphics community aim at describing deformation by *physically accurate models* that simulate the interaction of underlying structure [7, 38]. Such models are rarely applicable for detection since they rely on knowledge about the forces involved and demand expensive inverse kinematic calculations. They can, however, be used for learning models of other forms which are more applicable to detection.ˆFor the task at hand, structural recognition is especially appealing because, in addition to classification, the approach provides a description of how the given pattern is assembled from primitives. While templates are often used as these primitives of the hierarchy this work will employ statistical detectors as elementary input[20].

The following, last section of the review will focus on one particular physical model that, despite its formal simplicity, underpins the generation of *coat pattern textures* in a number of species. An analysis of the properties of the model will yield features that 1) support the uniqueness assumption which motivated the use of animal patterns as biometric entities, and that 2) can be exploited for designing more efficient and effective recognition techniques.

[20]Structural decompositions can then be built on top of the elementary features to reveal cues about the location of 1) the biometric entity, and 2) the animal's pose.

2.5 Generative Model of Animal Coat Patterns

2.5.1 Auto-generative Patterns on a Compartmentalised Surface

Animal textures represent an evolving, self-organising structure. In the literature, the generation of these ordered patterns (originally described by Turing [201]) has been analysed and exploited for various purposes including natural [177, 202, 215] and artificial [218] pattern synthesis in graphics, physical/chemical simulations [181] as well as biological population descriptions and predator-prey-models [31] – to name only a few.

Despite these efforts, the employment of pattern properties for the purpose of *visual recognition*, that is using structural knowledge gained from the formation domain to aid decisions in the recognition domain, has not been outlined explicitly as yet.

In this section it will be shown that the patterns generated by Turing systems carry a number of different structural and spectral properties which are suitable for the visual representation of species as well as individuals.

First, a generative model that deterministically describes Turing patterns is formally introduced. Let an animal surface be modelled as a discrete[21] 2D manifold $\mathbf{X} = [\mathbf{x}]$, a domain assembled from compartments specified by locations $\mathbf{x} = [x_1, x_2]$. Let the surface be observed through (discrete) time t and let it carry m different morphogene substrates, say A, B, C and so forth, which are distributed over the surface with variable concentrations, say a, b, c, respectively. The latter may be interpreted as representations of the intensity of skin colouration. Figure 2.13(a) schematically illustrates the concept of concentration-dependent surface colouration on a compartmentalised 2D system modelling skin with cellular division.

2.5.2 Reaction-Diffusion

The described system may be interpreted as a discrete, spatiotemporal image function $\mathbf{I}(\mathbf{x}, t) : \mathbb{I}^3 \rightarrow \mathbb{R}^m$ that maps from a position $\mathbf{x}$ on the surface and a time t of observation to a vector of local concentrations: $\mathbf{I}(\mathbf{x}, t) = \mathbf{I}([x_1, x_2], t) = [a(\mathbf{x}, t),\ b(\mathbf{x}, t),\ c(\mathbf{x}, t), ...]$.

After an initialisation with $\mathbf{I}_0 = \mathbf{I}(\mathbf{x}, 0)$ the further temporal development, that is the *dynamics* of the system, are modelled *deterministically* by combining the effects of two qualitatively different processes.

[21] The use of a discrete model does not affect the generality of the concept – results derived from a sufficiently dense model predict well the phenomena observable in real chemical, continuous systems. In fact, the cellular structure of skin suggests that a compartmentalised model is a reasonable choice.

Diffusion: Substances at a certain spatial position $\mathbf{x}$ *diffuse* into adjacent spaces of lower concentration with substance-specific rates formally expressed by a diagonal matrix $\mathbf{D}$ of diffusion coefficients, that is $\mathbf{d}_a$ and $\mathbf{d}_b$ in a 2D system. According to *Fick's Second Law* [61] applicable to continually changing state diffusion, the temporal change of concentration $\frac{d\mathbf{I}}{dt}$ can be described by the spatial Laplacian of concentrations weighted by the diffusion coefficients. For a *2-substance system* this diffusion model yields:

(DIFFUSION EQUATION)

$$\frac{d\mathbf{I}}{dt} = \begin{bmatrix} \mathbf{d}_a & 0 \\ 0 & \mathbf{d}_b \end{bmatrix} \left(\frac{d^2\mathbf{I}}{dx_1^2} + \frac{d^2\mathbf{I}}{dx_2^2} \right) = \mathbf{D}\nabla^2\mathbf{I} \tag{2.16}$$

where the del operator ∇ denotes the spatial image gradient.

Reaction: In addition, substances may locally *react* with each other and thereby progressively change the balance of their concentration according to fixed reaction kinetics[22] expressed by a function $f(\mathbf{I}) : \mathbb{R}^m \rightarrow \mathbb{R}^m$ that maps from the current configuration of concentrations to the predicted change of these concentrations after a reaction step.

The two processes are considered *simultaneously* to influence the substance concentrations as schematically illustrated in Figure 2.13(b) for the neighbourhood of a single spatial compartment at position $\mathbf{x}$.

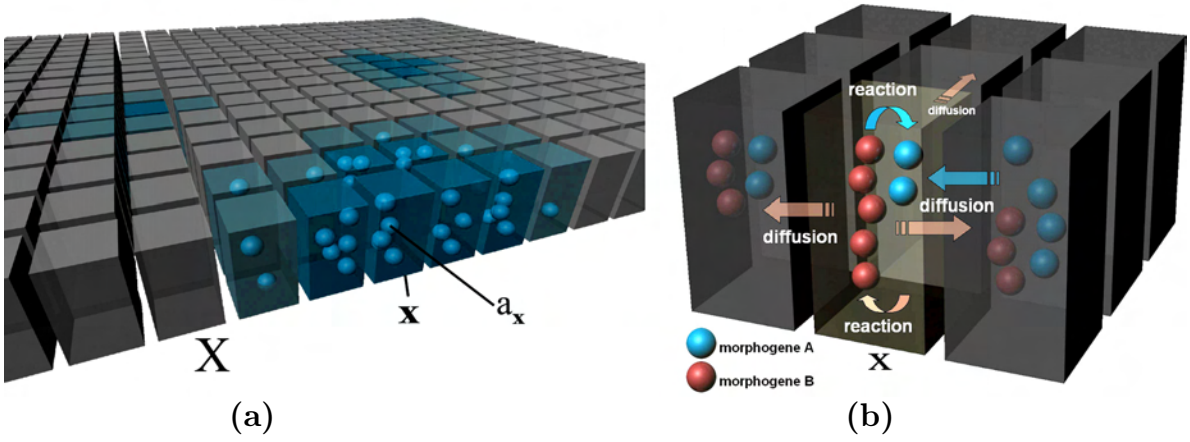


Figure 2.13: ***Reaction-Diffusion Model.*** **(a)** Multiple spatial elements $\mathbf{x}$ (shown as cuboids representing skin cells) form a 2D surface patch $\mathbf{X}$ that models skin where local colouration is induced by high/low concentrations $a_\mathbf{x}$ of morphogene A (visualised as cyan spheres). **(b)** Schematic activity sketch for a 2-chemical-system depicting simultaneous reaction and diffusion around a single spatial element $\mathbf{x}$ causing an ongoing adaptation of the local concentrations of the morphogenes A and B.

[22] See Appendix B.1 for a detailed method of deriving the reaction kinetics from (real) chemical equations employing *The Law of Mass Action* [61].

The resulting dynamical system can be compactly described by a partial differential equation (PDE) that explains the rates of local concentration change as the accumulated effects of both reaction and diffusion:

$$\begin{aligned}&\text{(GENERAL REACTION-DIFFUSION)}\\ &\frac{d\mathbf{I}}{dt} = f(\mathbf{I}) + \mathbf{D} \triangle \mathbf{I}\end{aligned} \tag{2.17}$$

For reasons of simplicity, systems with only two substances a and b are considered for further discussion. They carry all properties necessary for modelling the pattern generation[23] of interest. A general, normalised and non-dimensionalised form[24] of equation (2.17) for a system of such two substances may, after *Murray* [147], be written as follows:

$$\begin{aligned}&\text{(NON-DIMENSIONALISED, 2D REACTION-DIFFUSION)}\\ &\begin{bmatrix} a_t \\ b_t \end{bmatrix} = \gamma \begin{bmatrix} f_a(a,b) \\ f_b(a,b) \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & \delta \end{bmatrix} \begin{bmatrix} \triangle a \\ \triangle b \end{bmatrix}\end{aligned} \tag{2.18}$$

where a_t and b_t are the rates of concentration change for either substrate, γ is a scaling parameter that determines the size s of the domain by adapting the relative strength of reaction and diffusion (and thereby virtually scaling the domain), $\delta = \frac{d_b}{d_a}$ is the ratio of the involved diffusion coefficients and f_a, f_b represent the reaction kinetics separated for each of the substances. Thus, the complete 2-substance system can be described by a model $\mathfrak{M} = (\mathbf{X}, \mathbf{I}_0, f_a, f_b, \delta, \gamma)$ containing a manifold $\mathbf{X}$, an initial distribution $\mathbf{I}_0$ of concentrations, the kinetics f_a, f_b, the diffusion ratio δ and the scaling parameter γ.

The kinetics f_a and f_b establish a 2D vector-field $V(a,b) = [f_a, f_b]^T$. In the absence of diffusion, a *steady state* exists within the field at a location indicated by a zero-crossing of the norm of the state vector $[f_a, f_b]^T$. Depending on the local form of V, extrema may constitute a sink, a source, a saddle point, an isolated point or even a part of isolated curves such as stable orbits. Focussing on the long term development of the system $t \to \infty$, sinks guarantee convergence towards $[f_a, f_b]^T \to [0,0]^T$. Table 2.7 provides reaction kinetics (with one or more sinks) commonly used for the synthesis of Turing patterns.

[23] Surprisingly the overwhelming majority of real animal coat patterns are reproducible with two-chemical systems or combinations/iterations of such [177, 202, 215].

[24] For an in-depth discussion of the transformation between the standard reaction-diffusion Eq. 2.17 and the form as given in Eq. 2.18 see work by *Murray* [146, 147].

Gierer-Meinhardt (1972) [78]	$f_a = k_1 a^2 b^{-1} - k_2 a$	$f_b = k_3 a^2 - k_4 b$
Thomas (1975/76) [196]	$f_a = k_1 - a - \frac{k_2 ab}{1+a+k_3 a^2}$	$f_a = k_4(k_5 - b) - \frac{k_2 ab}{1+a+k_3 a^2}$
Schnakenberg (1979) [182]	$f_a = k_1(a^2 b + k_2 - a)$	$f_b = k_1(k_3 - a^2 b)$
Gray-Scott (1985) [81]	$f_a = k_1(1-a) + ab^2$	$f_b = b(k_1 + k_2) - ab^2$
Generic Activator-Inhibitor	$f_a = \frac{k_1 a^2}{b} - k_2 a + k_1 k_3$	$f_b = k_4 a^2 - k_5 b + k_6$

Table 2.7: **Specific Reaction Kinetics.** Selection of reaction systems that may contain sinks. Their governing kinetic equations are given and free parameters are represented by $k_1, k_2, ..., k_n$.

2.5.3 Where Patterns Emerge: The Turing Space

Diffusion is generally considered an equalising process. Yet, in co-occurrence with converging reactions it can *locally* counterbalance the sink effect of an attracting steady state. This regionally induced reduction of entropy can break the homogeneity of a system and amplify even infinitesimal small local differences in concentration towards persistent, non-constant pattern themes. First spotted by *Turing* [201], the phenomenon is known as *diffusion-driven instability* and can cause the evolution of non-trivial, spatial patterns from almost (but not totally) constant initial conditions. Experimental results in Figure 2.14 illustrate this formation of structure from randomness on a time series of evolving morphogene distributions.

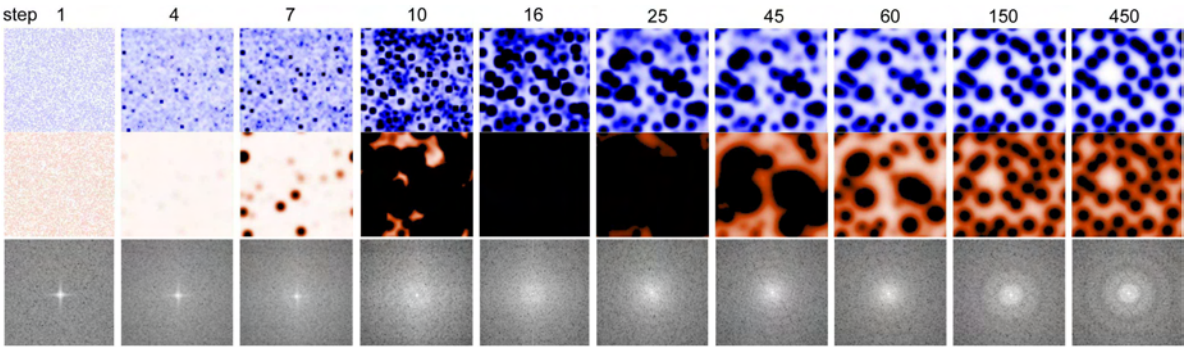


Figure 2.14: **Simulated Evolution of a Reaction-Diffusion System.** The above time series depicts simulation results of an evolving Reaction-Diffusion system on a closed, 128 × 128 domain using the generic activator-inhibitor kinetics given in Table 2.7. The morphogene populations A and B are visualised in the two top rows while the lower row shows the central area of the power spectrum. Note the transition from a initialisation $\mathbf{I}_0$ (at step 1) of Gaussian noise to a configuration of stable, approximately equally sized spots far away from a thermodynamical equilibrium. A system-specific frequency f and its harmonics become apparent as concentric rings in the Fourier spectrum. (parameters used for the simulation: $\gamma = 1, \delta = 10, k_1 = 0.05, k_2 = 0.045, k_3 = k_6 = 0, k_4 = 0.0004, k_5 = 0.2$)

Not all parameterisations lead to emerging structure as in the case depicted. Non-trivial patterns may evolve on the domain $\mathbf{I}$ if and only if 1) the system converges for $t \to \infty$ towards a steady state in the absence of diffusion and 2) diverges in its presence [146, 147]. These constraints define a compact subspace within the domain of free parameters – the

Turing Space[25]. The borders of the Turing Space can be quantitatively estimated utilising *linear stability analysis* as exercised by *Murray* [146, 147]. The process yields inequalities on the free parameters allowing for non-trivial pattern formation:

$$\begin{aligned} &\text{(TURING SPACE BOUNDARY CONSTRAINTS)} \\ &\mathbf{tr}(\mathbf{B}) < 0,\ |\mathbf{B}| > 0,\ F > 0,\ F^2 - 4\delta \cdot |\mathbf{B}| > 0 \\ &\text{where} \quad \mathbf{B} = \begin{bmatrix} \frac{df_a}{da} & \frac{df_a}{db} \\ \frac{df_b}{da} & \frac{df_b}{db} \end{bmatrix} \quad \text{and} \quad F = \delta \frac{df_a}{da} + \frac{df_b}{db}. \end{aligned} \tag{2.19}$$

By applying these inequalities at the steady state, limitations are imposed on the parameter δ and on the reaction parameters k_i of the kinetic functions f_a and f_b. The resulting inequalities circumscribe the Turing Space. Figure 2.15 depicts an experimental result plotting the concentration of one morphogene population of an evolved system against changing parameters (k_1, k_2) where the Turing Space becomes prominently visible as a sickle-shaped, patterned area between homogeneous regions of either high or low concentration.

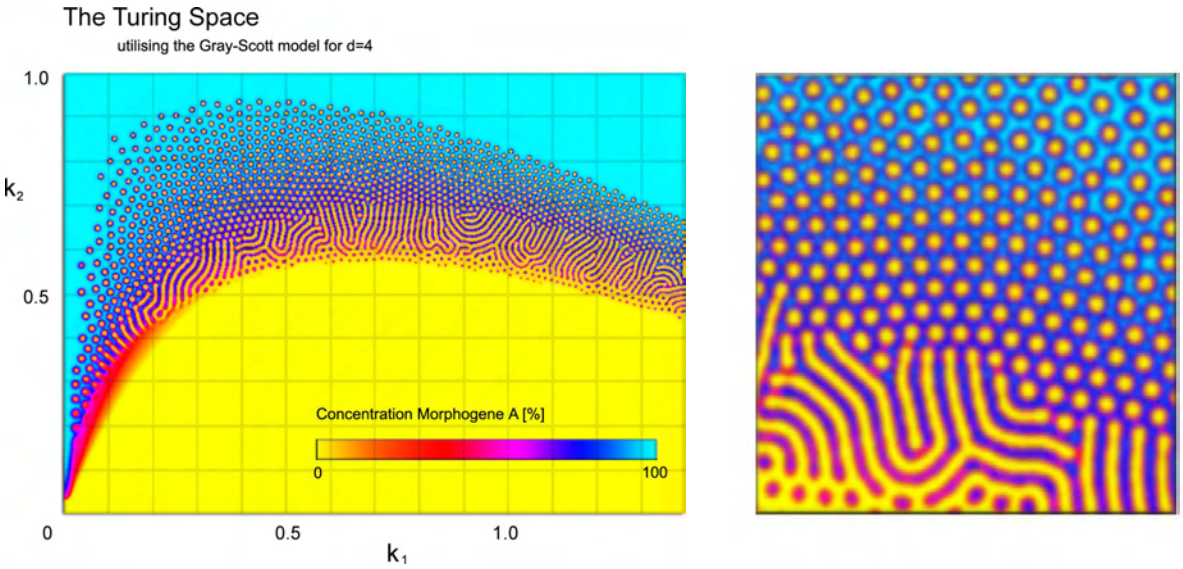


Figure 2.15: ***Visualisation of the Turing Space.*** The image on the right visualises locally occurring patterns superimposed on the pattern-supporting subspace (Turing Space) of the parameter space. Other Reaction-Diffusion systems produce similar patterns. On the left, several pattern classes are identified and associated with biological, structurally similar instances. The visualised system is calculated over 5,000 iterations using the Gray-Scott kinetics and a diffusion ratio of δ=4.

[25]Some sources use the equivalent term 'Turing Set' highlighting that it constitutes a 'true subset' of the set of possible parameter combinations.

2.5.4 Selective Spectral Amplification

Turing patterns exhibit a remarkable spectral property: in the idealised case of fully evolved patterns, that is for $t \to \infty$ and a generation on a simplistic manifold (e.g. on a sphere), a *single band* of dominant spatial pattern frequencies $f = R/\lambda$ is amplified in Turing patterns where R is the image resolution (per dimension) and λ is the dominant (wave)length inherent to elements of the pattern. Consequently, the power spectrum – shown in lower row of Figure 2.14 – exhibits rings[26] that represent this frequency band f and its harmonics. Using the dispersion relation[27] of the Turing model (see *Murray* [147, p.103ff] for details), the band of frequencies f can be approximated quantitatively as:

$$\frac{\gamma L_1}{8\delta\pi^2} < f^2 < \frac{\gamma L_2}{8\delta\pi^2} \qquad \text{where} \qquad L_{1/2} = F \pm \sqrt{F^2 - 4\delta|\mathbf{B}|}. \tag{2.20}$$

L_1 and L_2 represent the two zero crossings of the quadratic dispersion relation which constitute the boundaries of the amplified waveband[28]. Thus, assuming a single Turing system is causing the coat pattern development in animals, deviations from the dominant band over different body regions (e.g. deviations from an equidistant striping in zebras) are due to other factors including complex topology or disproportionate growth *after* pattern fixation (see Figure 2.16).

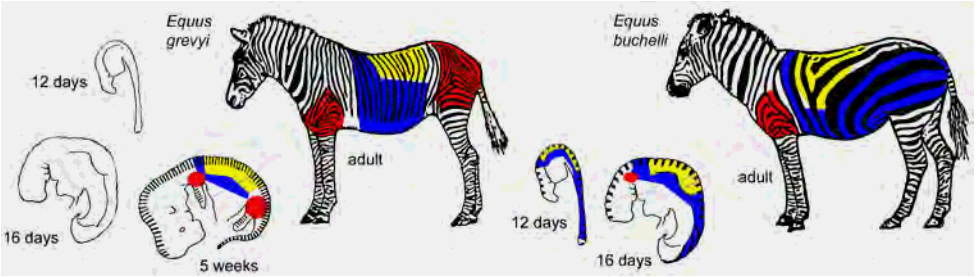


Figure 2.16: ***Growth-related Causes of Pattern Differences in two Species of Zebra.*** The body size at pattern creation determines the pattern frequency exhibited in later life (early lay-down induces lower frequency due to extended growth). Disproportionate growth promotes locally variable pattern frequencies. Regions at body junctions during lay down (red) and anisotropic scale changes during growth (yellow) exhibit highly variable patterns compared to regions of relatively homogenous growth (blue). [images based on work by *Murray* [147]]

[26] In general: The Fourier spectrum exhibits circular segments allowing for anisotropic patterns. Additional natural factors involved (projections, forces, drag etc.) may also cause the formation of elliptical patterns.

[27] A relation linking a spatial (here: domain size) and temporal (here: amplification rate of mode) quantity is known as a dispersion relation. In physics, it also represents a relation between energy and momentum.

[28] Note that wavenumbers $k = 2\pi f$ on finite domains I (as they actually occur in animal coats) are discrete measures. Therefore, in the extreme case of a domain not large enough to carry a single, full waveform, there will be *no* pattern emerging from the system despite the criteria (2.19) being satisfied.

2.5.5 Phase Singularities

Apart from a system-specific configuration of frequency and orientation, Turing patterns also exhibit randomly positioned *phase singularities*[29], that is configurations of local patterns which render the value of the local gradient direction indeterministic.

Overall, six classes of elementary singularities can be differentiated in Turing patterns, that is combinations of two different types and three categories illustrated in Figure 2.17.

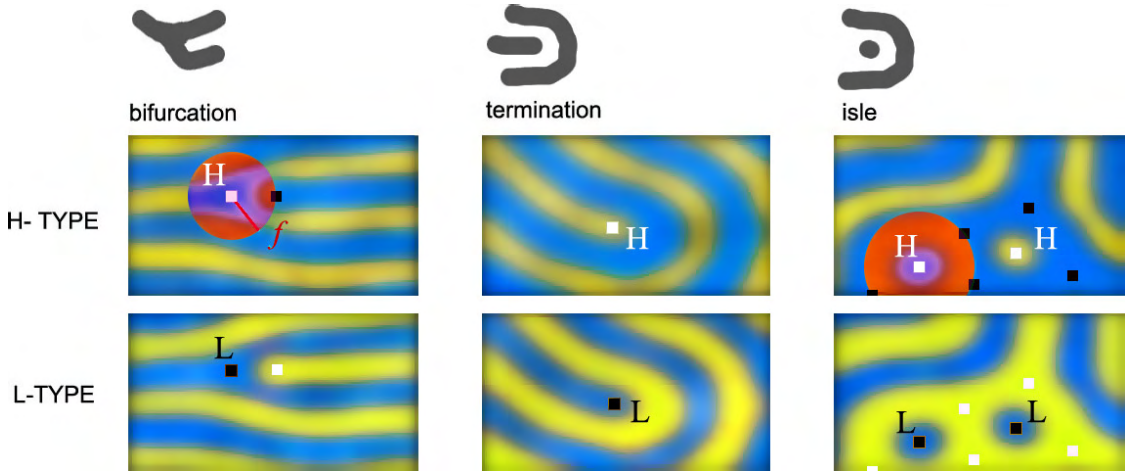


Figure 2.17: ***The Six Phase Singularities of Turing Patterns.*** Phase singularities in Turing patterns occur in areas of high morphogene concentration (H-type shown as white) and in areas of low concentration (L-type shown as black). Singularities are often accompanied by singularities of opposite type at a distance around the dominant frequency f (indicated as red discs). Three categories of singularities can be observed: 1) *bifurcations* where stripes fork, 2) *terminations* where lines end, and 3) *isles* where terminations have degenerated into a (symmetrical) spot.

On coat patterns, a number of regions (e.g. at body junctions) are more likely to develop phase singularities than others. Figure 2.18 illustrates a selection of regions where a local change of geometry or topology renders areas especially prone to disturbance. Similar to minutae in fingerprints [151], the configuration of singularities contained in these regions – e.g. terminations and bifurcations in a zebra's scapular stripes, a penguin's chest spots – will be used compactly to capture the visual characteristics of a individual's patterning. A small number of computer-aided, biometric animal identification systems already use specific types of singularities for the purpose of animal identification, e.g. spots in whales and sharks [6, 208]. However, the generic concept of using singularities to identify individual, Turing-patterned animals has neither been acknowledged nor exploited in the literature so far.

[29] A phase singularity is mathematically defined here as a location of undefined/indeterministic gradient direction. Note that an integration of the gradient direction along a closed path around the singularity yields $2n\pi$, $n \in N$. Morphologically, phase singularities resemble pixels of the skeletonised signal that do not have two neighbours.

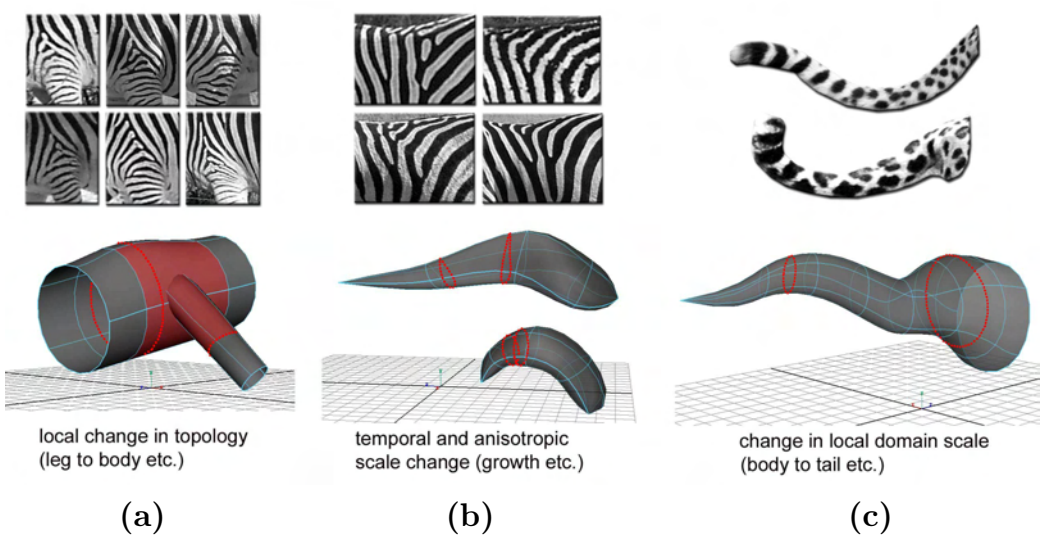


Figure 2.18: ***Geometry and Topology-related Causes of Singularities.*** The regional extent of singularity occurrence is correlated with the geometrical/topological configuration of the animal surface. **(a)** Changing local topology (e.g. at scapular stripes of zebras), **(b)** anisotropic growth (e.g. central body region at the embryonic bend), and **(c)** strongly changing domain size (e.g. between tail and body) trigger locally amplified pattern disturbances that result in singularities. These pattern areas are suboptimal to be used for species detection and ideal for individual identification.

2.6 Chapter Summary and Outlook

In this chapter a review of biometric and related recognition techniques has been presented, tracing the development of the subject from its ancient origins of manually inspected clay prints to state-of-the-art systems that aim for deformation robust, automatic identification. The *design, components* and *performance* measures used in automated biometric systems have been discussed and related to the raised problem of coat pattern identification.

The notion of *biometrical entities*, i.e. the very structures that carry individually unique features, took centre stage in the argument. It has been outlined that a uniqueness property on its own is insufficient for linking a pattern to the identity of an individual – at least, entities must also be species-universal, permanent, measurable and comparable.

Furthermore, it has been argued in this chapter that biometrics on non-human subjects are, despite a scientific need for their development, largely under-researched. The few existing, computer-aided *wildlife identification systems* have been reviewed and critically discussed. It has been concluded that the main drawback of existing systems is the lack of identification robustness and the time-consuming, human involvement during the recognition process. In particular, current systems demand the user 1) to collect the visual data, 2) to find and register the species in the images taken, 3) to normalise entities to some extent manually, and 4) to finally screen the machine-produced matching results.

In the following parts of this thesis it will be shown that these tasks can be automated. A proof of concept will be given by putting forward and evaluating a hybrid vision framework that takes advantage of the reviewed visual properties of *Turing patterns*. More specifically, the configuration of phase singularities is inherently linked to a specific initialisation of a Turing system. Therefore, they will be used as an integral ingredient for differentiating between different individuals. On the other hand, pattern properties which are independent from the initialisation and evolution of a Turing system hold species-characteristic information, i.e. the *type of pattern elements* (e.g. spot or stripe), the *dominant local scale*, and the *orientation of pattern elements*.

In Chapter 3 Haar-like features are used to approximate elementary pattern elements (e.g. spot and line features). Boosting is then utilised to learn the scale, orientation and relative alignment of multiple Haar-like features to characterise a species' appearance. The chapter will explore in how far detectors built on these grounds can be used to find key surface points on animals given multi-view recognition in unseen, largely unconstrained imagery. Chapter 4 suggests a *structural recognition* approach for grouping key points and for associating them to individual animals as well as for extracting and perspectively correcting the biometric texture of interest by posing surface models in the scene.

Chapter 5 finally deals with individual animal identification. It proposes robust methods for the extraction of phase singularities and for the matching their configuration. In conclusion of the chapter the final results of applying the system described to detection tasks in a real animal populations will be discussed. ■

Chapter 3

APPEARANCE-BASED SPECIES DETECTION

'A representation is a formal system making explicit certain entities or types of information, together with a specification of how the system does this.' [134]

(David Marr, 1945 - 1980)

3.1 Chapter Overview

The following chapter focusses on the first subtask of the proposed 'locate-pose-identify' paradigm: it is concerned with representing and rapidly detecting a species. The approach proposed captures a species by estimating locations of key reference points strategically spread over their coat pattern. Each of these points is characterised by the surrounding image structure and learned from sample images complemented by local gradient features for fine localisation. The neighbourhood classifiers operate in a Haar-like feature space since dimensions of this domain approximate macro-elements of Turing-like patterns.

For implementation, the framework by Viola and Jones [212] is expanded to a perspectively guided point-surround descriptor, providing for close-to-realtime recognition of multiple key points. The *effectiveness and efficiency*[1] of the method is quantified and critically discussed for the task of single and multi-pose species detection in natural scenes that contain deformation, clutter, lighting changes as well as intra-species resemblance.

In addition, the effect of Turing patterning on the technique's performance is documented. Throughout the chapter, to underline a generic suitability of the approach, the analysis focusses on three species of very different genus and appearance, namely plains zebras, African penguins and lions[2].

[1]The term 'effectiveness' is used to indicate the need for *high detection robustness*, while the term 'efficiency' refers to the need for a *fast and sparse* classifiers. The latter is of particular importance to avoid buffering potentially weeks worth of high-resolution media streams created by environmental cameras.

[2]Scientific species names of the animals analysed are *Equus burchellii, Spheniscus demersus*, *Panthera leo.*

3.2 Species-characteristic Key Points on Coat Patterns

3.2.1 Object Description via Clouds of Annotated Points

The human visual system can often reconstruct the presence and pose of a species from a few locally confined texture patches and their spatial arrangement – as visualised in Figure 3.1(a) – where missing data is substituted with a-priori knowledge about the species and body pose consistent with the observation. Note that, as illustrated in Figure 3.1(b), only the combination of both local components and their global structure define the object.

Sparse patch descriptions can, neglecting pose information encoded in local features, be interpreted as a shape of class-annotated reference locations (see Figure 3.1(c)). Following this modelling concept, the semantics of single points in the cloud are defined by visual information on: 1) the location within the point cloud, 2) the motion properties of points,

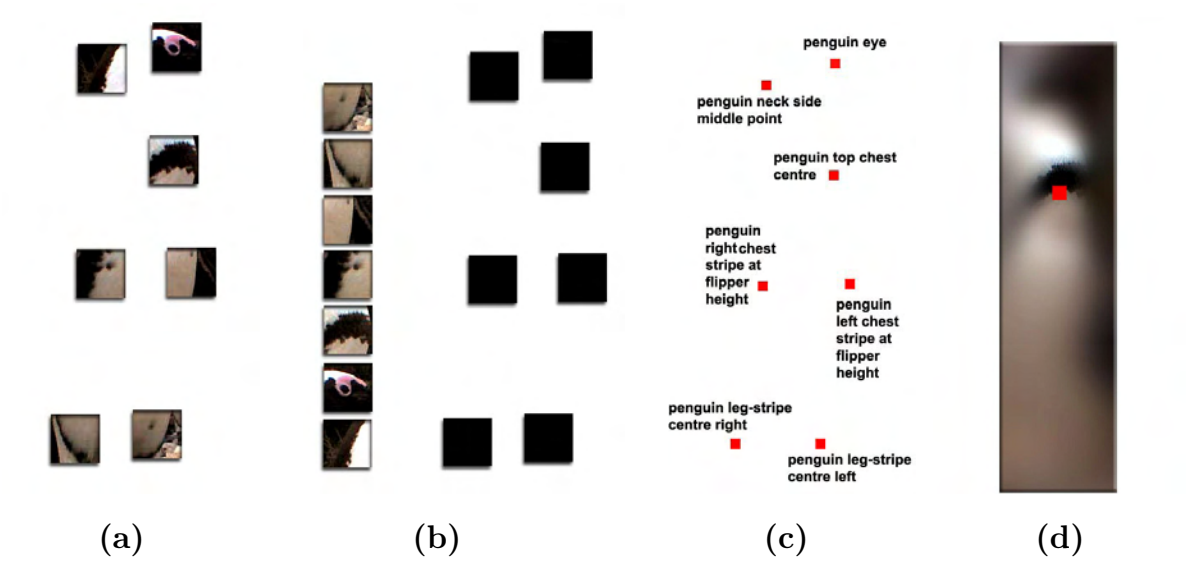


Figure 3.1: ***Localised Semantics.*** **(a)** A small set of texture samples contain sufficient information to show an African penguin in a specific pose; **(b)** Separating the local features from their arrangement renders the object unidentifiable. **(c)** The example visualised as class-annotated reference locations, that is a shape of semantic key points. **(d)** The description of single key points is not confined to the local context, but reaches out to distant, yet class-characteristic features (e.g. the relatively large, homogenously white chest below the neck stripe of African penguins). Peripheral features are often of low spatial frequency (indicated by blur) since variance increases with distance.

and 3) the local visual context of points. This chapter deals with exploiting the last. As indicated in Figure 3.1(d), the context used for description will be widened beyond strictly local information since distant features support the recognition process. They often carry important, class-defining characteristics. In further contrast to *prototypical* local descriptors [130, 161], salient [144] or interest points [180], semantic key points represent

location information valid for an *entire class of variable objects*, i.e. a species.

In fact, most species carry visual features that exhibit properties universal to the population. Species-typical landmarks are often related to specific organs[3] (e.g. eyes, nose etc.) or topological body junctions (e.g. base of the leg). In coat patterned animals, there exist additional, purely textural[4] landmarks that are also idiosyncratic of the species. Figure 3.2 illustrates a selection of species-typical landmarks (labelled in blue).

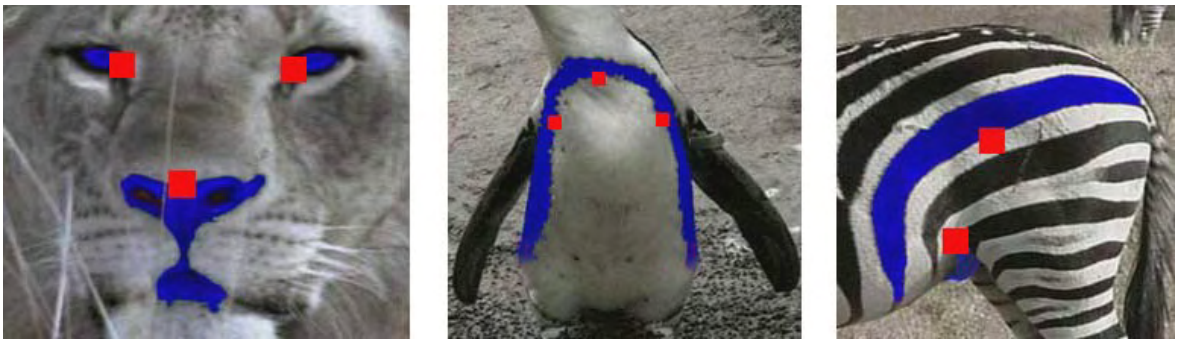


Figure 3.2: ***Landmarks and Key Points on Animal Surfaces.*** Facial features of lions, the chest band of African penguins and the central hindquarter stripe on zebra coats (all marked in blue) are examples of species-typical, visual landmarks that occur on all members of the population. A selection of key points linked to the edges of these landmarks are visualised in red. [original images: I10, I01, I02]

Distinctive reference points (exemplified in red) associated to these contrast landmarks are proposed to serve as a set of appearance-based anchors for species modelling. They will be referred to as *semantic key points*. A chosen set of key points should – ideally – exhibit the five properties[5] summarised in Table 3.1. Compliance with the five features portrayed

Property	**Explanation**
universality	the key points can be found on all species members
stationarity	key points provide a stable, absolute reference on the animal surface
localisability	truly local image information exists that allows for pinpointing key points
distinctiveness	key points can be disambiguated robustly from other content by their context
surface coverage	key points are distributed sufficiently dense to probe the surface area of interest

Table 3.1: ***Proposed Properties of an Ideal Set of Semantic Key Points.***

[3] Note that locations of the eyes, eyebrows, nose and mouth are traditionally employed in component-based, human facial recognition systems [97] or as reference shapes in active appearance models [36].

[4] A generation of species-wide similar Turing patterns is a genetically controlled procedure. Exact, generative reaction-diffusion models of this process exist, for instance, for a number of butterflies [147,p.161].

[5] Describing interest points, *Schmid et al.* [180] split the property 'distinctiveness' into *repeatability of detection* and *entropy of representation*. Classification processes, however, interlink the two properties.

transforms a set of 'ordinary' surface locations into class-specific, sparse landmark points that clearly capture information on 1) the actual presence of species members in an image, on 2) the configuration[6] of body parts, and on 3) an object-centred reference system, which helps relating *individual* pattern features in different individuals.

In essence, semantic key points capture a set of elementary locations on variable content combining sparse point representation [180] for a spatial description with the concept of component-based class representation [97] for a semantic description.

3.2.2 Selection of 'Good' Key Points

Before creating actual descriptors and detectors for key points, one needs to stipulate their locations on a species' surface.

Traditionally, accurate point positioning is achieved by exploiting corner locations, which are defined by a significant signal change at the reference point in all signal dimensions [180]. Given the set $\mathcal{X} \subseteq \mathbb{I}^2$ of all signal locations, the change of the signal $\mathbf{I}$ at a particular position $\mathbf{x} = [x_1, x_2]^T \in \mathcal{X}$ can be approximated by the eigenvalues $\lambda_{1,2}$ of the signal's auto-correlation matrix $\mathbf{A}$ over a small neighbourhood $S \subseteq \mathcal{X}$ around $\mathbf{x}$ [90]:

(AUTO-CORRELATION MATRIX)

$$\mathbf{A} = \begin{bmatrix} \sum_{\mathbf{x}\in S} \frac{\partial^2 \mathbf{I}}{\partial x_1^2} & \sum_{\mathbf{x}\in S} \frac{\partial^2 \mathbf{I}}{\partial x_1 \partial x_2} \\ \sum_{\mathbf{x}\in S} \frac{\partial^2 \mathbf{I}}{\partial x_1 \partial x_2} & \sum_{\mathbf{x}\in S} \frac{\partial^2 \mathbf{I}}{\partial x_2^2} \end{bmatrix}. \tag{3.1}$$

Two large eigenvalues ($\lambda_{1,2} \gg 0$) provide superior, full 2D localisation in the form of *corners* as, for instance, utilised in the detection systems by *Harris* [90], *Förstner* [64] or *Shi and Tomasi* [187].

Coat patterns exhibit strong changes of signal intensity, mainly for the provision of disruptive colouration [89, 47]. However, in most cases corners constitute phase singularities of the gradient's tangent. Thus, the position of corners in Turing patterns is individually unique, that is neither *universal* nor *stationary* with respect to the species[7]. Figure 3.3 illustrates this variability of corner locations observed on upper chest patterns of African penguins. Thus, corner points do not contribute towards positional stability of reference points on a species' surface. However, the corner constraint can be weakened to a basic edge

[6] Note that for stationary key points, the relative location of fixed landmarks on the surface is, given a certain pose, also stable. Therefore, the relative configuration of (ideal) key points determines the pose.

[7] See Section 2.5 for details.

Figure 3.3: ***Variability of Corner Measures in Animal Coats.*** Images show the upper chest area of four different African penguins. Interest points (extracted using the Harris detector [90]) are superimposed on the two rightmost images. It can be seen that, comparing different animals, cornerness is not a characteristic feature of any specific pattern location.

criteria Λ seeking to stipulate key points along strong signal changes, which are present and for a number of features stable in all disruptive camouflage patterns of a population:

$$\begin{aligned} &\text{(RELIABLE 1D-LOCALISATION CONSTRAINT)} \\ &(\Lambda = \max(\lambda_1, \lambda_2)) \gg 0 \end{aligned} \tag{3.2}$$

Key points from this candidate set are positioned to cover the body area of interest with a net of probing points. Figure 3.4 visualises the procedure of key point selection for the chest area of African penguins.

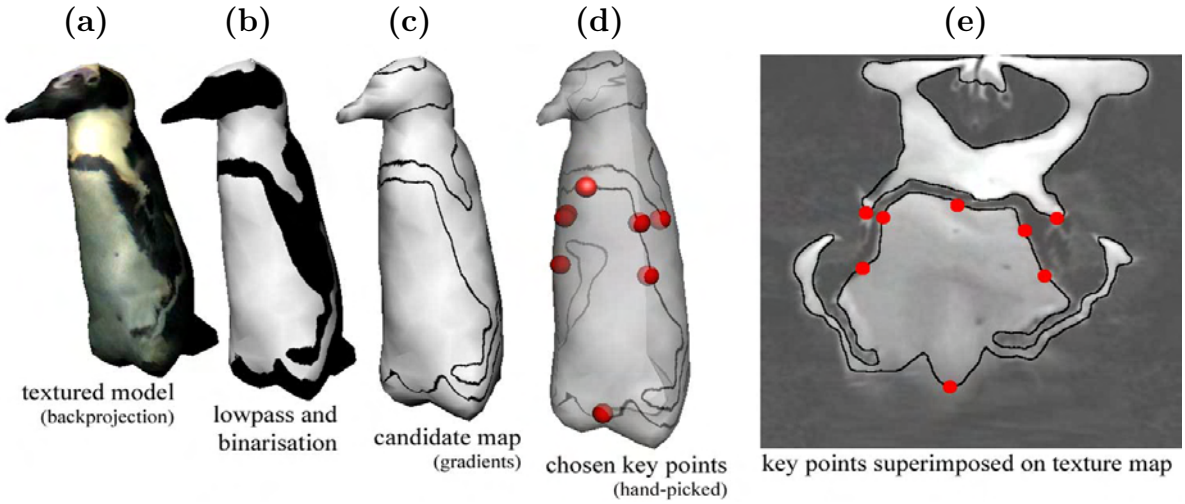


Figure 3.4: ***Key Point Selection.*** **(a)** a textured species model obtained by manual 3D modelling and backprojection of the texture from a real animal (stitching of several 2D images); **(b)** version of the texture, low-pass filtered at the dominant frequency of the chest band followed by optimal thresholding; **(c)** candidate map of strong gradient locations; **(d)** superimposition of 9 key points manually chosen to cover and approximate the chest area; **(e)** model texture with key points and candidate map;

None of the five conditions described in Table 3.1 is met fully when considering the variability of coat patterns in real animal populations. However, a gradient-confined candidate selection promotes a species-wide applicable fine localisation of key points in *one spatial direction* (that is along the gradients normal vector). The location in the remaining dimension is approximated from the surrounding image structure.

In order to produce a context descriptor in a robust yet compact fashion, the local signal neighbourhood needs to be confined in terms of its size and shape relative to the object, its resolution, and the form of its representation. Once these parameters are fixed, the aim is to learn class-distinctive, local context descriptors which determine a certain type of key point *globally*.

First, the question of appropriate sizes and proportions of contexts are investigated where the methodology proposed will be exemplified on a small set of specific key points acting as representative samples.

3.2.3 Class-dependent Neighbourhood Confinement

The appearance of coat patterns notably varies over a population, that is both the shape of the residing landmark as well as the appearance of structure in a wider context around a key point differ from individual to individual. Figure 3.5 illustrates this natural phenomenon of 'intra-species variability' in hindquarter patterns of a sample zebra population.

Figure 3.5: ***Sampling Zebra Patterns.*** Depicted are 60 different, hand-labelled samples registered at a key point from an image set of 600 individual plains zebras. [image source used for experiment: I02]

Despite the global variance present, a regional analysis of these patterns over a population shows that a specific landmark point is strongly correlated with *local features* while the variance aggravates with increasing distance[8] to the reference. Figure 3.6(b) visualises the standard deviation of key-point-registered hindquarter patterns over a population of zebra. It can be seen that there is a rapid drop of correlation when moving away from the reference where only a few local stripe features are stable (dark colouration) with regard to

[8]Proximity relations are commonly linked to structural and geometrical alterations (e.g. viewpoint change, pose change) as, for instance, modelled by *Berg and Malik* [22, 21] through the concept of 'geometrical blur', that is uncertainty which increases with the distance to the reference. For coat patterns, however, the texture itself exhibits proximity-dependent variability.

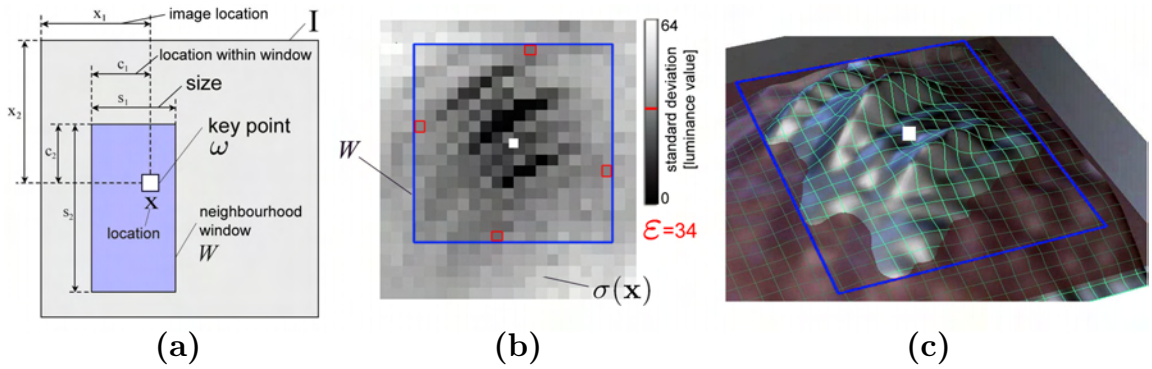


Figure 3.6: ***Confinement of the Neighbourhood Window.*** **(a)** window components; **(b)** image variance function of zebra hindquarters calculated over a population of 200 and the stipulated neighbourhood window (blue) by using a bounding box circumscribing the class-specifically thresholded variance function; **(c)** 3D visualisation of the thresholding plane cutting the correlation functional.

the key point (white block). This observation suggests bounding the object area used for the description of a key point to a *local region* potentially smaller than the object.

In order to quantify the confinement, a neighbourhood window $W(\mathbf{x}) \subseteq \mathcal{X}$ is introduced (marked in blue). It is defined by its key point location $\mathbf{x} \in W$, the relative position $\mathbf{c}$ of this point within the window W and the window size $\mathbf{s}$ (see Figure 3.6).

Using a representative image set Ψ_{pos} containing a species, centred at the key point and taken at approximately the same view aspect, the intra-population standard deviation $\sigma(\mathbf{x})$ over contrast normalised versions of the images Ψ_{pos} is constructed as:

$$\text{(INTRA-POPULATION VARIABILITY)} \qquad \sigma(\mathbf{x}) = \frac{1}{|\Psi_{pos}|}\sqrt{\sum_{\mathbf{I}\in\Psi_{pos}}(\mathbf{I}(\mathbf{x}) - \mu(\mathbf{x}))^2} \quad \text{where} \quad \mu(\mathbf{x}) = \frac{1}{|\Psi_{pos}|}\sum_{\mathbf{I}\in\Psi_{pos}}\mathbf{I}(\mathbf{x}). \tag{3.3}$$

Thresholding the function σ cuts out some compact region around the key point $\mathbf{x}$ (see Figure 3.6(b)) where the threshold value ε is chosen such that the cut area is connected and does not cover content outside the object. The bounding box circumscribing this region is used as a neighbourhood window whose structure will be exploited for learning and detecting a key point in cluttered environments. Thus, W is constructed on the basis of class-dependent[9], sample-driven features. Figure 3.7 illustrates examples of neighbourhood windows.

3.2.4 Learning Key Points from Population Samples

Is possible to robustly identify key points using the information within the neighbourhood windows despite the high degree of intra-class variance and the presence of visual clutter

[9]Most systems that engage neighbourhood descriptions for classification purposes utilise either a constant [161] or an image-dependent, yet class-independent window size for description, e.g. SIFT [130].

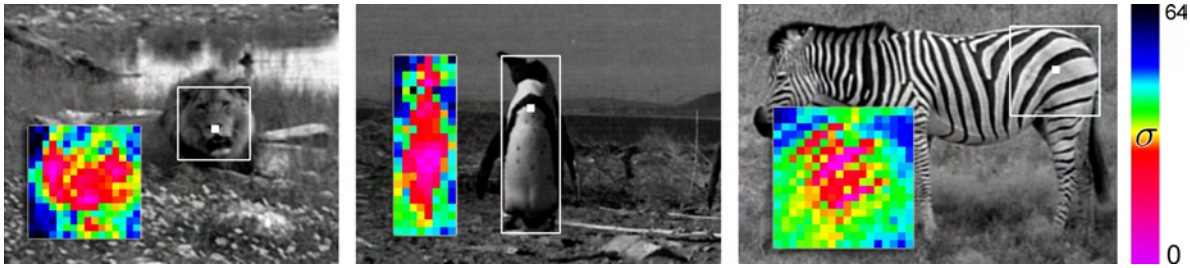


Figure 3.7: ***Neighbourhood Windows.*** Depicted are examples of key points (white blocks) and their associated neighbourhood windows (white rectangles) constructed using the method described where 200 images were used per species as positive samples. Coloured superimpositions show the cut section of the variance function σ [measured in luminance values] within the neighbourhood window. Note that the window may cover the main area of the body (e.g. penguin) or only small fractions (e.g. zebra). In this context, notice that textural regions are found to be structurally stable even far away from the reference landmark, provided the features are either homogenous and large (penguin's white chest) or rigidly connected (lion's eyes). In Turing-patterned zebra coats, however, stable correlations are – as predicted by the Reaction-Diffusion model – confined to a local region around the reference. [wildlife images used for the experiment: I10, I01, I02]

in natural environments? According to the properties of Turing's generation mechanism, coat patterns are predicted to contain a population-universal appearance component with regard to the local arrangement of the dominant 1) orientation, 2) feature types, and 3) size (i.e. spatial frequency) of pattern elements.

The *learning task* of extracting the species-characteristic pattern information, disambiguating it from clutter and the 'noise' of individual influences has to be solved before effective descriptors and detectors can be constructed. In particular, class-stable relations have to be discovered between the image space $\mathbf{\Psi}$, covering the visual evidence, and the semantic space $\mathbf{\Omega}$, containing the species-specific key point classes ω (see Figure 3.8).

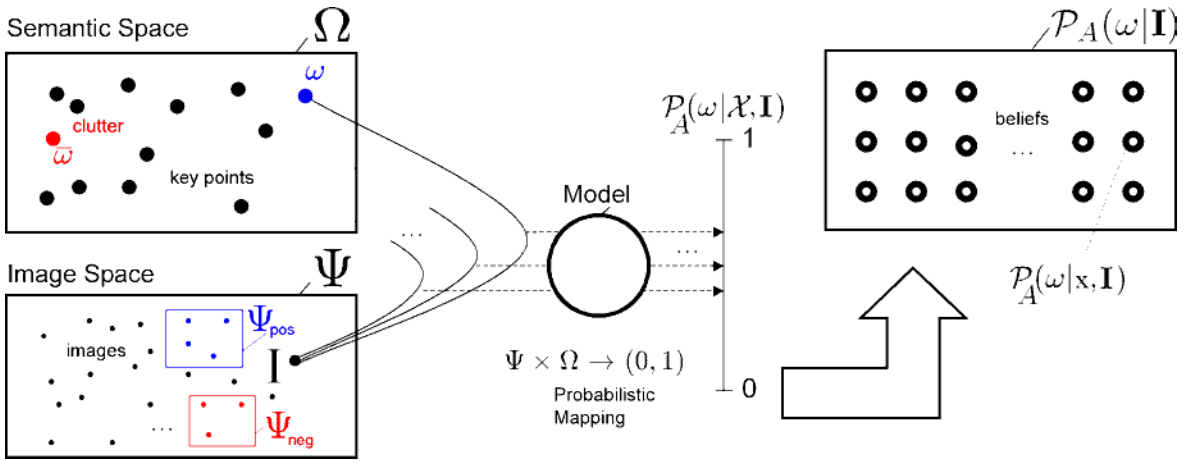


Figure 3.8: ***An Abstract View on the Domains Linked by Supervised Learning Applied.***

Clearly, the full domain $\mathbf{\Psi}\times\mathbf{\Omega}$ spanned by combinations of images $\{\mathbf{I}\}$ and meanings $\{\omega\}\bigcup\{\bar{\omega}\}$ (where $\bar{\omega}$ represents clutter) is – as repeatedly acknowledged for most vision applications of practical relevance [157] – too large to be sampled exhaustively for the training of classifiers.

Supervised learning is employed, modelling a key point class ω by analysing *comparatively small, yet representative sample sets* containing some hundred or thousand (pre)classified images.

The necessary expert knowledge is then given in the form of annotation, that is spatial selection and subsequent classification into image sets Ψ_{pos} and Ψ_{neg} generated by manual annotation of image data.

In particular, the set Ψ_{pos} contains hand-labelled image contexts registered at key points ω. The images are taken from real-world populations (as shown earlier in Figure 3.5) where the neighbourhood window is scaled to normalise for different object sizes. The negative set Ψ_{neg} covers randomly sampled patches of the natural habitat and neighbourhoods of other key points.

3.3 Local Description of Key Points

3.3.1 Extensions to the Viola-Jones Framework

The Viola-Jones framework[10] is utilised and extended to learn the species information contained in the sets Ψ_{pos} and Ψ_{neg}. In contrast to the standard approach [126, 211, 212, 213], a number of modifications and extensions will be applied in order to tailor the technique to the problem of key point description and to reduce the impact of a number of major drawbacks including poor localisation properties of the detector.

Note that the detector will be used to extract *appearance evidence in video* using both a frame-by-frame analysis as well as a tracking framework.

Thus, the real-time capabilities of the Viola-Jones framework are essential for an efficient online-operation and can, therefore, not be compromised for the achievement of other detector features. Table 3.2 summarises the major modifications and extensions proposed.

[10] The technique has been reviewed and motivated in detail earlier in Section 2.4.4.

Feature	Motivation/Explanation
point-surround description (Section 3.2)	In most applications of the Viola-Jones framework [126, 212, 211, 213] the bounding box of the object is used as the image patch to be learned. It will be demonstrated that modelling the neighbourhood of single key points, that is using the Viola-Jones classifier as a *class-specific point-surround descriptor*, improves the localisation performance with regard to a reference.
resolution prediction (Subsection 3.3.5)	Instead of heuristically choosing the resolution [8, 111, 126] of the window used for training, it will be shown that the presence of a dominant spatial frequency band in Turing patterns can be exploited for estimating a suitable detector resolution.
perspective constraints (Subsection 3.5.5)	It will be demonstrated that the introduction of perspective constraints can significantly reduce the sub-domain of the scale-location space processed during detection. Thus, both runtime speed as well as detection performance can be improved significantly.
smart labelling (Subsection 3.3.2)	Instead of hand-labelling all the thousands of samples necessary for a representative description of a pattern class, a form of supervised bootstrapping is proposed to reduce labelling efforts.
dense belief maps (Subsection 3.5.1)	A real-valued, dense approximation of the appearance-based observation density is proposed instead of utilising binary classification outputs. This approach will provide the basis for integrating the technique with structural recognition.

Table 3.2: ***Overview of Extensions/Modifications Applied to the Viola-Jones Detector.***

3.3.2 Iterative Training via Supervised Bootstrapping

In order to avoid hand-labelling several thousand images for the creation of a sufficiently large set Ψ_{pos} as a learning base, an iterative scheme is engaged that progressively improves premature classifiers by supervised bootstrapping[11]. The method suggested replaces – at least for parts of the sample space – the need for manual labelling of neighbourhood windows by a simple flagging of misclassifications.

Initially, several hundred positive images $\Psi_{pos}^{i=0}$ are manually produced in order to capture the main variance occurring in the image set, e.g. lighting situations and the broad spectrum of patterns. Training the system on $(\Psi_{pos}^{i}, \Psi_{neg}^{i})$ then yields a classifier H_i which is (using the tracking extensions proposed in the next chapter) applied to a sequence Ψ^i of several hundred novel images containing the species at some unknown image location. This classification

[11]The term *bootstrapping* refers back to the German legend about Baron Münchhausen, who claimed to have lifted himself out of a swamp by pulling up his own boot straps [1]. In its current scientific meaning, the term is used for the prediction, construction or generation of a complex system or model from a more simple, yet given dataset. For the problem at hand, a small set of hand-labelled data Ψ_{pos}^{0} is given together with a large set of unlabelled data Ψ^i. The task is to help producing a classifier by smartly exploiting – i.e. bootstrapping from – the pre-labelled set, i.e. to reduce manual labelling efforts.

process reveals two new sets of automatically classified images $(\bar{\Psi}^i_{pos}, \bar{\Psi}^i_{neg})$, resulting from the scale- and translation invariant detection on Ψ^i. Manual viewing of $\bar{\Psi}^i_{pos}$ reveals[12] the set of false positives $\Psi^i_{mis} \subseteq \bar{\Psi}^i_{pos}$ that is to be flagged in a next step (see Figure 3.9(b)).

This strategy is convenient for the user since the set of positive detections is relatively small, that is $\bar{\Psi}^i_{pos} \ll \bar{\Psi}^i_{neg}$. In addition, the set of misclassifications generally decreases during further iterations and, hence, $\bar{\Psi}^i_{pos} \ll \bar{\Psi}^{i+n}_{pos}$ is likely to hold for $n \gg 0$. Most importantly though, labour-intensive labelling can be replaced by simple flagging.

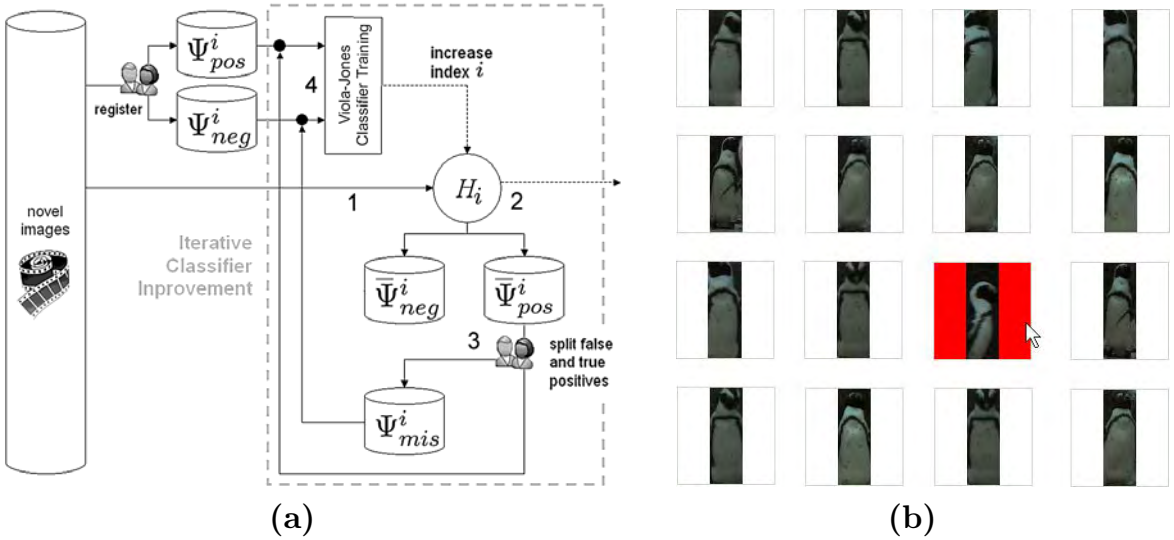


Figure 3.9: ***Iterative Training.*** **(a)** progressive cycles of 1) drawing novel samples, 2) performing classification based on present classifier, 3) supervised flagging of the resulting positive classifications, and 4) retraining the classifier on the novel training sets, steadily increases the classifier's performance. **(b)** For flagging, multiple images can be viewed at once to increase the user's efficiency in spotting false positives (exemplified on a side view detected as a frontal).

At the end of an iteration step, a new set of training samples is created as $(\Psi^{i+1}_{pos}, \Psi^{i+1}_{neg}) = (\Psi^i_{pos} \bigcup \bar{\Psi}^i_{pos} \setminus \Psi^i_{mis}, \Psi^i_{neg} \bigcup \Psi^i_{mis})$ which is utilised to produce a classifier H_{i+1} during the next learning loop. The process is stopped at i_{max} once the rate of misclassified positives, that is $|\Psi^i_{mis}|/|\bar{\Psi}^i_{pos}|$, falls below a user chosen threshold or no more samples are available for training. Figure 3.9(a) shows a schematic flow chart summarising the method.

The asymmetry produced by not screening $\bar{\Psi}^i_{neg}$ is tolerable since 1) the set $\bar{\Psi}^i_{neg}$ is not used for further training, 2) it is easy to produce a large initial set Ψ^0_{neg} by random filming of clutter, and 3) the application to individual identification demands a low false positive rate, that is exclusively species members are to be detected, while a minor increase in the rate of

[12] A detection is counted as successful if, and only if the detection was made not further away from a true landmark than 25% of the size of the neighbourhood window.

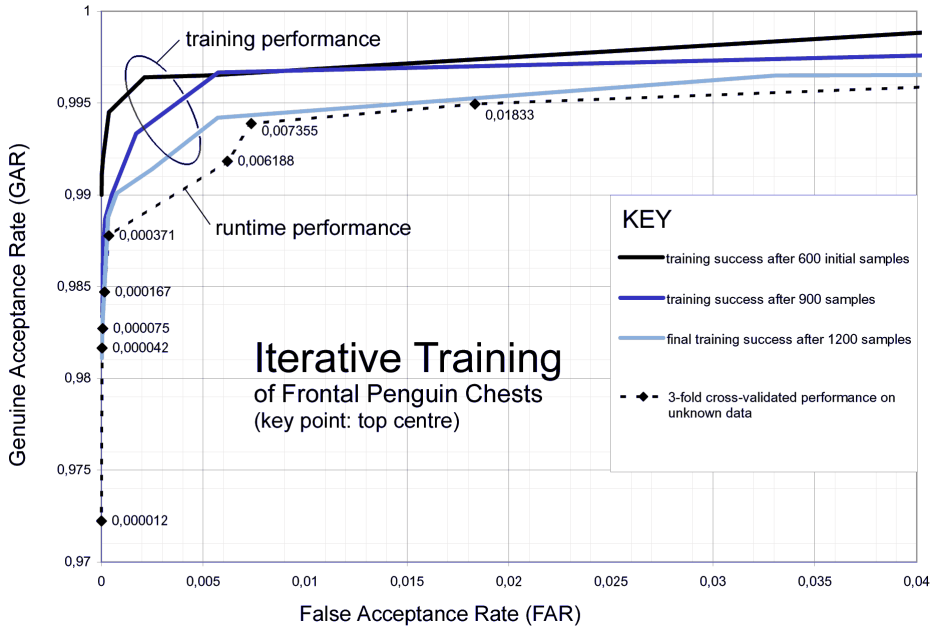


Figure 3.10: ***Performance Improvements During Training.*** The graph shows receiver operating characteristics (ROC) at different stages of the training process for penguin chests. The increasing difficulty of incorporating more and more samples causes a progressive drift of the performance curves away from the ideal (top-left) position in ROC space. The application to novel data induces a further drop in performance (dashed line). However, the overall performance of the classifier shows – despite an iterative classifier construction – a sufficiently high performance characteristic similar to common measures in human face detection [126]. For instance, at around 98% genuine acceptance rate the false acceptance rate is measured below 4×10^{-5}. [image source used for the experiment: I01]

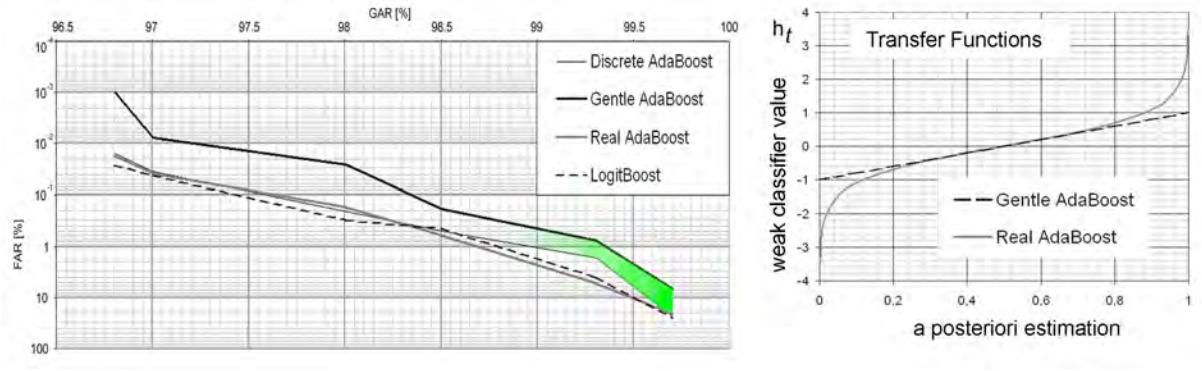
false negatives (resulting in lower population coverage) is acceptable. Figure 3.10 quantifies and discusses the training performance at different stages of the learning process[13].

3.3.3 Performance of Different AdaBoost Variants

The results of the learning procedure are crucially dependent on the choice of the boosting algorithm applied. Therefore, different variants of AdaBoost are tested and compared in order to optimise the disambiguation capabilities of the resulting classifiers.

[13]Learning is conducted using luminance samples quantised at 8×24 pixels. Gentle AdaBoost and CART structures (limited to a depth of two branching levels) are used as the learning model. 600 positive samples are initially labelled and trained against negative image patches containing random parts of the habitat. The resulting classifier is progressively applied to detect penguins in videos. Two flagging iterations – covering 300 screenings of positive detections and re-training per round – are used as further supervision steps, yielding a total of 1200 positive samples. For estimating the classification performance on novel data, 3-fold cross-validation is applied: the 1200 images are partitioned into three sets. Classifiers are then trained over all pairs of partitions (i.e. over three sets of pairs containing 800 images each) where the retained set (of 400 images) is used for testing the performance. The average performance of the three classifiers on novel data is used for constructing a runtime characteristic (dashed line). Other forms of testing could be considered, however, holdout validation has a higher risk of containing a bias, while higher n-fold cross-validation is time consuming (estimated at $\approx 1.2 \cdot n$ hours of training/testing time).

Previously, *Lienhart et al.* [126] documented the superiority of Gentle AdaBoost over discrete and non-gentle boosters when disambiguating human faces from other content. Their findings are confirmed here for the animal patterns investigated. Figure 3.11(a) visualises the performance results for learning zebra hindquarters using different boosters.



(a) **(b)**

Figure 3.11: ***Booster Performance.*** **(a)** The receiver operating characteristic of different boosting algorithms generated by averaging over threefold cross-validation trained on the zebra hindquarters data. It can be seen that Gentle AdaBoost outperforms other variants tested. **(b)** The functions used to transform a posteriori measurements (abscissa) into weak classifier outputs $\mathbf{h}_t$ (ordinate) are shown. Note that gentle boosting (dashed line) adopts a linear function - instead of a logistic relationship - by mapping the probabilistic interval $(0, 1)$ straight onto a weighting interval $(-1, 1)$. Limiting the trust in the data at the extremes of the spectrum results in a better generalisation compared with logistic characteristic (solid line).

It can be seen that, at a constant genuine acceptance rate, Gentle AdaBoost outperforms the other boosters tested in reducing the false acceptance rate by an average of a factor of 5 as indicated by the green band.

As illustrated in Figure 3.11(b), a linear transfer function is used by Gentle AdaBoost for the generation of weak classifier values $\mathbf{h}(\mathbf{I}_j)$ from the a posteriori estimates $\mathcal{P}(y_j = 1|h_1(\mathbf{I_j}), ...h_n(\mathbf{I_j}))$ which are calculated over the training set, that is:

$$\begin{aligned} &\text{(GENTLE, LINEAR TRANSFER FUNCTION)} \\ &\mathbf{h}(\mathbf{I}_j) = 2\mathcal{P}(y_j = 1|h_1(\mathbf{I_j}), ...h_n(\mathbf{I_j})) - 1 \end{aligned} \tag{3.4}$$

where the $\mathbf{I}_j$ represent samples, $y_j \in \{-1, 1\}$ holds the ground truth assigning the samples to either the negative or positive set, respectively, and the h_i stand for the output of threshold classifiers of the CART used for testing. Gentle AdaBoost reduces the maximum impact of single weak classifiers compared to logistical transfer functions[14] and, thus, avoids an overwhelming importance of any *single CART classifier*.

[14]For example, Real AdaBoost employs a logistical function $\mathbf{h}(\mathbf{I}_j) = \frac{1}{2} \ln \left(\frac{\mathcal{P}(y_j=1|h_1(\mathbf{I_j}),...h_n(\mathbf{I_j}))}{1-\mathcal{P}(y_j=1|h_1(\mathbf{I_j}),...h_n(\mathbf{I_j}))} \right)$.

Assuming the existence of isolated samples, that is instances containing rare features, the observed strength of Gentle AdaBoost can be explained by the resulting 'gentle' handling of strongly performing classifiers during learning. Since animals show both strong pattern variations and commonalities, the algorithm feature directly applies to the task of learning characteristics of pattern themes in animal populations: instead of relying on the few strongest features (at the extremes of the a posteriori spectrum), gentle boosting limits the trust in any single feature, allowing the technique to deal with the variability of single features and, thus to generalise better.

Nevertheless, Gentle AdaBoost focusses on strong characteristics: the first CART classifiers produced by gentle boosting reflect the (intuitively) most typical features of the pattern class investigated. Figure 3.12 exemplifies this observation on classifiers trained around a lion's nose tip (frontal view). Characteristics such as the species-typical, vertical nose-line and the dark eye region below a brighter forehead are captured by the classifier.

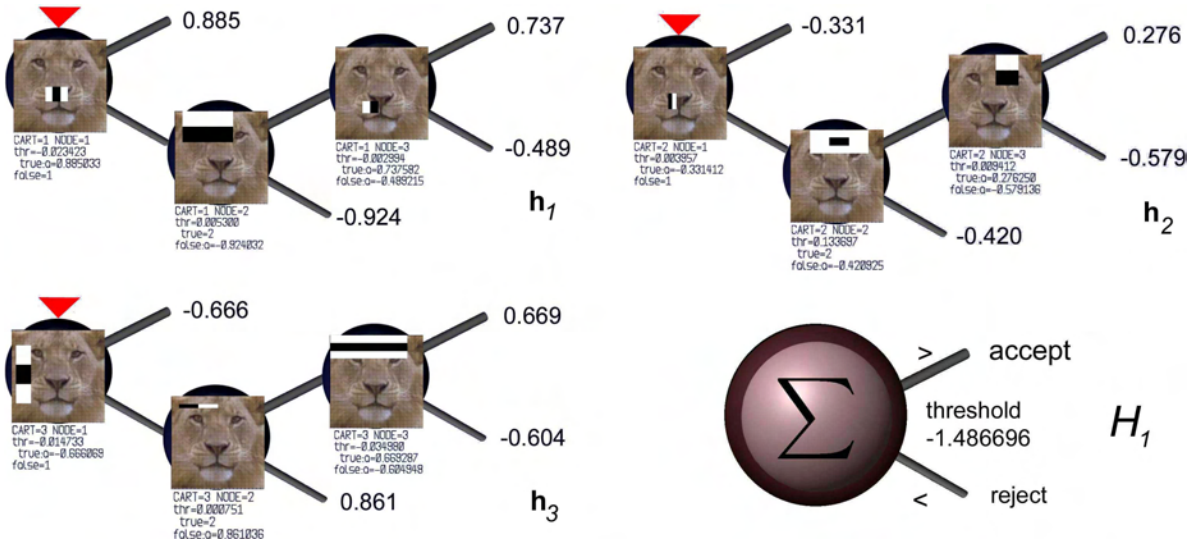


Figure 3.12: ***Example of a Key Point Descriptor.*** The figure illustrates the first out of 30 classifier stages trained in a boosted cascade for the description of a lion's nose tip (in frontal view). The first stage classifier H_1 is applied to a novel image by evaluating the 3 CART's $\mathbf{h}_1$ to $\mathbf{h}_3$ shown (top node indicated by red triangles) and comparing the accumulated leaf values against a threshold $\beta = -1.486696$. The classifier illustrated comprises 9 Haar-like features. It contributes to the overall classification by filtering out 99.7% of positives while accepting only 20% of negatives given novel data of frontal lion faces and clutter. [wildlife images used for training: I10]

3.3.4 Compact Description of Turing Patterns

The Haar-like functions employed for feature representation imitate lines, spots etc. These structures constitute primitives of Turing patterns, which is a strong motivation for choosing this specific feature set as kernel primitives. One can hypothesise that, as a result, Turing

patterned animal coats can be represented highly efficiently by Haar-like features. This hypothesis is tested through an analysis of classification performance and descriptor size, comparing patterned and non-patterned species. Figure 3.13 provides a visualisation of the analysis, plotting the ROC performance curves (solid lines) for the three species together with the associated descriptor sizes (dotted lines).

It can be seen that, at a common genuine acceptance rate (GAR plotted on the abscissa), the classification of zebra patches is achieved employing less than half the weak classifiers used to describe lion faces (note the 'green gap'). A comparison at a constant false acceptance rate (FAR plotted on the right ordinate) produces – as indicated by grey bars – a similar result: it confirms the suitability and 'elegant' nature of Haar-like descriptors for explaining biological Turing patterns, thus, satisfying Ockham's razor[15].

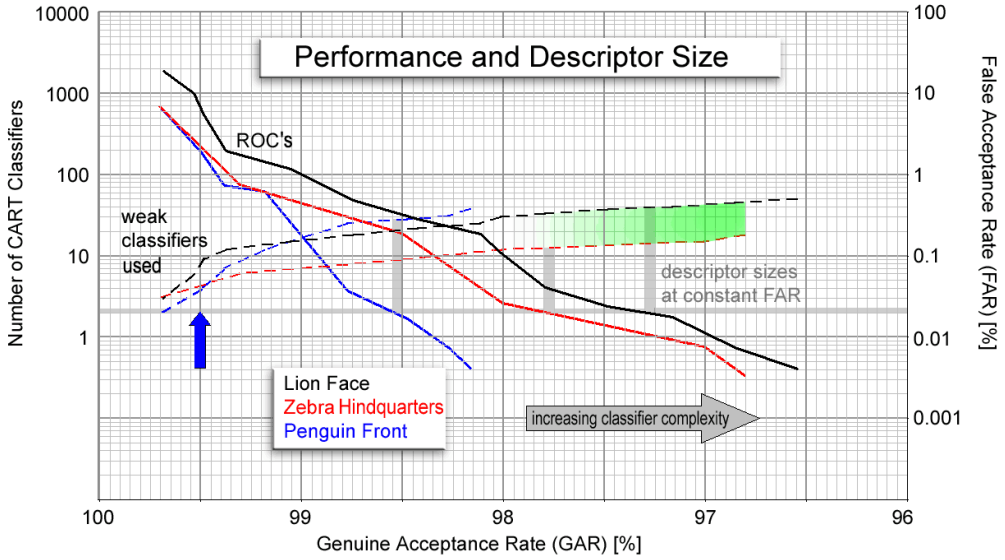
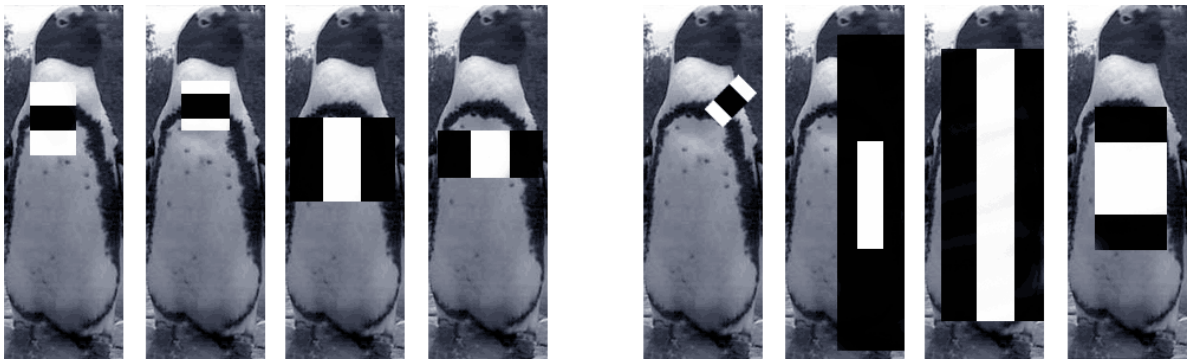


Figure 3.13: ***Performance vs. Descriptor Size.*** Plotting both the descriptor size (dotted) and the false acceptance rate (solid) against the genuine acceptance rate provides information about the balance between a classifier's performance and its complexity. A steeply falling receiver operating characteristic (ROC) and a slowly rising descriptor size indicate effective and efficient learning, respectively. Whilst penguins can be learned most effectively, zebra hintquarters are learned most efficiently. Note that each CART classifier used contains three Haar-like features.

In contrast to zebras, penguins carry limited Turing patterning, e.g. chest stripe, neck band, side bands etc. Once these features are exhaustedly exploited for description (as approximated by a blue arrow) the boosting procedure is forced to focus on non-Haar-like, more variable characteristics (white chest shape, feet etc.) which can not be explained by using

[15]The term refers to the law of parsimony, first outlined by William of Ockham in the 14^{th} century. It states the – now long-standing – science tradition that, "...other things being equal, simple theories are preferable to complex ones." [219, p.180], where the descriptor size traded against the encoded information (or absence of descriptive error with regard to the class) is commonly used as a measure of simplicity [85].

small kernel sets any more. As a result, the rate of classifier employment, i.e. the gradient of the function representing the classifier count, increases steadily (note the logarithmical scaling of the graph). The actual Haar-like features learned confirm this exhaustion of elementary representable features for penguin content as shown in Figure 3.14.



(a) **(b)**

Figure 3.14: ***Decreasing Feature Suitability.*** **(a)** the first four, most characteristic Haar-like features learned by gentle boosting describe stripe and chest; **(b)** Four features picked later in the process (after the 20^{th} CART) show, in some cases, little resemblance to the described penguin – combinations of multiple features are now necessary to further advance the descriptive accuracy.

However, the feature combinations used for describing penguin chests are highly class-characteristic throughout the spectrum analysed (as reflected by a steeply falling ROC curve in Figure 3.13).

It is concluded that the Haar-like feature descriptor employed is a suitable model for explaining the species-specific component of Turing patterned animal coats since: 1) it describes Turing patterns with particularly *high efficiency* with regard to the trade-off[16] between goodness-of-fit and complexity, 2) while preserving the effectiveness and generalisation capabilities of the descriptor previously confirmed in domains such as human faces[17] and non-organic, rigid objects [111].

[16] Note that an optimally compact descriptor is uncomputable for a specific pattern without full class knowledge, that is there is no general, calculable algorithm which, given a subset of samples of a pattern class, produces its shortest description [85].

[17] *Lienhart et al.* [126] describe the performance of the single-view Haar-like detector on human faces to be 0.5% FAR at 95.3% GAR (one specific point on the ROC curve). For the task at hand, the same FAR is achieved for lion faces at 98.75% GAR and for penguins and zebras at over 99% GAR. Clearly, the effectiveness of the descriptor is maintained in the domain of the coat patterns investigated, hinting towards a slightly smaller pattern class complexity than found in arbitrary human faces.

3.3.5 Estimating a Classifier's Spatial Resolution

Turing systems suppress features that exhibit spatial frequencies outside a band around a dominant frequency f typical for the system. Features of non-dominant frequency may exist in individuals, but are predicted to be uncharacteristic with respect to the pattern class spanned by Turing systems with equal parameterisation typical of a species.

Exploiting this observation, it is proposed to control the quantisation[18] of a key point's neighbourhood window by using the dominant band f as an upper resolution bound.

First, the band's existence is confirmed in local patches of coat patterns in real-world animal populations. Figure 3.15 illustrates the dominant frequencies of zebras and penguins.

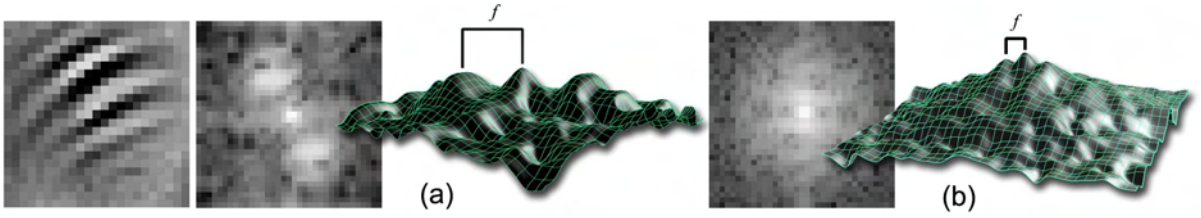


Figure 3.15: ***Dominant Local Frequencies in Patches of Natural Coat Patterns.*** Next to the population average (depicted at the left for zebras), the power spectra are shown for different species: **(a)** plains zebra hindquarter patches and **(b)** frontal neck patches of African penguins. The visualisations reveal the presence of peaks at a dominant frequency f. [wildlife images used: I01, I02]

Attempting a lossless representation below the dominant frequency f, *Shannon's theorem* [186] suggests sampling above the critical frequency f_s determined as twice f:

$$\text{(NYQUIST-SHANNON CONDITION)} \qquad f_s > 2f. \tag{3.5}$$

Assuming the species-specific information is locked below f, sampling below f_s is then predicted to weaken significantly a classifier's performance due to loss of species-characteristic information, while sampling above f_s should lead to no major increase in descriptiveness[19].

The hypothesis is tested by experiment on the zebra example, measuring the influence of the sample resolution on the performance of the detectors built. The necessary test data is created by spatially quantising both positive and negative sample sets at different resolutions (see Figure 3.16).

[18] Traditionally, the resolution of the detection window is heuristically determined [212, 126].

[19] It is the authors conviction that the principle can be generalised and is applicable to provide a guidance for choosing the resolution of detection windows in all detection scenarios where class-specific information is residing exclusively within some upper bounded region of the frequency domain.

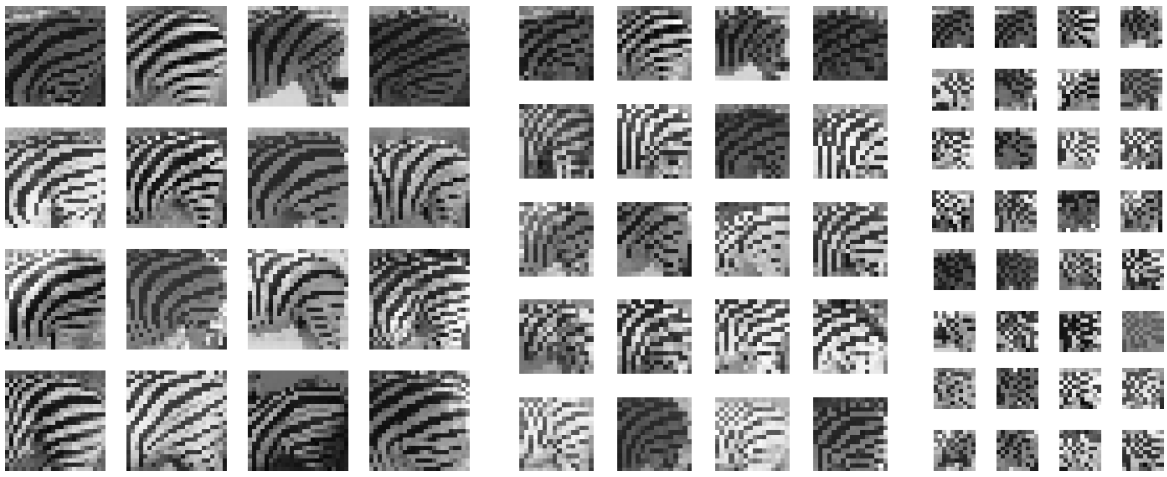
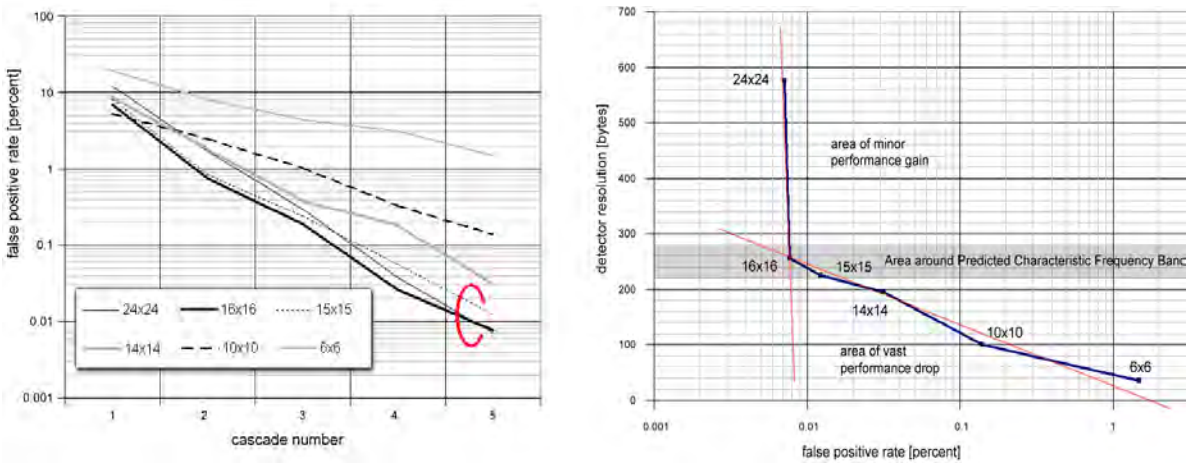


Figure 3.16: ***Resolution-quantised Sample Sets.*** The images show representative subsets of the positive training data for zebra hindquarters sampled **(a)** above twice the characteristic Turing frequency (24×24 pixels), **(b)** at twice the characteristic frequency (16×16 pixels) and **(c)** below twice the characteristic frequency (10×10 pixels) of the local coat pattern neighbourhood. Note that, intuitively, stripe patterns are clearly visible in the two image sets quantised above and at the critical value, whilst the final set shows a clear drop in recognisability. [original images: I02]

The performance of classifiers trained on this data is evaluated and compared at points of common genuine acceptance rate (GAR). The results for zebra hindquarters are plotted in Figure 3.17. It can be seen that the area of highest gradient change of the resolution-dependent learning performance coincides with the critical frequency $f_s \approx 2f$ which is indicated by a grey band at a resolution of 16×16 pixels in Figure 3.17(b).

It can be seen that decreasing the image resolution below this critical value causes the classification performance to drop significantly (note the change of gradient visualised by red lines). On the other hand, it can be seen that an increase of the sample resolution above the critical value does not significantly improve the detector's performance.

These observations indicate that there exist, as predicted by Turing's theory, only limited species-characteristic information in the frequency band above f. The measured dominant frequency and experimentally found values of detector resolutions are given in Table 3.3. For practical considerations note that inaccurate labelling (e.g. using bounding boxes etc.) potentially degenerates the relevance of f with respect to the classifier trained.



(a) **(b)**

Figure 3.17: ***Sampling Resolution vs. Classifier Performance.*** Graphs characterise the performance of gently boosted classifiers for plains zebra hindquarters with respect to the sample resolution. **(a)** Performance curves for different resolutions: shown is the false acceptance rate (at common genuine acceptance rate) at different learning cascades. The performance of classifiers operating at or above twice the dominant pattern frequency ($f_s = 15.2$) lies closely together (red circle). **(b)** Plotting the FAR at constant GAR (at cascade 5, that is at 97% hit rate) against the used detector resolution reveals a sudden change of the classifier's performance around the predicted critical resolution (cross-section of red lines).

Image Set	Resolution Used	Dominant Frequency Measured [1/width]
Zebra Hindquarters	16×16	7.6
Penguin Neck	8×24	3.8
Lion Face	16×16	no dominant frequency

Table 3.3: **Examples of Detector Resolutions.** Since characteristic pattern features reside within a dominant sub-band, detectors should resolve the signal at a frequency above twice this band.

3.4 Practical Detection of Key Points

3.4.1 Lighting Normalisation

Having discussed the training, size stipulation and resolution confinement of coat pattern classifiers, the focus is now shifted towards applying classifiers as pattern detectors.

Natural environments comprise a highly variable set of lighting conditions, considering different times of the day and the seasons. Figure 3.18(a) illustrates some of the lighting scenarios witnessed in a South African penguin colony. In order to compensate for the variance in luminance induced by different, non-specularly reflected light, the signal measured is modified (for both training and detection).

Figure 3.18: ***Lighting Correction using z-Scores.*** Illustrations depict the Lambertian lighting correction applied. **(a)** original RGB sensor images after aperture and gain optimisation; **(b)** isolated luminance channel; **(c)** luminance channel, z-Score corrected over neighbourhood window; [wildlife image source: I01]

Following an earlier application by *Lienhart and Maydt* [126, 37], the z-Score measure – that is essentially contrast centring and stretching – is employed for normalising the signal contained in each neighbourhood window $W(\mathbf{x})$:

$$\begin{aligned} &\text{(Z-SCORE LIGHTING CORRECTION)} \\ &\bar{W}(\mathbf{x}) = \tfrac{W(\mathbf{x})-\mu}{2\sigma} \end{aligned} \tag{3.6}$$

where μ is the mean and σ is the variance taken over the window W. This form of correction can be applied without significant reduction of the runtime speed since both mean and variance are calculated efficiently using the basic and square integral image [126]. Therefore, z-Score lighting correction demands just 4 extra table lookups[20] for each μ and σ.

Note that a full lighting normalisation that also accounts for specular reflection (e.g. wetness of feathers or fur, changes in sun-angle and environmental conditions) is not pursued. Covering this form of variation would require both the availability of a dynamically measured BRDF (Bidirectional Reflectance Distribution Function) of the animal surface and the direction and parameters of *all* the light sources and environmental reflectors.

[20] See Section 2.4.4 for details on the efficient calculation of the integral image during preprocessing.

3.4.2 Detection Performance under Natural Conditions

Classifiers used for practical detection are selected amongst potential candidates by defining a 'working point' on the receiver operating characteristic, that is pinpointing a single position on the ROC curve by fixing all parameters[21] of the training process. Figure 3.19(a) visualises the working areas (red squares) chosen for the three key points in focus. Note that the working points are stipulated at particularly low false positive rates around $4 \cdot 10^{-3}$ %

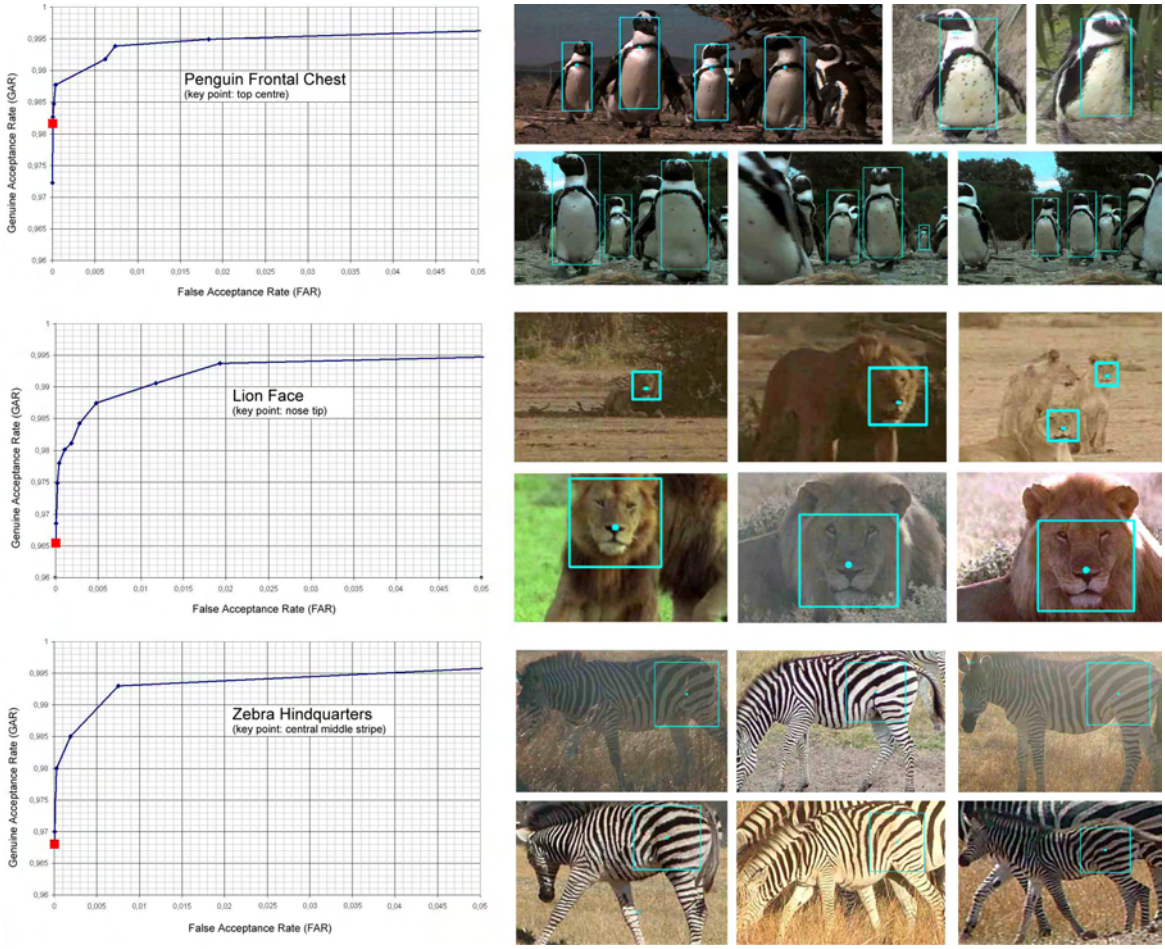


Figure 3.19: ***Performance Plots of Key Point Classifiers.*** **(a)** The graphs depict the receiver operating characteristics of the classifiers trained. Again, they were created using threefold cross-validation. Red squares mark the manually chosen working point on the curves. **(b)** The images show examples of true positives detected in novel data. It can be seen that instances of key points are recognised across a wide range of lighting conditions (see lions), individual patterns (see zebras), deformations and minor variance in pose (see penguins). [images used from: I01, I10, I02]

[21] For the classifiers at hand, the working area is stipulated by setting two parameters: 1) the *length of the boosting cascade* where longer cascades move the working area left in ROC space, and 2) the *spatial density of classifier hits* required to spark a detection. The latter parameter is fixed for all key points (see Section 3.5.1 for details), however, higher densities would also move the working area left in ROC space.

in order to produce very reliable positive detection sets. Clearly, this approach trades recall performance for precision[22], favouring a robust decision on true positives over a wide coverage of the population of positives, i.e. the species. Hence, detections produced by the classifier can be treated with high confidence by later, higher-level stages of the recognition process. Table 3.4 summarises the performance of single-view detectors at working point.

Species/ Image Source (key point described)	**Performance (⌀GAR/⌀FAR [%])**	**Feature Count/ Resolution**	**Training Set Sizes (pos | neg/validation)**
Zebras/I02 (hindquarters)	96.83/0.0034	54/16×16	400 | 11,764,706/200
Penguins/I01 (top chest)	98.25/0.0042	114/ 8×24	800 | 19,047,619/400
Lions/I10 (nose tip)	96.56/0.0041	150/16×16	600 | 14,341,146/300

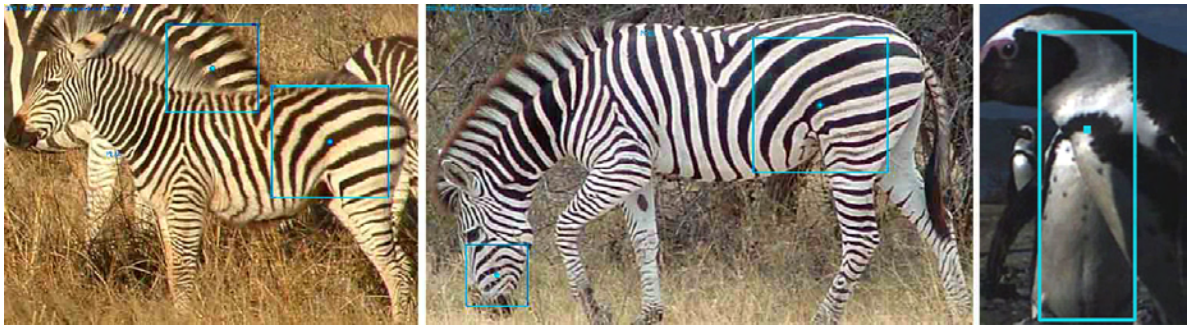
Table 3.4: **Classification Performances at Working Point.** The table shows measured performance characteristics, the number of Haar-like features used and the resolution for key point classifiers at the working point. All classifiers achieve a ⌀GAR above 96% at a ⌀FAR around $3 \cdot 10^{-5}$. Results are generated averaging over 3-fold cross-validation where the cardinalities of the used sample sets are shown in the rightmost column. Negative samples are generated from a set of 10,000 habitat images by choosing patches at random scale and position.

3.4.3 Limitations of the Key Point Detector

Despite the fact that millions of sample patches are used to generate the above performance landmarks, the performance measured takes into account only a fraction of the true variance present in natural environments. Given the operation of the classifier in uncontrolled conditions, the cross-validation employed can, therefore, only approximate the generalisation capabilities of the learning methods and should be seen as an 'optimistic estimate'. To provide practically applicable insights into the limitations of the classifiers, a number of categories of false classifications are discussed now, focussing on the question: what are repeatedly noted conditions that lead the detectors astray?

Cryptic Resemblance. Applying the classifiers to groups of patterned animals reveals one major cause for false positives: zebras and penguins alike exhibit disruptive patterning [89, 47] and cryptic resemblance [54, 139] to conspecifics. As a result, pattern combinations of several animals can imitate regional patterns of a single animal, misleading detectors (see Figure 3.20). Regional appearance descriptors are not suited to counter this form of camouflage. Thus, the issue will be tackled later by adding another layer of abstraction, modelling an individual by a species-typical arrangement of *multiple* key points.

[22]The term '*recall*' is synonymous to the genuine acceptance rate (GAR) while '*precision*' is defined as the ratio of the true positives and the sum of true and false positives: $precision = \frac{TP}{TP+FP}$.



(a) **(b)** **(c)**

Figure 3.20: ***Camouflage through Cryptic Intra-specific Resemblance.*** **(a)** A configuration of two zebras creates an image region that resembles the pattern of a hindquarter patch, leading the detector astray, **(b)** A false positive is induced by (a rarely observed, pose-dependent) cryptic intra-individual resemblance. **(c)** Two penguins create a visual configuration similar to the visual structure of a single penguin – the detector is mislead. [wildlife images used: I02, I01]

Environmental Clutter. Some of the background content present in wild habitats comprises highly variable structures such as trees or bushes. Depending on the viewing angle, wind, the time of day and season etc., they create widely random patterns which often replicate features inherent to the animal coats learned. As a consequence, the detector misclassifies some of these regions as positives. Figure 3.21 illustrates three examples of false positive detections due to environmental resemblance. Both the use of structural recognition and a perspective constraint (that binds the size of a detection to its image position) will be used to further reduce the number of false positives resulting from clutter.

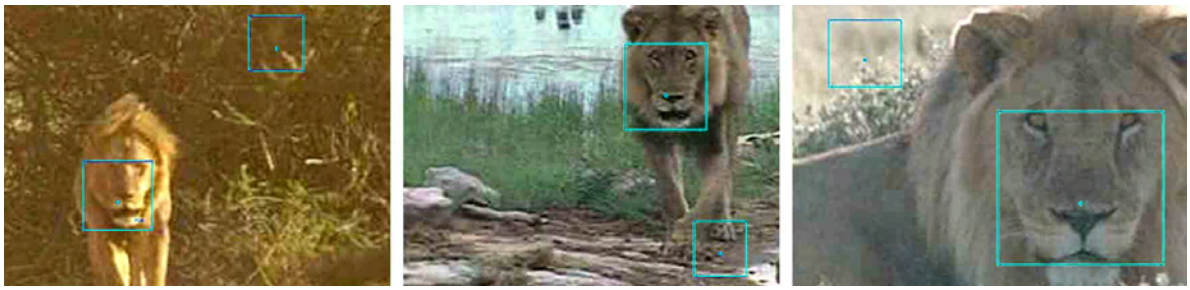


Figure 3.21: ***Clutter in Natural Environments.*** The images depict false positive detections due to clutter in images that also contain a true positive. Note that the false detections exhibit some basic features of a lion's face (e.g. darker regions where eyes should be situated). Although such clutter-induced misdetections are rare, the sheer variety of random structures present in natural environments increases the probability of class-resembling clutter. [wildlife images used: I10]

Lighting and Shadows. The combination of diverse positive training samples, the generalisation capabilities of AdaBoost and the lighting correction employed promote a successful classifier application to a wide range of different lighting scenarios. Figure 3.22 illustrates a selection of conditions covered using the facial lion detector.

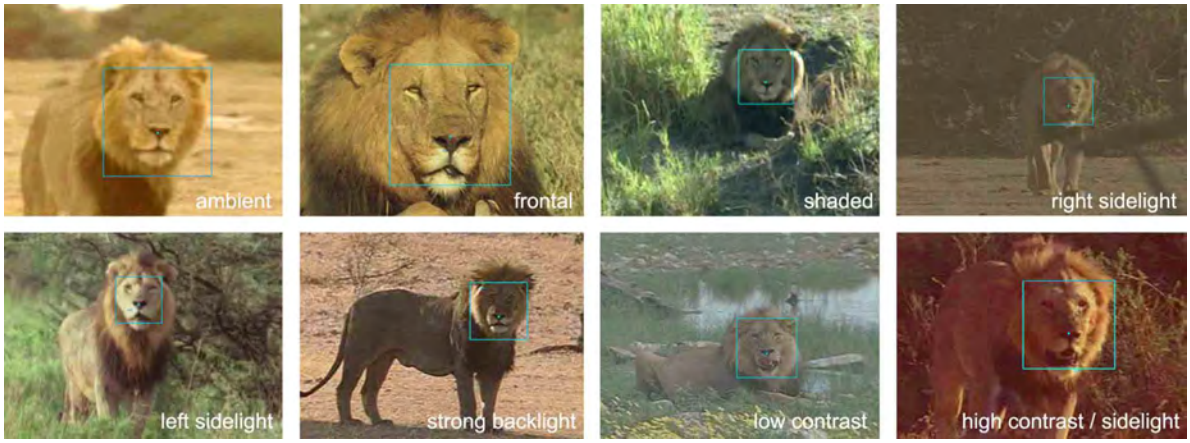


Figure 3.22: ***Range of Accepted Lighting Conditions.*** The images illustrate the range of lighting situations covered by the facial lion detector. Note detections in scenes with frontal, side and back-lighting as well as high and low values in contrast and luminance. [wildlife images used: I10]

Despite the wide spectrum of applicability, the lighting normalisation employed does neither account for deep shadows nor for the non-Lambertian reflective component. As a result, lighting conditions that significantly alter the texture in areas of key features cause the classifiers to fail, producing false negatives. Figure 3.23 documents such missed detections where shadows and specular reflections create novel surface textures (e.g. shadow of the beak in penguins). Unaccounted lighting variations constitute a form of *information limitation* [103] where the hidden or altered data results in a drastic decrease in informational overlap between samples of the same semantic class.

Iterative retraining of the classifier after including spotted misclassifications of this kind into the positive training set has been successful in a number of cases to increase coverage. Nevertheless, an exhaustive sampling of all possible lighting situations is considered infeasible for this project given a highly changeable environment.

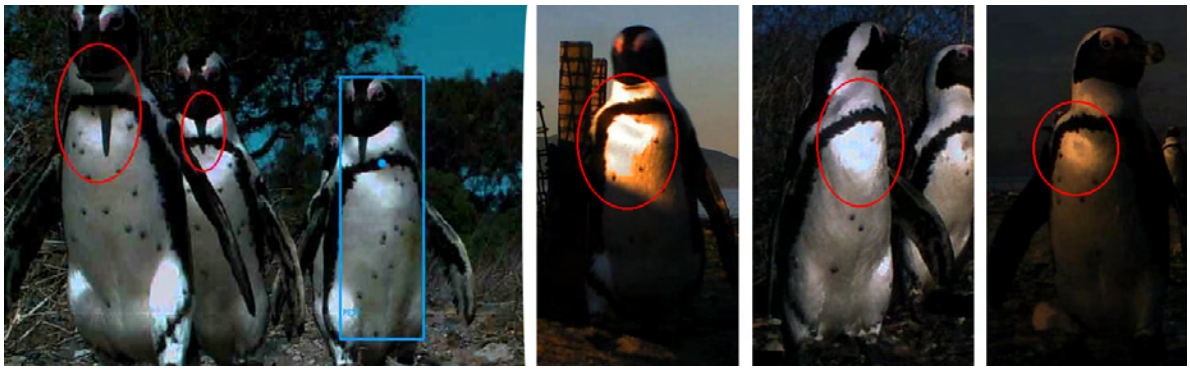


Figure 3.23: ***Lighting Conditions Inflict False Negatives.*** Shadows, specular reflections and over-exposure (all marked with red ellipses) form novel image textures which degenerate the visual properties of the object of interest, causing the detector to fail. [wildlife images used: I01]

3.5 Improving Performance, Localisation and Speed

3.5.1 Constructing Dense Belief Maps

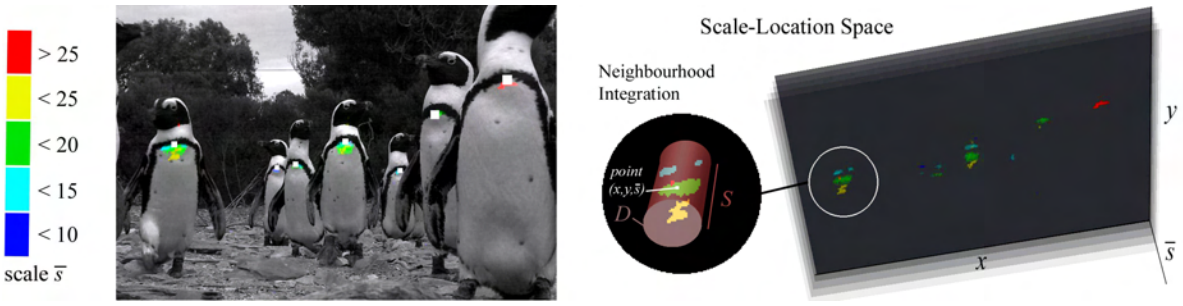
So far, appearance detections have merely been based on directly interpreting a binary evidence function $\mathcal{H}_\omega(\mathbf{x}, \bar{s}) : \mathcal{X} \times \bar{S} \rightarrow \{0 \equiv false, 1 \equiv true\}$ that holds suspected presences of key points ω in the space of image locations $\mathbf{x} \in \mathcal{X}$ and object magnifications $\bar{\mathbf{s}} \in \bar{S}$. In accordance with Viola and Jones' principle of attentional cascades [212], the function $\mathcal{H}_\omega$ is assembled dependent on a unanimous agreement of m booster-generated stage classifiers H_i:

(VIOLA AND JONES' CASCADE CLASSIFIER)

$$\mathcal{H}_\omega(\mathbf{x}, \bar{s}) \quad = \quad \underbrace{\bigwedge_{i=1}^{m} H_i}_{\text{attentional cascade}} \quad = \quad \bigwedge_{i=1}^{m} \underbrace{\left(\sum_{t=1}^{t=n_i} \mathbf{h}_t^i(\mathbf{I}, \mathbf{x}, \bar{s}) \right) \geq \beta_i}_{\text{linear stage classifier } H_i} \tag{3.7}$$

where $\mathbf{h}_t^i$ is the output of the t^{th} weak classifier of the i^{th} detection stage in the cascade, n_i is the number of weak classifiers and β_i is the acceptance threshold learned for the i^{th} stage, m is the number of stages in the detection cascade and $\mathbf{I}$ represents the image.

Figure 3.24(a) visualises a map $\mathcal{H}_\omega$ as scale-coloured superimpositions on the input image (red encodes close, blue distant object instances). It can be seen that key point estimations are widely dispersed around the ground truth (white blocks) in both scale and space.



(a) **(b)**

Figure 3.24: ***Likelihood Map $\mathcal{H}_\omega$ and Integration in Location-Scale Space.*** Classifier hits for key points ω are resolved as point clouds in the location-scale space $\mathcal{X} \times \bar{S}$. **(a)** Projection of classifier hits $\mathcal{H}_\omega$ as scale-coloured superimpositions onto the original luminance image. **(b)** Stacked plane plot of the underlying x-y-$\bar{s}$ space. The 5 bins of scale used are, again, colour-coded for better visibility. Note that classifier hits associated to a single object are dispersed in both scale and space, covering sub-volumes of the domain. It can be seen that the ground truth (white blocks in (a)) does not always occupy the centre of local density of the map $\mathcal{H}_\omega$ (for the particular image it lies above the density maxima due to elongated animal pose). Thus, a cylindrical integration of local evidence for maxima extraction only poorly reflects the spatial statistics of perturbations from the ground truth. *Tu et al.* [200], for instance, try to counter the problem by estimating real-valued detector outputs. However, their method ignores the spatial instability of the detector. [wildlife image: I01]

Hitherto, detections – that is density maxima of the $\mathcal{H}_\omega$-map – have been extracted by integration and maximisation over cylindrical volumes $D \times S$ around classifier hits (as performed by Intel's OpenCV implementation[23] [37] and visualised in Figure 3.24(b)). As shown, the procedure yields sufficient detection performance. However, it shows poor localisation in the spatial domain. Figure 3.25 illustrates the phenomenon on representative detections.

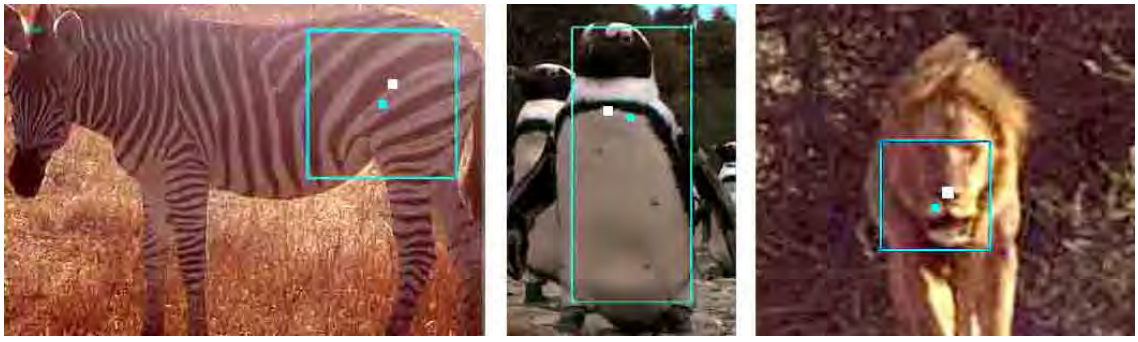


Figure 3.25: ***Poor Localisation.*** The images show detection examples (cyan blocks) that exhibit a strong spatial perturbation from the ground truth (white blocks). Such poor spatial localisation causes inaccurate registration, fuzzy pose estimation and shaky tracking. [images used: I01, I10, I02]

Aiming for a more realistic interpretation of the observation map $\mathcal{H}_\omega$ produced by the classifier, it is proposed to model explicitly a prior for the *spatial component* of the feature-specific perturbation of localisation, i.e. the specifics of the classifier's 'spatial fuzziness'.

A probability distribution $E_\omega : \mathbb{I}^2 \rightarrow [0, 1]$ is used to describe the prior. Given a key point ω, it is built to capture the likelihood of perturbation in the reference system centred at the ground truth. As exemplified in Figure 3.26, it is constructed experimentally by

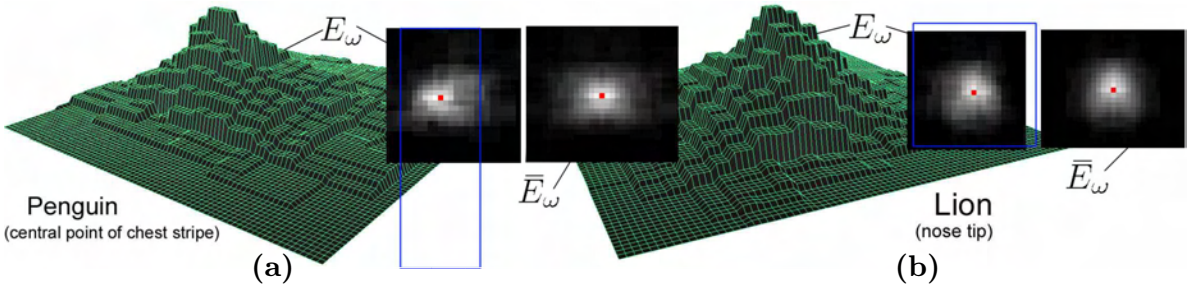


Figure 3.26: ***Localisation Prior.*** The illustrations show accumulations E_ω of classifier hits $\mathcal{H}_\omega$ with respect to the true key point (red point) for penguin chests in **(a)** and lions nose tips in **(b)**, each averaged over 400 training samples. Distributions are size-normalised with respect to the detector resolution superimposed in blue. Symmetricalised versions $\bar{E}_\omega$ are also shown. [images used: I01, I10]

[23] Following the cylinder paradigm, the detection posterior is defined as the local maxima of density, which is calculated for a point $(\mathbf{x},\bar{s})$ as the integral over binary evidence in a neighbourhood volume $D \times S$:, that is $\sum_{s_j \in S(\bar{s})} \sum_{\mathbf{x}_i \in D(\mathbf{x})} H_\omega(\mathbf{x}_i, s_j)$ where $D(\mathbf{x}) = \{\mathbf{x}_i \in \mathcal{X} \mid p \geq ||\mathbf{x} - \mathbf{x}_i||\}$ is the circular spatial neighbourhood, $S(\mathbf{x}) = \{s_j \in \bar{S} \mid q \geq ||1 - \bar{s}/s_j||\}$ is the linear magnification neighbourhood, $\mathbf{I}$ is the image, $||.||$ denotes the vector length and (p, q) are thresholds. Common values are $(p, q) = (\frac{s_1 s_2 s_j}{4}, 0.5)$ where $s_1 \times s_2$ is the detector's resolution [37].

accumulating and normalising the spatial component of binary estimates $\mathcal{H}_\omega$ over a population. In the penguin example, for instance, a clearly visible, horizontal anisotropy (see Figure 3.26(a)) indicates that the x-localisation of the top-centre chest point (along the near-horizontal chest stripe) is especially prone to error. The depicted function E_ω captures this spatial divergence of classifier hits from the true key point location.

Clearly, the function E_ω models the spatial spread of relevant measurements and can, therefore, be used as a kernel for estimating a dense detector function $\mathcal{P}_E(\omega|\mathbf{x},\mathbf{I})$ from the sparse, spatial classifier output $\mathcal{H}_\omega$ illustrated Figure 3.27(a). The dense map is built by convolving[24] scaled versions[25] of the spatial prior E_ω with the classifier evidence $\mathcal{H}_\omega$ yielding:

(DENSE DETECTION MAP)

$$\mathcal{P}_E(\omega|\mathbf{x},\mathbf{I}) = \underbrace{n}_{\text{normalisation}} \underbrace{\sum_{s_j\in\bar{S}}}_{\text{scales}} \underbrace{\sum_{\mathbf{x}_i\in\mathcal{X}}}_{\text{locations}} \underbrace{E_\omega(\mathbf{x}-s_j^{-1}\mathbf{x}_i)}_{\text{scaled spatial prior}} \underbrace{\mathcal{H}_\omega(\mathbf{x}_i,s_j)}_{\text{classifier evidence}} \tag{3.8}$$

where $\mathbf{x}$ is the image location to be evaluated, ω is the key point class under review, n is an empirical measure that ensures $\mathcal{P}_E \in [0,1]$, and $\mathbf{I}$ is the image with locations $\mathcal{X}$ and object magnifications $\bar{S}$. Applied over the entire image, the result is interpreted as an image again, a map that captures the overall *observation density* $\mathcal{P}_E(\omega|\mathcal{X},\mathbf{I})$ as illustrated Figure 3.27(b).

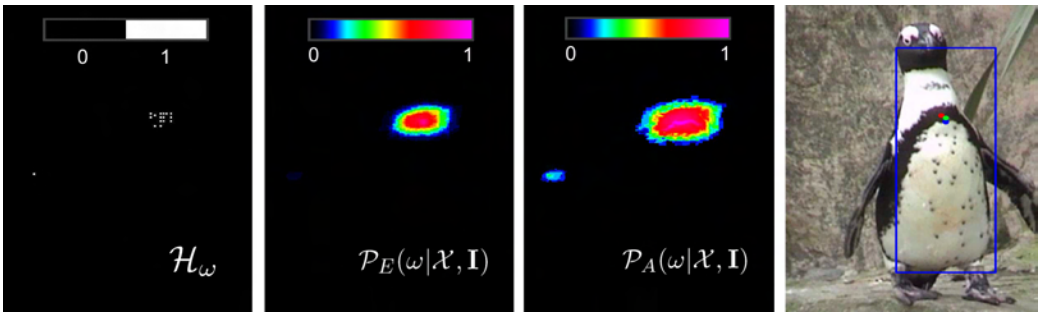


Figure 3.27: ***Observation Density and Detection Generation.*** The images illustrate different models for the observation density which, for the above sample, can be interpreted as the system's belief in the presence of a 'top-centre point of a penguin's chest stripe'. **(a)** Spatial component of point cloud $\mathcal{H}_\omega$ created by the point-surround version of the Viola-Jones' method sparsely applied to the image; **(b)** dense belief map $\mathcal{P}_E(\omega|\mathcal{X},\mathbf{I})$ created by convolving the spatial likelihood E_ω with the binary evidence $\mathcal{H}_\omega$; **(c)** gradient-supported belief map $\mathcal{P}_A(\omega|\mathcal{X},\mathbf{I})$; **(d)** Maxima suppression plus thresholding yield detections (in this case a single one) visualised for each of the three density models shown as a red, green and blue key point estimate. [wildlife image used for example: I01]

[24]Note that a normalised convolution [109] cannot be applied since the density of *binary evidence*, which would be lost in the normalisation process, encodes vital information on the presence of key points.

[25]The prior E_ω is scaled by the object magnification s_j since uncertainty is assumed to scale linearly with object size due to a constant, scale-independent classifier resolution. The use of symmetricalised kernels yielded worse performance indicating a natural asymmetry of the classifier.

The belief map can be fine-crafted further by using the high gradient property of key points as an independent source of localisation information. Weighting $\mathcal{P}_E(\omega|\mathcal{X},\mathbf{I})$ by a gradient map – interpreted as the likelihood of gradient presence – adds valuable fine structure (as shown in Figure 3.27(c)). An advanced observation density can, therefore, be modelled as:

$$\begin{aligned}&\text{(GRADIENT-SUPPORTED BELIEF MAP)}\\&\mathcal{P}_A(\omega|\mathbf{x},\mathbf{I}) = n\mathcal{P}_E(\omega|\mathbf{x},\mathbf{I})\ \Delta(G*\mathbf{I})\end{aligned} \tag{3.9}$$

where n is a normalisation factor ensuring $\mathcal{P}_A \in [0,1]$, $*$ denotes convolution, G is a Gaussian sized at half the dominant Turing wavelength, Δ is the derivative operator, $\mathbf{x}$ is the image location to be evaluated, ω represents the key point class under review and $\mathbf{I}$ is the image. Figure 3.28 visualises the map (green structures around the function's maxima) in some more complex detection scenarios – including pose variation, partial occlusion and deformation.

Note that, in contrast to the spatial localisation prior E_ω, the maps $\mathcal{P}_E$ and $\mathcal{P}_A$ are 'rankings of evidence' rather than a valid probability distribution. Thus, $\sum_{\mathbf{x}\in\mathcal{X}} \mathcal{P}_A(\omega|\mathbf{x},\mathbf{I}) = 1$ does *not* hold since the true number of underlying causes (e.g. how many penguins are present?) is unknown. Consequently, global normalisation cannot be applied without loss of information. Nevertheless, normalised *regions* of $\mathcal{P}_A(\omega|\mathcal{X},\mathbf{I})$ – explicitly known to contain exactly one key point – will be used as the spatial probability distribution of local evidence.

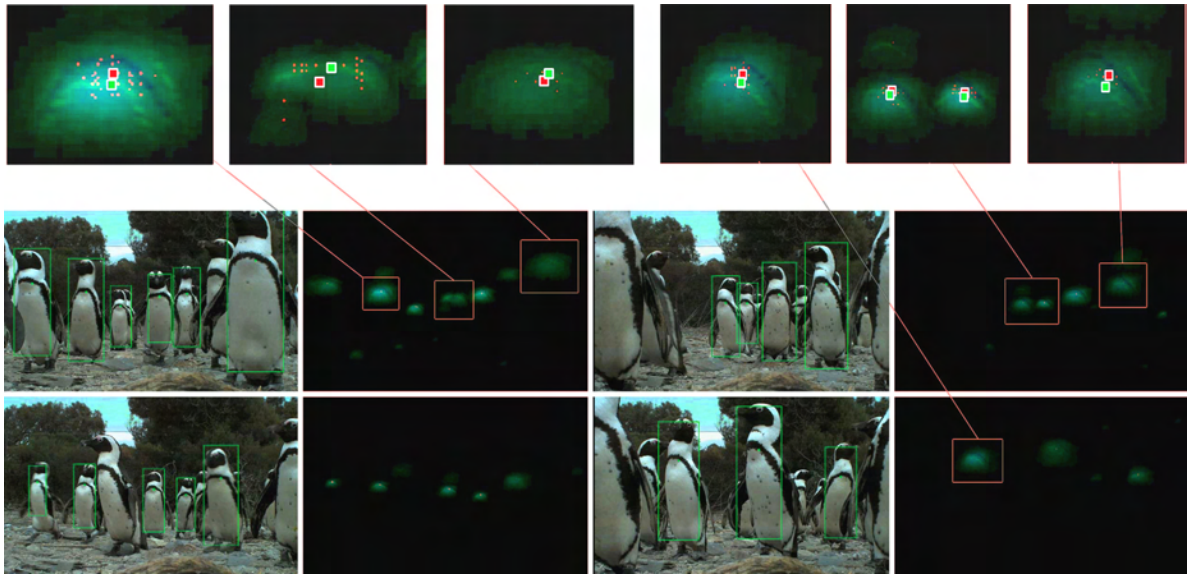


Figure 3.28: ***Gradient-supported Belief Maps.*** The images show the sparse binary classifier output $\mathcal{H}_\omega$ (red points), the dense belief maps $\mathcal{P}_E$ (blue) and the gradient-supported belief maps $\mathcal{P}_A$ (green) in different colour channels. Magnifications of regions (top) clearly reveal the fine structure added in the gradient-supported case. Functional maxima (green blocks/detections) are firmly located at high gradients – the cylinder method yields coarse estimates only (red blocks). [images: I01]

3.5.2 Quantitative Analysis of Localisation Improvements

As shown in the previous section, exploiting gradient information in conjunction with a spatial prior increases the localisation accuracy of the detector. In order to estimate the quantity of this improvement, the localisation error[26] is now estimated for sample key points on each of the three species (plains zebras, lions and African penguins) in focus.

Three different detection paradigms are compared: 1) the original Viola-Jones method based on bounding boxes, 2) point-surround detection of key points using cylindrical accumulation of evidence (see Section 3.5), and 3) detection of maxima of the gradient-supported belief map $\mathcal{P}_A$ as proposed.

The figure on the next page visualises the experimental results, comparing the average localisation error over 400 detections. It can be seen that, compared to the standard bounding box approach (dashed lines), point-surround description enforced during training (dotted lines) and spatial prior & gradient support applied during runtime (solid lines) significantly narrow the error band.

In particular, the standard deviation σ of the error distribution is decreased by 35 to 40 percent as shown in the table below.

The highest gain of accuracy achieved by using point-surround registration alone (shown in blue) can be observed at key points where, untypically for Turing patterns, stable corner measures sit on or close by the key point (lion's nose tip).

Data Set (Resolution)	**V:** σ	**K:** σ	$\mathcal{P}_A$**:** σ	**V:** μ	**K:** μ	$\mathcal{P}_A$**:** μ	**Decrease of** σ
Lion Face (16×16)	**5.13**	**3.47**	3.19	**3.96**	**2.56**	2.28	35.87%
Penguin Frontal (8×32)	5.44	**4.18**	**3.26**	4.38	**3.14**	**2.55**	40.07%
Zebra Hindquarters (16×16)	5.59	4.26	3.47	4.26	3.03	2.26	36.79%

Table 3.5: **Localisation Error.** The table shows the mean localisation error μ and the standard deviation σ of the distribution of the spatial detection error measured in pixel based on the Euclidean distance to the true location in a normalised detection. It can be seen that an improvement of 35 - 40% can be achieved by extending the bounding box approach suggested by Viola-Jones [212] (marked by 'V') to a dense, gradient-aided measure $\mathcal{P}_A$. Relying on only point-surround appearance for key points (K), that is without using a dense representation, yields improvements with respect to method (V) but fails to reach accuracy levels exhibited by method ($\mathcal{P}_A$). [image sets: I01, I03, I10].

[26] The localisation error – measured as the Euclidean distance to a labelled ground truth – is normalised for both the scale of the detected object and the resolution of the detection window. Thus, the error measure refers to the distance of detections from the ground truth at a fixed scale and resolution.

The positional stability of corners close to the centre of the surround descriptors enables weak learners – underpinning CART construction – to take advantage of their spatial distinctiveness.

On the other hand, strongly deforming patterns without the presence of stable corners (penguin centre point of chest stripe) can be stipulated more effectively using actually measured gradients (outlined in red), which prove more significant for improving localisation in this case.

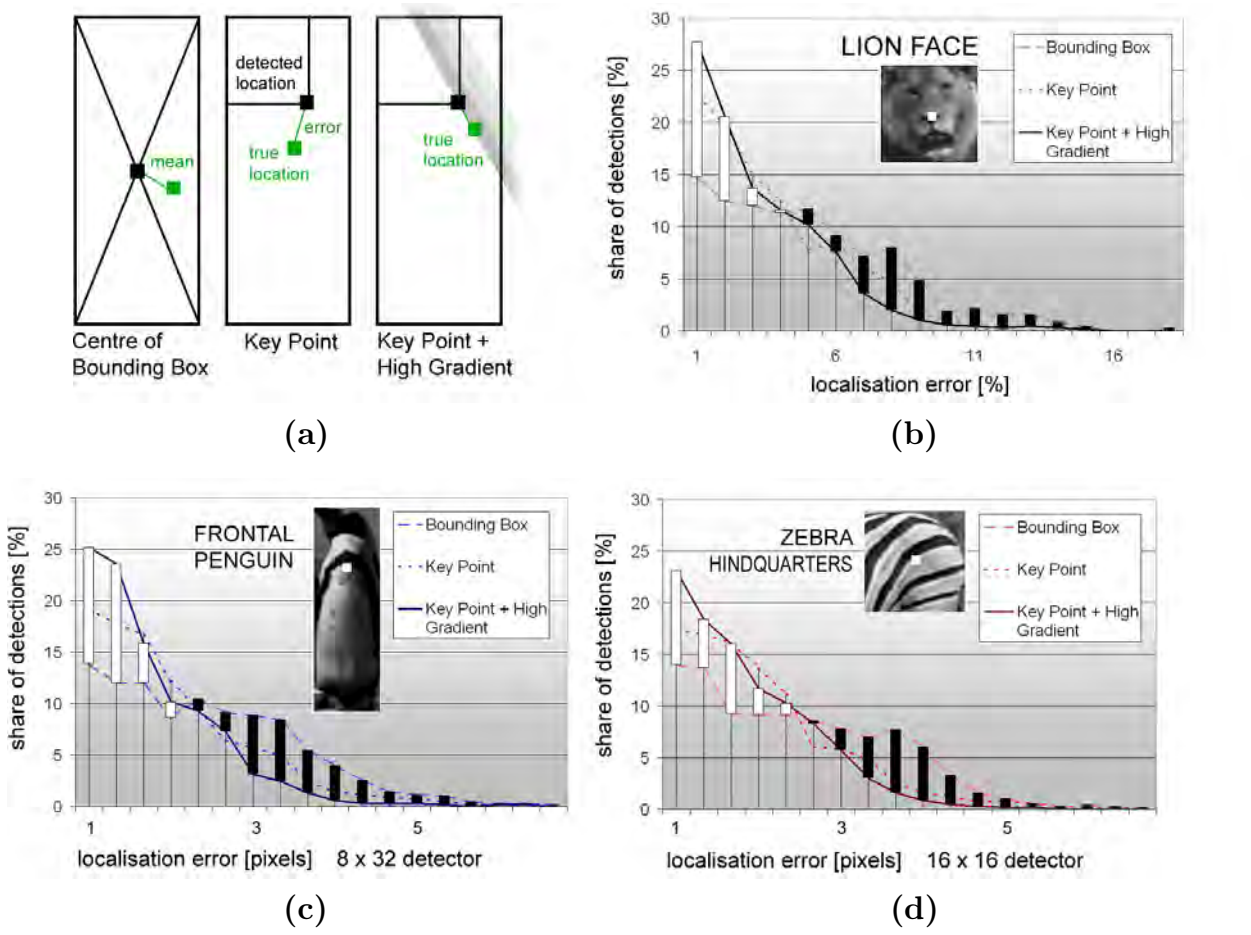


Figure 3.29: ***Localisation Accuracy.*** The localisation performance of the three different detector versions is illustrated with respect to a hand-labelled position of the key point. In the case of the bounding box approach, this ground truth is substituted by the mean over all detections. The localisation error is given in percent of the neighbourhood window size. **(a)** registration paradigms tested; **(b)-(d)** visualisations of the distribution of the localisation error in different sample sets.

3.5.3 Multi-Component Description by Sets of Belief Maps

A belief map $\mathcal{P}_A$ represents the detector evidence for the presence of a single key point class. Now, sets of belief maps are used to form a key point model $\mathfrak{B}$, describing a species' surface in a component-like fashion. Figure 3.30 visualises $\mathfrak{B}$ for African penguins.

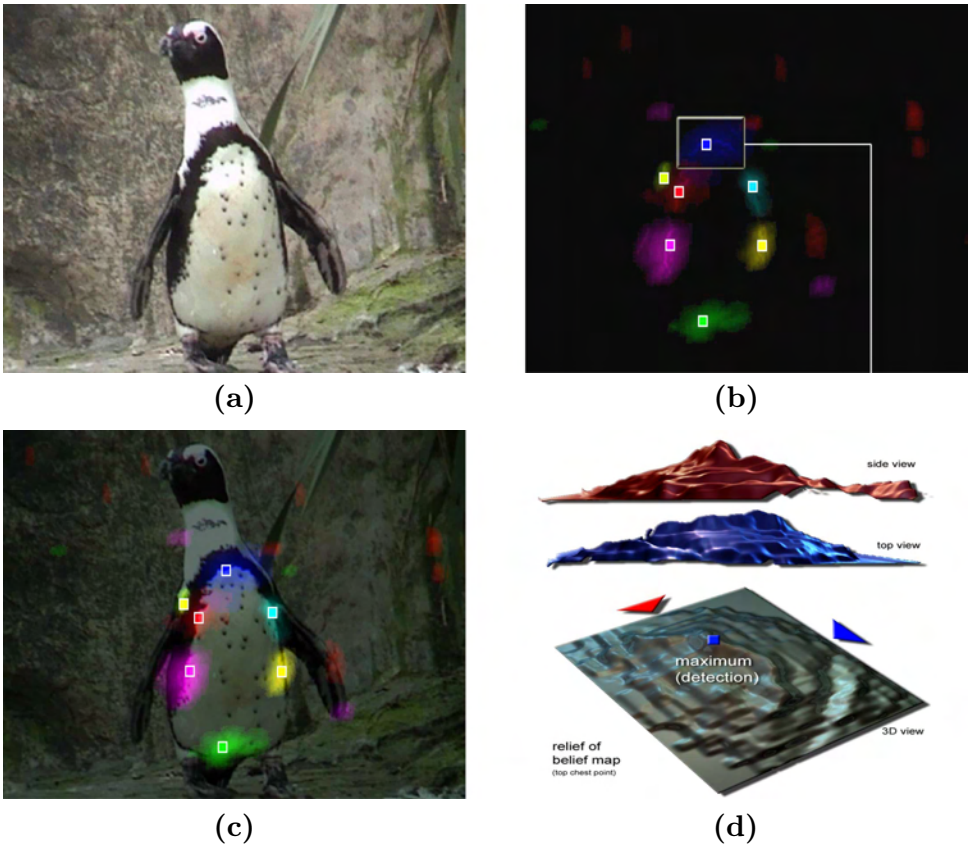


Figure 3.30: ***Component Description by Sets of Belief Maps.*** **(a)** original image; **(b)** model $\mathfrak{B}$ visualised by coloured superimpositions of its various belief maps; **(c)** original image superimposed on the belief maps; **(d)** 3D relief of the belief map representing the top-chest key point. Fine information added by gradients (peak line) can be seen clearly as sharp peaks. [images: I01]

Using this model $\mathfrak{B}$, the complexity of the space spanned by the object's presence and its different (in the above case: near-frontal) poses is compacted into a set of 2-dimensional image descriptions.

The model provides a solid foundation for fitting spatial models as described later in the theses. However, building belief maps for larger sets of key points decreases the execution speed of the detector. Yet close-to-realtime operation on high-resolution imagery is essential to avoid buffering terabytes worth of video during prolonged autonomous operation in wild habitats. The next sections focus on improving the runtime performance of the detector.

3.5.4 Blessing and Curse: Invariance from Exhaustive Search

Scale- and shift invariant species detection[27] is required to detect various animals in the field of view, that is to maximise detections in suitable poses for later fingerprinting.

Hitherto, the paradigm of 'exhaustive search' has been applied for detection as suggested by *Viola and Jones* [212]. That is the content has been parsed for the presence of key points by exhaustively scaling and shifting the neighbourhood window over the entire image, progressively creating values of the map $\mathcal{H}_\omega$ as shown in Figure 3.31(a)-(b).

Clearly, this approach comes at the cost of substantially growing computational efforts needed for the classification of windows when increasing the resolution of the image parsed. Given an image $\mathbf{I}$ of resolution $r_1 \times r_2$ and a neighbourhood window W of resolution or size $s_1 \times s_2$ pixels, there exist n potential search windows:

(FULL NUMBER OF DETECTION WINDOWS)

$$n = \underbrace{\sum_{\bar{s}\in\bar{S}}}_{\text{window scales}} \underbrace{\prod_{i=1}^{i=2}}_{\text{dimensions}} \underbrace{(r_i - \left\lfloor \overbrace{s_i}^{\text{base size}} \underbrace{\bar{s}}_{\text{scale}} \right\rfloor + 1)}_{\text{no. of windows at particular size and dimension}} \tag{3.10}$$

where $\bar{S} = \{1.2^0, 1.2^1, 1.2^2, ...\}$ is the set containing the stretch factors $\bar{s}$ chosen for scaling[28] the original neighbourhood window W over the image. Note that the selected set $\bar{S}$ uses exponential scaling to account for the drop in localisation accuracy in large scale-ups.

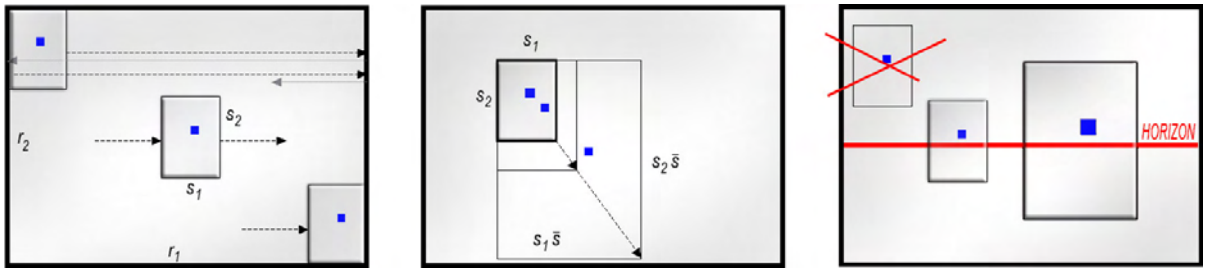


Figure 3.31: ***Intra-Image Search Dynamics.*** An image is probed by placing the neighbourhood window (portrait-sized rectangles) of a key point (blue blocks) at different scales and locations over the image. **(a)** The window is shifted over the image, **(b)** the window is scaled by $\bar{s}$ at each location, and, **(c)** dealing with walking animals of equal size and static cameras, it is suggested to restrict location-scale combinations to a subspace consistent with a weak perspective model.

[27] Methods that provide full rotation invariance of the detection [96] are neglected here since the animals observed maintain a near-vertical orientation throughout the process of locomotion.

[28] Naturally, all scaled windows are required to reside within the image area, satisfying $\forall_{\bar{s}\in\bar{S}} : s_i\bar{s} \leq r_i$.

3.5.5 Perspectively Constrained Search

To achieve close-to-realtime operation on a (single) detector despite a high computational matching cost, most applications [37] process only every other window in each image dimension, quartering the number n of effectively evaluated search windows.

It is proposed to further constrain the search space by exploiting the (approximately) equal proportions of the objects of interest, i.e. the members of the species. In particular, it is suggested to restrict combinations of location and scale to a perspectively feasible parameter subspace, assuming static cameras and animals walking on flat terrain.

Acquisition conditions are approximated by a horizontally pointing camera (see Figure 3.32(a)).

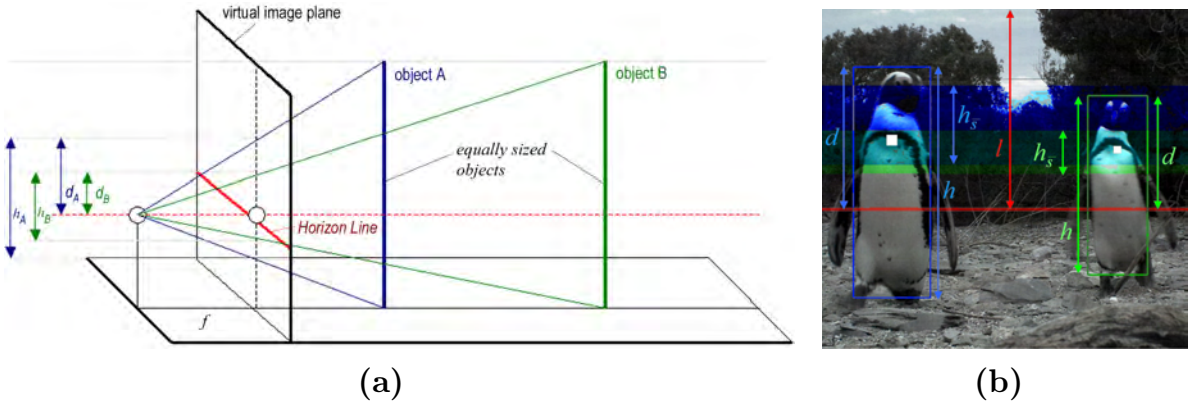


Figure 3.32: ***Perspective Constraint.*** **(a)** Given equally sized objects the ratio of object height h and distance d to the horizon line is a constant. **(b)** Calibration requires defining the horizon line l and *some* correct detection samples to establish an object-specific size parameter c. Applying the constraint in Eq.(4.10), the search space for a key point (white blocks) can be confined to a set of horizontal bands of heights $h_{\bar{s}}$ (coloured regions) where a band exists for each scale $\bar{s} \in \bar{S}$. The sum of all location-scale pairs in the $|\bar{S}|$ bands then spans the search space. [wildlife image: I01]

The size of an object is then bound to be proportional to the projected distance d between a fixed point on the object (e.g. the top point, a key point) and the horizon line, that is:

$$\begin{aligned} &\text{(PERSPECTIVE CONSTRAINT)} \\ &c(1-\varepsilon) < h/d < c(1+\varepsilon) \quad \text{since} \quad h_A/d_A \approx h_B/d_B \end{aligned} \tag{3.11}$$

holds for any two equally sized objects A and B, where $h = s_2\bar{s}$ is the projected object height measured in the image, c is an object-dependent constant and ε is a relaxant which allows for some degree of natural variance in the object size.

An application of the constraint effectively decreases the number of search windows to:

$$\text{(REDUCED NUMBER OF WINDOW CANDIDATES)}$$
$$n_{\text{reduced}} = \underbrace{\sum_{\bar{s}\in\bar{S}}}_{\text{scales}} \underbrace{h_{\bar{s}}}_{\text{vertical positions}} \underbrace{\lfloor r_1 - s_1\bar{s} + 1 \rfloor}_{\text{horizontal positions}} \tag{3.12}$$

where $h_{\bar{s}} = \min\left(l, \left\lfloor \frac{s_2\bar{s}}{c(1-\varepsilon)} \right\rfloor\right) - \left\lfloor \frac{s_2\bar{s}}{c(1+\varepsilon)} \right\rfloor$ is the height of a band of perspectively acceptable locations (see Figure 3.32(b)) for a specific scale $\bar{s}$, and l is the ordinate of the horizon in image space, and r_1, s_1 are used from Eq.(3.10). Thus, the weak perspective constraint reduces computational efforts by decreasing the cardinality of the search space[29]. Table 3.6 gives a quantitative overview of this reduction achieved by comparing the number of windows processed at different image resolutions with and without the constraint in place.

Image Resolution [pix]	**Full Window Count** n / $\lvert\bar{S}\rvert$	**Reduced Count** n_{reduced} / $\lvert\bar{S}\rvert$
160×120	87,085 / 9	17,567 / 9
320×240	591,467 / 13	83,634 / 13
640×400	2,643,770 / 16	305,152 / 16
1280×800	14,248,155 / 20	1,289,472 / 20

Table 3.6: **Search Space Confinement.** The table illustrates the count of search windows for different image resolutions before and after enforcing the weak perspective constraint. An 8×24 neighbourhood window (according to the penguin example) is used. The object-specific parameters are set to $c = 1.562$ and $\varepsilon = 0.2$. Scaling is performed employing sets $\bar{S} = \{1.2^0, 1.2^1, 1.2^2, ...\}$ where $|\bar{S}|$ gives the number of scales tested for a specific resolution.

3.5.6 Runtime Speed vs. Image Resolution

The Viola-Jones framework uses integral convolution in order to increase the speed of the process of feature extraction. It will be shown now that this technique proves especially effective in high-resolution scenarios.

In order to quantify the additional performance gain induced by integral convolution for high-resolution imagery, practical experiments on key point detectors are carried out, operating (frame-by-frame) on 2300 frames of a video sequence, quantised at different resolutions. The key point detector is then applied for comparing standard convolution with integral convolution, thus measuring the processing time used by the two methods.

Table 3.7 illustrates the results. Focussing on the last column, it can be seen that the relative speedup substantially increases with the image resolution $r = r_1 \times r_2$.

[29] As a side effect, the constraint also removes (potential) false detections and, therethrough, contributes to the robustness of the detector, approximately reducing false positive detections by a factor $n/n_{reduced}$.

Image Resolution $(r)/\lvert\bar{S}\rvert$	**Speed CC [fps]**	**Speed IC [fps]**	**Relative *Speedup(r)***
352×264(92,928)/13	5.74	39.09	6.81
400×300(120,000)/14	2.92	31.44	10.77
480×360(172,800)/15	1.01	22.99	22.76
640×480(307,200)/17	0.16	14.04	87.75
1280×800(1,024,000)/20	0.01	8.06	806

Table 3.7: **Speedup via Image Integration.** The table summarises experimental results comparing multi-scale conventional convolution (CC) in the frequency domain with multi-scale integral convolution (IC) for the frontal penguin detector (resolved at 8×24 pixels). An Intel DualCore T2300 running on 1GB of RAM is used for the experiment. It can be seen that the relative speedup increases with the resolution r used. Thus, the integral technique proves especially suitable for speeding up convolutions in scenarios of high resolution r and a large number of tested scales $|\bar{S}|$.

A polynomial approximation of the measured values in the domain spanned by speedup and resolution r shows a basic, quadratic relationship:

$$\begin{aligned} &\text{(RELATIONSHIP BETWEEN SPEEDUP AND RESOLUTION)} \\ &Speedup(r) = r(ar + b) + c \end{aligned} \tag{3.13}$$

where the parameters are measured as $(a, b, c) = (0.0018, -1.1886, 208.55)$ for the system setup employed. Figure 3.33 illustrates both the speedup (purple curve) and the underlying measurement series.

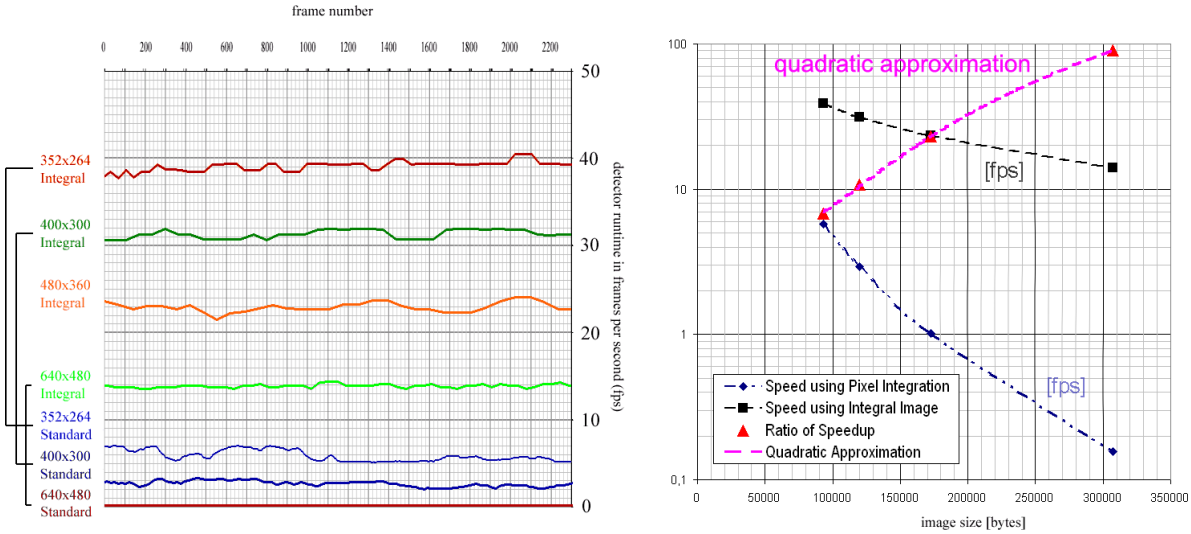


Figure 3.33: ***Integral Convolution, Image Resolution and Speed.*** Illustrations depict runtime characteristics of the frontal penguin detector with and without the use of integral convolution applied to video sequences of different image resolution (using an 2.3GHz Intel DualCore with 1GB of RAM). **(a)** frame rate for different video resolutions recorded over a sequence of 2300 frames; **(b)** Plotting the average performance of each method against the image sizes used and computing the relative speedup (red) reveals the suitability of image integration for the analysis for particularly large images. (Note that the weak perspective constraint is not applied in these measurements.)

While the exact parameterisation may change using different system configurations, the quadratic nature of the relationship is intrinsic to the image integration process. Thus, it can be concluded that the effects of integral convolution are most significant for high image resolutions, enabling detection at about 8 frames per second at megapixel scale.

3.5.7 Improved Runtime Results

A combined application of the weak perspective constraint and integral convolution can, as shown in Table 3.8, further increase the runtime benchmarks of the detector. It can be seen that, using both constraints in conjunction, a speed of about 8 frames per second can be maintained when applying a sequence of 3 detectors at 1280×800 pixels.

This gain in performance establishes the basis for an application of multi-pose detectors which are capable of covering a wide range of poses in close-to-realtime.

Image Resolution	**Speed IC [fps]**	**Speed WP+IC [fps]**	**3 Detectors WP+IC [fps]**
352×264	39.09	49.33	30.02
400×300	31.44	38.42	24.27
480×360	22.99	29.58	18.85
640×480	14.04	19.11	13.93
1280×800	8.06	12.24	8.17

Table 3.8: **Combined Runtime Performance.** The table summarises experimental results comparing the Viola-Jones method [212] of integral convolution (IC) and the proposed method of combining a weak perspective constraint with the integral convolution (WP+IC). An Intel DualCore T2300 running on 1GB of RAM are used for the experiments. All tests are conducted using 2,300 frames. Only every other detection window is checked in each dimension, effectively quartering the number of detection windows. Note that, for the multi detector scenario, the integral image is calculated only once and reused for each of the detectors. [wildlife images used for the experiment: I01]

3.6 Multi-Pose Appearance Detection

3.6.1 Generalisation Limitations in Single-Pose Detectors

Analysing the single-view detectors built in the previous chapter, it can be observed that poses covered by them exceed the subspace of pose spanned by the positive training images used for construction. Thus, the detection subspace is widened[30] to unseen, yet 'similar' visual configurations. The extent of this generalisation will now be quantified by measuring the detection confidence (reflected by the value of belief map $\mathcal{P}_A$ at the detection point) with respect to the location in pose space. Figure 3.34(a) documents the generalisation of one of the key point detectors for African penguins. The illustration plots the average detection confidence $\varnothing\mathcal{P}_A$ – collected from detections in a video sequence – against two hand-labelled parameters[31] of pose.

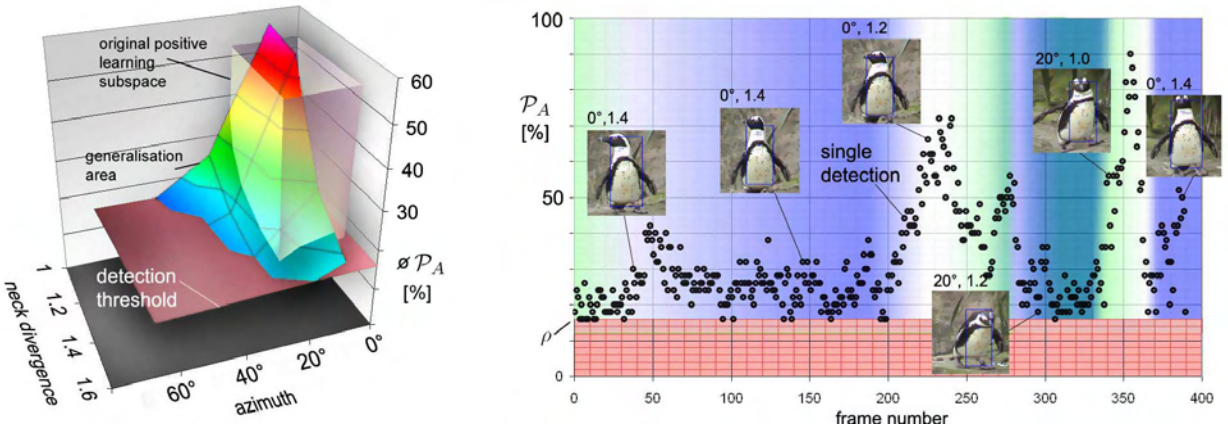


Figure 3.34: ***Pose Coverage of a Single Detector.*** **(a)** visualisation of the detector confidence (vertical axis) averaged over 2300 detections plotted against azimuth (4 bins resolution: first bin ranges from 0° to 20° etc.) and neck divergence (4 bins resolution); The detector used was originally trained on positive samples from a subspace indicated by a highlighted box. It can be seen that the detection area above the detection threshold (determined by the working area) exceeds the training space and, thus, generalisation occurs. **(b)** The graph visualises the distribution of detector confidence (black dots) over the first 400 frames of the test video. Overlayed images depict specific detections with annotations of azimuth and neck divergence. [images used for experiment: I01]

[30] It can be argued that AdaBoost's well documented generalisation [8, 111, 212, 211, 213] performance on novel data is the reason for a widened detection space, assuming that over-fitting can be avoided.

[31] The pose space is projected onto two parameters: 1) azimuth and 2) a measure that captures the extension/contraction of the animal's very flexible neck section with respect to its relaxed pose. For the task at hand, the azimuth is interpreted as the approximate angle between the direction faced by the animal and the camera axis. (Generally, the azimuth is defined as the angle between a reference plane and a point. Here, the plane is the one parallel to the image plane at the animal and the point is any point on the ray from the animal along its direction faced by the body.)

Two qualitative observations are of special interest for designing a multi-view detector. First, single-view key point detectors deliver fuzzy results with respect to pose parameters, that is an exact stipulation of a specific pose is not a straight forward process.

Second, key point classifiers are capable of covering an extended area that reaches outside the original training space. Aiming for a robust operation over natural animal populations, can a coverage of large sets of poses be achieved by *sparse* combinations of a sparse set of key point detectors?

3.6.2 Designing a Multi-Pose Architecture

A number of architectures have been suggested for implementing multi-pose detection. Common arrangements of multiple detectors of small pose coverage or weak precision in a multi-view framework include: 1) parallel cascades (arrays) [221], 2) trees [96, 56], and 3) hierarchical cascades (pyramids) [125]. Figure 3.35 visualises the designs mentioned.

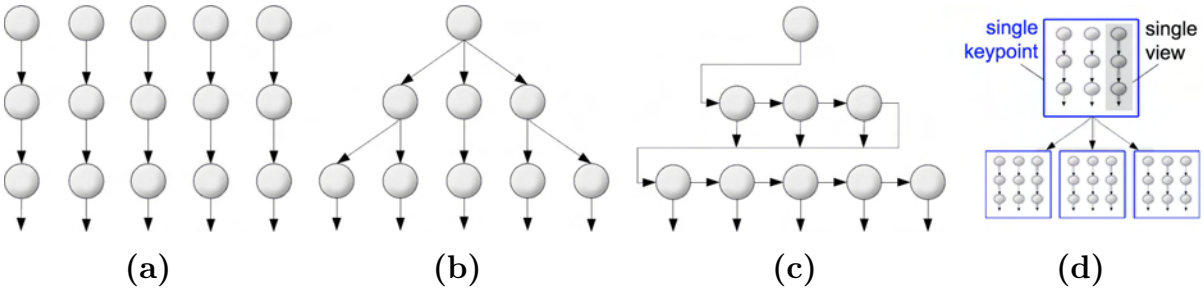


Figure 3.35: ***Selection of Architectures for Multi-pose Detection.*** Circles represent classifiers and arrows indicate the sequence of application. **(a)** parallel array of classifiers, that is a set of independently operating cascades; **(b)** tree structure, which enforces rigorous coarse-to-fine coverage starting with an 'overall object detector' at the root; While the approach promotes good pose resolution at leafs, the root classifier has to cover the *full* pose space; **(c)** pyramid structure, that is a hierarchy of cascades; **(d)** The hybrid structure proposed in this work: multiple cascades of single-view classifiers model a key point while several key points (which are connected in a tree-based, geometrical model as described in the next chapter) explain the pose of the object.

A number of recent publications, e.g. work by *Huang et al.* [96] or *Everingham et al.* [56], argue in favour of tree structures in order to minimise the computational cost of detection. The tree architecture implements a rigorous coarse-to-fine recognition concept. It promotes an especially rapid processing since the tree depth places an upper bound on the number of classifiers tested during runtime (along a branch from root to leaf). The technique has, nevertheless, a major bottle-neck that proves – as will be shown now – critical for the case of complex content classes: the strong classifiers situated in high-up nodes, foremost the root itself, are exposed to modelling very large fractions of the pose space.

Describing a wider range of poses, however, increases the variance – and thereby the complexity – of the associated class of sample images to be modelled. Figure 3.36 exemplifies this correlation between the spectrum of pose covered and the variance found in (manually registered[32]) image sets of lion faces for three different spectra of azimuth.

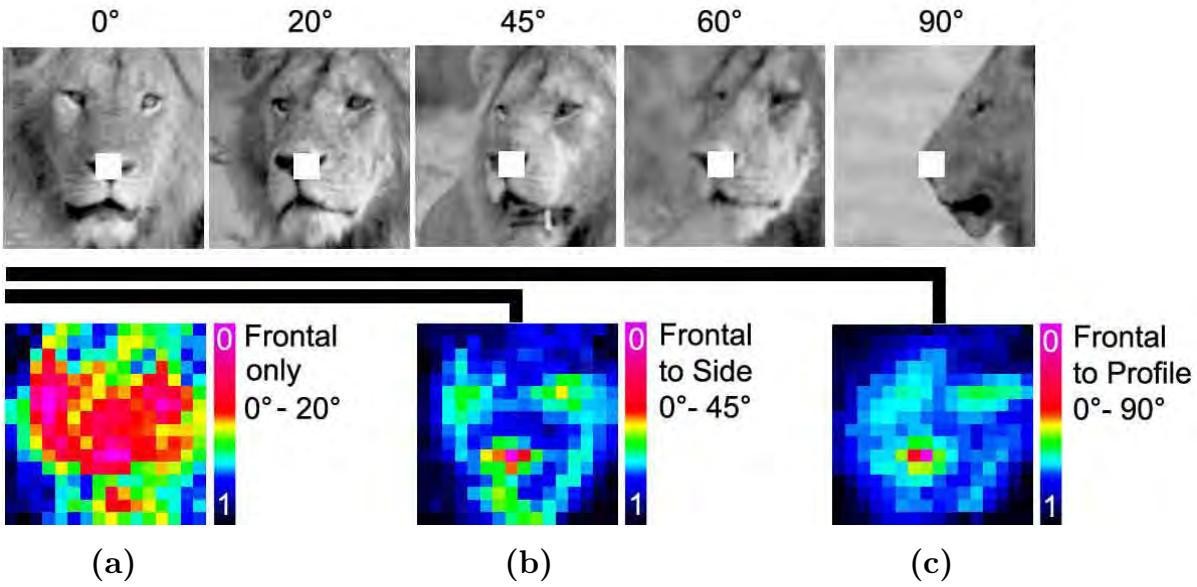


Figure 3.36: ***Pose Coverage vs. Image Variance.*** The images visualise the standard deviation of the z-Score normalised luminance function in the region of the lion's nose tip (indicated by white squares in top row). No variance is shown as 0 (purple) while 1 (black) represents the maximum deviation observed at any location ($\sigma(\mathbf{x}) = 46.87$ pixels). For each of the three azimuth ranges studied the variance is estimated based on 200 images resolved at 16×16 pixels sampled from that specific azimuth range. In general, it can be seen that the variance increases with the spectrum of pose covered. **(a)** variance plot constructed using near frontal views of lion faces only; **(b)** variance plot for samples ranging from azimuth 0°- 45°; **(c)** variance plot for azimuth ranges from frontal to full left profile (0°- 90°). [wildlife images used for experiment: I10]

Thus, classifiers trained on more variable data sets are required to model an increasingly complex distribution in pattern space. Classifiers for recognising poses of single objects (as, for instance, described in *Everingham et al.* [56]) can often cope with this extra degree of complexity.

However, the generalisation behaviour of classifiers that model entire object classes can, as it has been recognised in the literature, become '*somewhat unstable*' (quote from [96]) when large fractions of the pose space are monolithically modelled.

[32]Manual labelling ensured an accurate registration at the nose tip and a scaling to a normalised object size/resolution. Images were also normalised for variable lighting using the z-Score approach.

3.6.3 Performance Costs of Coverage

In order to quantify this 'instability' for the task at hand, classifiers are trained on lion samples covering images from different azimuth spectra. As illustrated in Figure 3.37 and shown in Table 3.9, the performance of classifiers drops significantly when trained to cover the full frontal-to-profile spectrum.

However, the performance decrease in the category frontal-to-side is, compared to frontal views, infinitesimal. This observation can be interpreted as resulting from the qualitative difference of structural changes occurring: most features, e.g. eyes, jaw etc., undergo a mere perspective transform in the frontal-to-side spectrum while some of the features disappear completely in *some* examples of the category frontal-to-profile due to occlusion.

Key point classifiers can accommodate linear transform during learning simply by adjusting the sizes of Haar-like features. Accounting for occlusions, in contrast, demands a number of complex adaptations in the tree classifiers (CARTs) that, similar to the avoidance of pruning trees, limits the potential for generalisation.

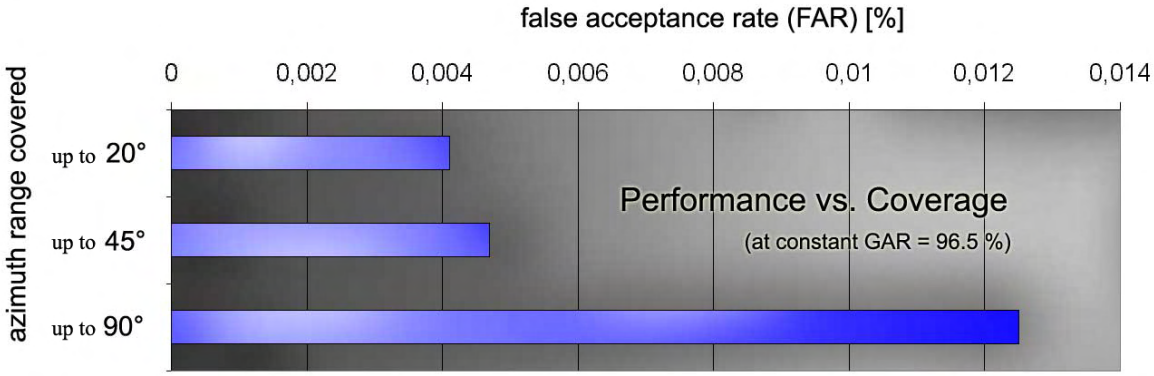


Figure 3.37: ***Detector Performance vs. Azimuth Range Covered.*** The graph depicts classifier performance for different ranges of pose-space coverage. Fixing the genuine acceptance rate (GAR), the false acceptance rate (FAR) is plotted against the azimuth range captured. It can be seen that the false acceptance rate increases significantly when covering the full frontal-to-profile spectrum. The increase for the frontal-to-side category compared to frontal detection is, on the other hand, infinitesimal small. [wildlife imagery used for the experiment: I10]

Species/Image Source (key point described)	**Front up to** 20° **(GAR/FAR** [%]**)**	**Front-Side up to** 45° **(GAR/FAR** [%]**)**	**Front-Profile up to** 90° **(GAR/FAR** [%]**)**
Lions/I10 (nose tip)	96.5/0.0041	96.5/0.0047	96.5/0.0125

Table 3.9: **Detector Performance vs. Azimuth Range Covered.** The table shows measured performance characteristics of boosted key point classifiers for the three different ranges of azimuth investigated. Results are averaged over three-fold cross validation using – in each of the 3 runs – 200 positive samples for training and 400 positive samples (plus negatives) for generating the landmarks.

A fast operating tree-based detector would require the presence of a monolithic classifier at root level, covering the full pose space.

Since it has been shown that such a wide band classifier comes with considerable performance trade-offs a more conservative approach will be pursued: *detector arrays* will be used for achieving a wider coverage of the pose space.

3.6.4 Covering Pose Space using Detector Arrays

What fraction of the pose space is to be covered? – Clearly, different recognition applications have different demands. For instance, when aiming for individual identification, only poses that (fully) expose the biometric entity are relevant. For identifying African penguins the azimuth spectrum can, therefore, be limited to $-30°$ to $30°$ since the chest pattern is fully visible only within this range.

On the other hand, behaviour detection in lions demands a maximum degree of pose coverage. However, according to the results presented in Figure 3.37, a *single* lion detector should cover significantly less than $90°$ azimuth range.

In close approximation to this result, a frontal-side-profile description[33] is constructed to cover the profile-to-profile range. A total of 5 narrow band detectors are trained for the same key point (nose tip) whereas, exploiting facial symmetry, 2 of them are derived by feature mirroring.

During learning the working areas of narrow band detectors are chosen such that a common genuine acceptance rate (GAR) is reached. Table 3.10 illustrates the performance of the three classifiers that are contained in the frontal-side-profile array.

	Front $-20°$ **to** $20°$	**Side** $20°$ **to** $60°$	**Profile** $60°$ **to** $100°$
Positive Image Set [I10]	1,200	1,200	1,200
GAR [%] (3-fold cross-validated)	96.5	96.5	96.5
FAR [%] (3-fold cross-validated)	0.0041	0.0056	0.0083

Table 3.10: **Training and Performance of Detectors Recruited.** The table shows measured performance characteristics of boosted key point classifiers for the three different classifiers (columns). Results are averaged over three-fold cross validation using – in each of the 3 runs – 800 of the 1200 positive samples for training and 400 positive samples (plus negatives) for generating the landmarks. [wildlife imagery used for the experiment: I10]

[33] This form of description is a classic frequently applied in human face detection [37].

Note that, instead of having one belief map per key point, each narrow band detector now produces a separate estimate. Figure 3.38 visualises detections resulting from an application of the frontal-side-profile array to a video sequence of a lion turning sideways.

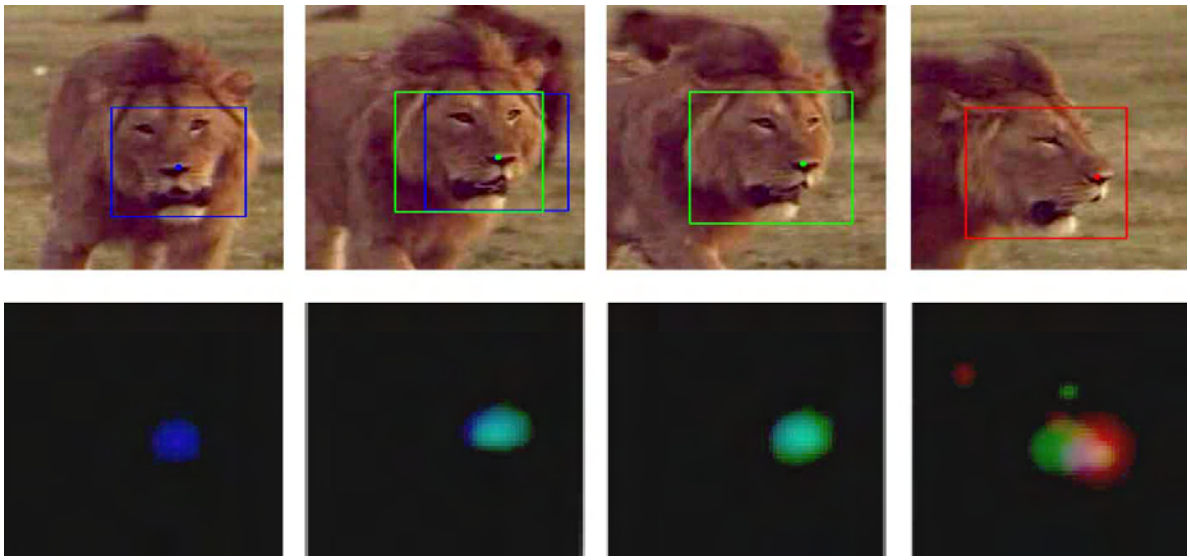


Figure 3.38: ***Frontal-Side-Profile Detection.*** The top row images show example detections resulting from the parallel application of frontal (blue), side (green) and profile (red) detectors. The bottom row visualises the according sets of belief maps $\mathcal{P}_A$. The three separate RGB colour channels visualise the belief maps of respectively colour-coded detectors. Note that the key point locations of different narrow band detectors vary relative to the surrounding neighbourhood window. [wildlife imagery used for the experiment: I10]

3.6.5 Linearly Integrated Multi-Pose Detector

False positive rates (FAR) vary for different classifiers. The profile detector, for instance, shows the highest false positive rate, indicating that profile images of lion faces contain the least class-characteristic set of features.

When integrating diverse narrow band detectors into a single recognition framework there is a need for scaling the classifiers in order to normalise the performance differences found. It is proposed to use the likelihood ratios between two detectors as a relative, linear scaling constant α. Picking one detector as the reference (here: the frontal view indexed with 0), a weighting scalar α_i is calculated as the ratio of false acceptance rates at constant GAR:

$$\text{(LINEAR SCALING FACTOR)} \qquad \alpha_i = \frac{\mathsf{GAR}/\mathsf{FAR}_0}{\mathsf{GAR}/\mathsf{FAR}_i} = \frac{\mathsf{FAR}_0}{\mathsf{FAR}_i} \tag{3.14}$$

The measure reflects the performance of the i^{th} classifier compared to the reference classifier ($i = 0$). Table 3.11 shows the resulting values for the lion example.

	Front $-20°$**to** $20°$	**Side** $20°$ **to** $60°$	**Profile** $60°$ **to** $100°$
FAR [%] (3-fold cross-validated)	0.0041	0.0056	0.0083
Linear Scaling Factor	$\alpha_0 = 1.00$	$\alpha_0 = 0.73$	$\alpha_0 = 0.49$

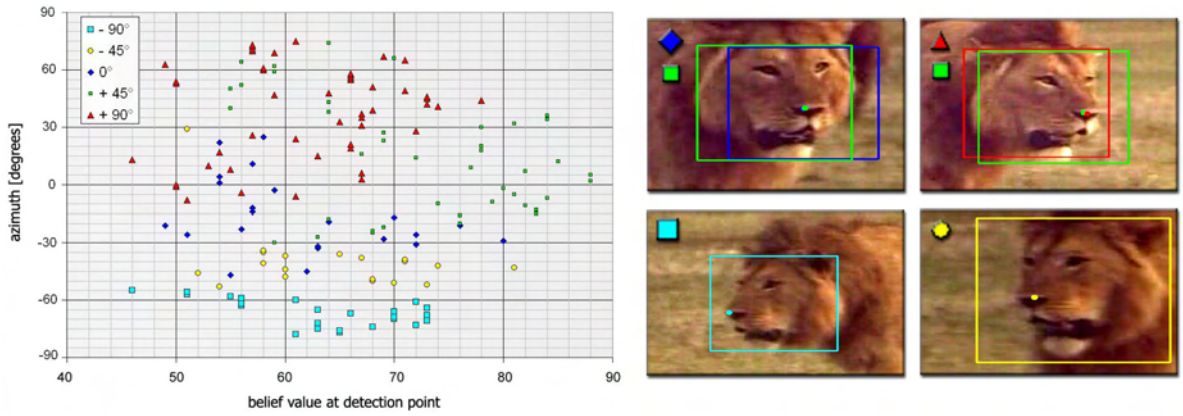
Table 3.11: **Linear Scaling Factors.** The table shows linear scaling factors α_i with respect to the frontal detector as reference. [wildlife imagery used for the experiment: I10]

Using the constructed set of linear weights, a set of n narrow band detectors can now be merged into a single detection space (captured in one, pose-invariant belief map $\mathcal{P}_\omega$) by element-wise integration of their weighted belief maps $\mathcal{P}_A^i$:

$$\begin{aligned} &\text{(LINEARLY INTEGRATED DETECTOR)} \\ &\mathcal{P}_\omega(\mathbf{x}) = \tfrac{1}{\bar{\alpha}} \textstyle\sum_{i=1}^{n} \left(\alpha_i \mathcal{P}_A^i(\mathbf{x})\right) \quad \text{where} \quad \bar{\alpha} = \textstyle\sum_{i=1}^{n} \alpha_i \end{aligned} \tag{3.15}$$

and $\mathbf{x} \in \mathcal{X}$ represents the image location. The α_i are the weights reflecting the relative reliability of classifiers. Intuitively, $\mathcal{P}_\omega$ is patched together by a set of narrow band detectors that cover various areas along one dimension in pose space, i.e. the azimuth. However, regions of response are only fuzzily defined for the narrow band detectors contained.

Figure 3.39 visualises the recognition activity of the 5 narrow band detectors over a video sequence of a lion changing direction. The graph plots the detection confidence (i.e. the value of the belief map at the detection point) against the azimuth. Clearly, there exists a high degree of overlap between detectors in the profile-to-profile spectrum. This phenomenon of '*pose blur*' results from the fact that positive samples are not crossed during training, e.g. side views are not part of the negative training set for frontals. Thus, classifiers disambiguate habitat patches only, permitting a generalisation over different lion poses.



(a) **(b)**

Figure 3.39: ***Pose Blur in Detector Arrays.*** **(a)** The graph plots the confidence of single narrow band detectors (disambiguated by different colours/shapes) against azimuth over a 10 second video sequence. **(b)** Images show a selection of corresponding detections. [imagery used for experiment: I10]

The actual calculation of $\mathcal{P}_\omega$ can be optimised by factoring out components common to all narrow band detectors. Dissecting Eq. (3.15) by using Eq. (3.9) reveals the inner mechanics of the recognition process:

(DETECTOR NOTATION EXPANDED)

$$\mathcal{P}_\omega(\mathbf{x}) = \tfrac{1}{\bar{\alpha}} \sum_{i=1}^{n} \left(\alpha_i \underbrace{\overbrace{\mathcal{P}_E^i(\omega|\mathbf{x},\mathbf{I})}^{\text{dense belief map}} \quad \overbrace{\Delta(G*\mathbf{I})}^{\text{gradient map}}}_{\text{gradient-supported map } \mathcal{P}_A^i(\mathbf{x})} \right) \tag{3.16}$$

where $\mathbf{I}$ is the image, G is a Gaussian kernel, $*$ denotes convolution and Δ is the derivative operator. The gradient map is clearly independent from the detector index i and can, therefore, be factored out yielding:

(OPTIMISED APPEARANCE DETECTOR)

$$\mathcal{P}_\omega(\mathbf{x}) = \tfrac{\Delta(G*\mathbf{I})}{\bar{\alpha}} \sum_{i=1}^{n} \alpha_i \, \mathcal{P}_E^i(\omega|\mathbf{x},\mathbf{I}) \tag{3.17}$$

The above calculation suggests to integrate the dense belief maps $\mathcal{P}_E^i$ of the narrow band detectors first before *multiplying once* the resulting map with the noise-filtered gradient map of the image. Figure 3.40 illustrates the optimised construction process.

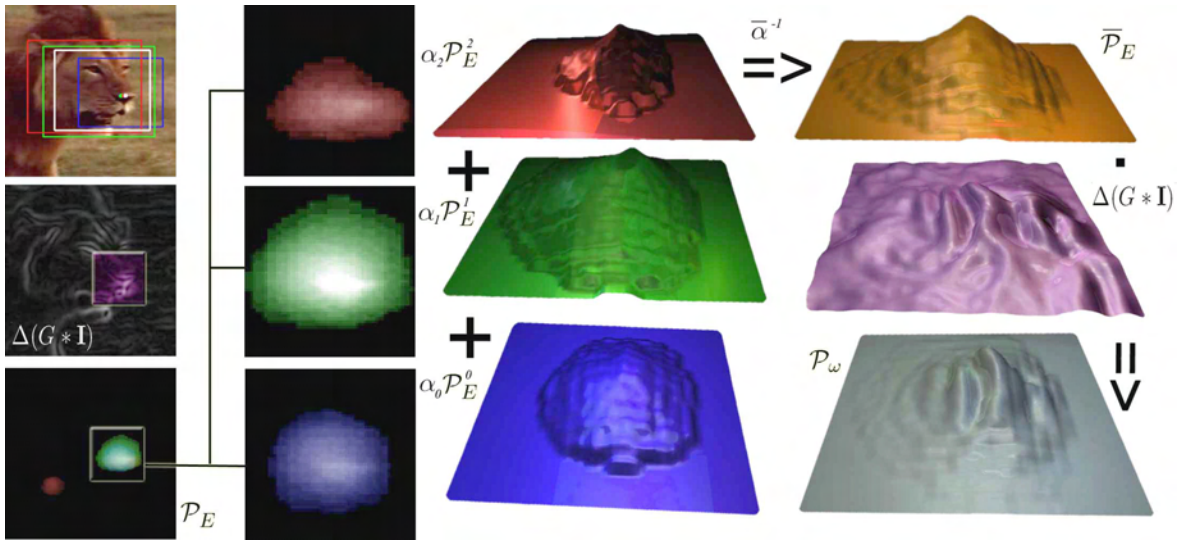
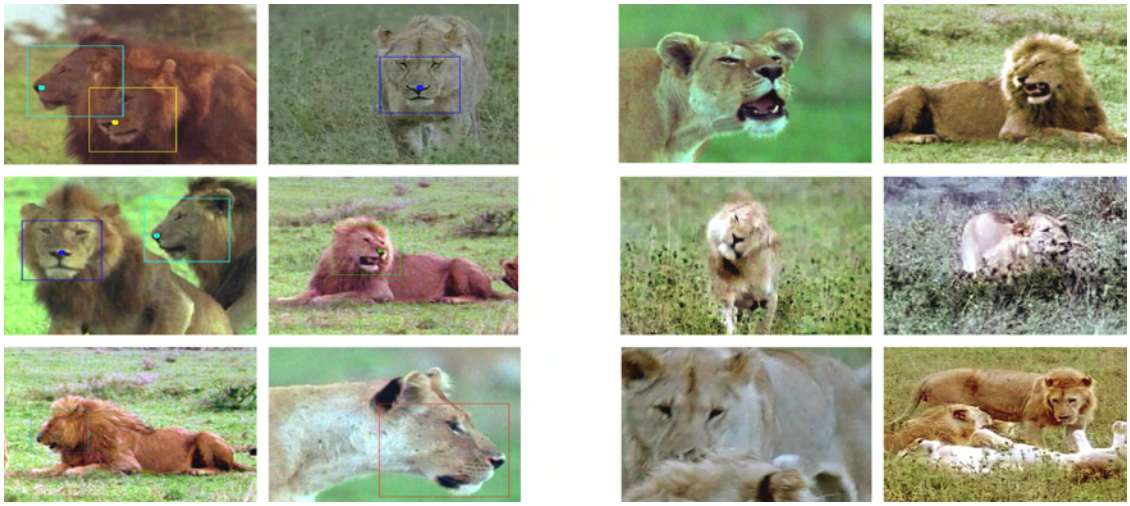


Figure 3.40: ***Wide Pose Detector for Key Points.*** **(a)** original image with detections of 3 narrow band classifiers (RGB) and the wide pose detection (white), derivative map, and multi-channel dense belief map capturing the three $\mathcal{P}_E^i$; **(b)** the three belief maps visualised as 3D surfaces; **(c)** The weighted, normalised sum of the belief maps yields an intermediate image $\bar{\mathcal{P}}_E$. Element-wise multiplication with the gradient map produces the final detection map $\mathcal{P}_\omega$. [wildlife imagery used for experiment: I10]

3.6.6 Virtues and Limitations of Detector Arrays

The detector arrays built cover a significantly wider fraction of the pose space compared to single view classifiers. Since generalisation occurs in all the learned detectors, the resulting array coverage extends far, allowing to track of key points despite natural variance from learned views over several frames in standing, walking, running and turning sequences.

Nevertheless, the array fails to detect species members whenever an animal's projected appearance departs somewhat significantly from the poses trained. Such scenarios occur during severe body deformation (e.g. roaring in lions, bending in penguins), significant rotational variance (e.g. lion rolling on the ground), change of proportions (e.g. penguin stretching) or partial occlusion (e.g. high grass partially covering a lion's face). Figure 3.41 illustrates the coverage and limitations of detector arrays on representative lion examples.



(a) **(b)**

Figure 3.41: ***Coverage and Limitations of the Multi-pose Lion Detector.*** **(a)** The images illustrate the range of robust applicability of the detector array. The detected key points (thresholded maxima of $\mathcal{P}_\omega$) are visualised as coloured dots where the neighbourhood windows and the colouration are selected according to which single view classifier is dominant. **(b)** The frames shown contain false negatives, that is unrecognised lions. The failure to detect the instances is due to untrained facial expressions, facial rotation, partial occlusion or facial tilt. [wildlife images used: I10]

3.7 Chapter Summary and Outlook

During the course of this chapter an approach to describing the visual *appearance* of a species has been discussed. The strategy proposed (see Figure 3.42) explains a species by modelling a set of key reference locations within a framework that expands on the basic detection methodology put forward by Viola and Jones [212]. In particular, the learning model (bootstrapped boosting of point-surround descriptors) has been confirmed as success-

ful in extracting visual features typical of a species despite a considerable degree of intra-population variance. The resulting sets of detectors have been tested on real-world animal populations. It has been noted that the detectors effectively cover a wide spectrum of

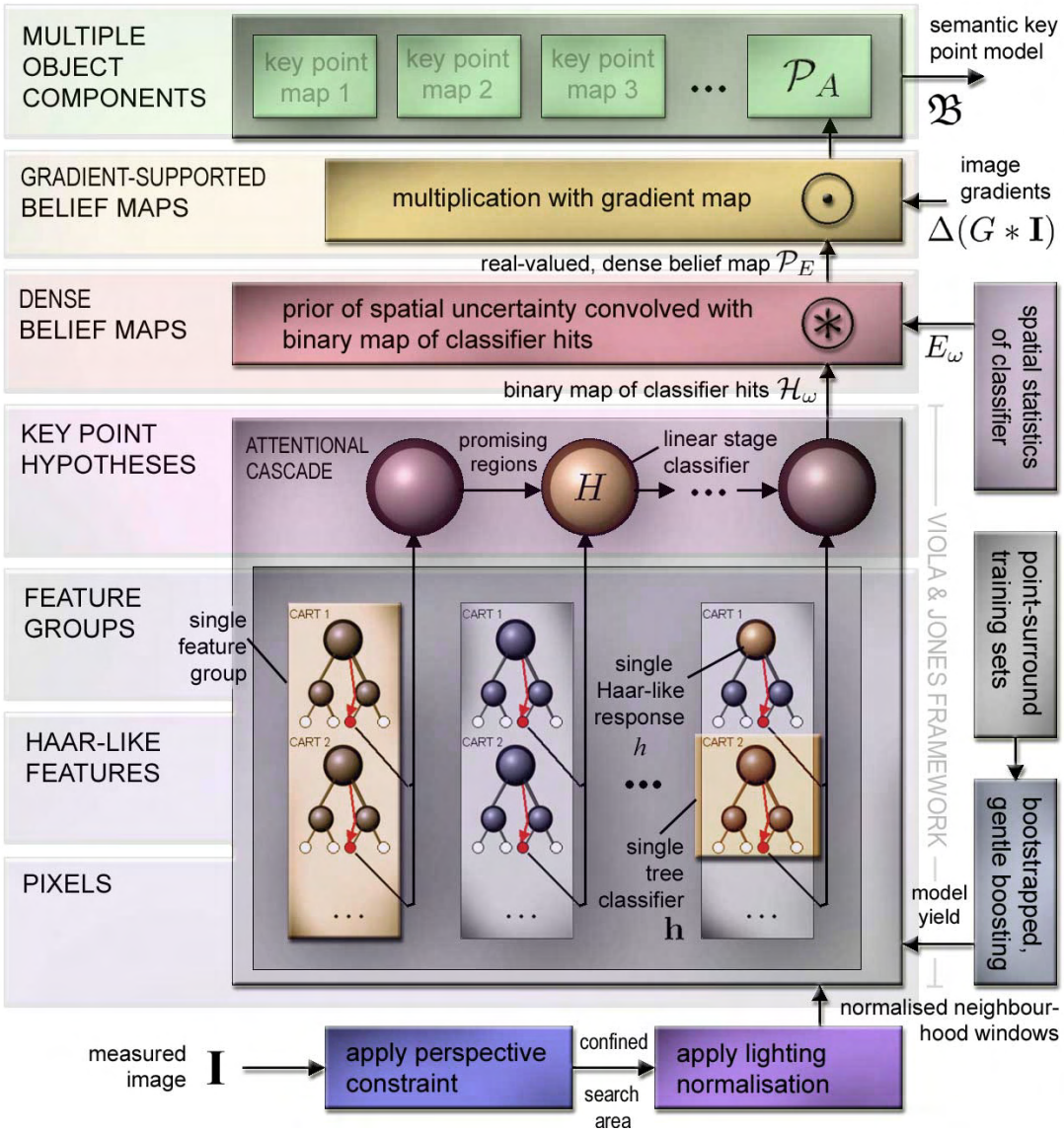


Figure 3.42: ***Summary of the Framework of Semantic Key Points.*** The chart explains the flow of information in the single-view detection framework proposed.

individual pattern variation within a species under various lighting conditions. In addition, methods for predicting the object-related size, proportions and resolution of detectors have been proposed. It has been demonstrated that the detectors are highly efficient for the task at hand: they compactly encode Turing patterns, show supreme localisation properties, and can be executed in close-to-realtime on high-res images. Overall, it has been shown that the cross-validated classifier performance is similar to state-of-the-art detectors for human faces.

However, clear limitations of single detectors have been disclosed, including high susceptibility to deep shadows, cryptic resemblance, occlusion, specular reflection, and coarse pose changes. The latter problem class was then tackled using multi-pose detector arrays to significantly widen the pose space covered.

When stepping aside and interpreting the model shown in Figure 3.42 through the eyes of *Gestalt theory* (e.g. summarised in [65, p.305]), the representation of a species can be understood as a hierarchy of visual wholes (the 'Gestalten'): responses h of *Haar-like features* around a key point are first combined to shape *complex features* coded as tree classifiers $\mathbf{h}$. Sets of them form *feature groups* coded as linear stage classifiers H, which are arranged in attentional cascades to produce a *point-surround classifier* $\mathcal{H}_\omega$ engineered further into a belief map $\mathcal{P}_A$. Finally, an object is described by a set $\mathfrak{B}$ of various *belief maps*. The approach constitutes a practical example of actually modelling unifying concepts (the 'Gestaltqualitäten'), which link elements in a visual hierarchy of growing complexity to form ever higher-level components that finally provide evidence on a 'believed presence' of objects. Overall, the 'Gestalt'-rule of *familiar configuration* forms the governing backbone for generating the links of the hierarchy: all Gestaltqualitäten are statistically learned as population-stable *similarities* of appearance in some *proximate neighbourhood*.

Closing the frame of this chapter by referring back to its opening quote, the model mechanics constructed also agree with *David Marr's* 'computational model of visual representation' [134]: the key point model expands 'primal sketches' (that is basic features such as block features and edges) to '2D sketches' (key point classifiers) which are input to more complex classifiers.

It follows Chapter 4 in which the paradigm of *familiar configuration* will be exploited once again – however, this time in order to group the extracted key points and to associate them with instances of single animals in the scene. This will allow for the extraction and normalisation of surface textures in order to identify and compare their *individually characteristic* coat pattern information. ■

Chapter 4

EXTRACTION OF COAT TEXTURES BY FITTING SPATIAL MODELS

'The essence of truth reveals itself as freedom.' [92]

(Martin Heidegger, 1889 - 1976)

4.1 Chapter Overview

The following chapter is dedicated to the extraction and perspective normalisation of texture patches from regions of the animal surface that are of biometric interest for individual identification.

Grouping Key Points. In the first part of the chapter a technique for grouping the previously extracted key points is described. All key points belonging to an individual animal instance are grouped into a common entity. A viewer-centred, spatial model structure termed *feature prediction tree* is utilised to create these associations based on learned, species-typical spatial configurations of key points.

Rigid Model Fitting. Each group of key points is then interpreted as an anchor set on a *single individual's* surface, providing a *correspondence set*, i.e. a sparse mapping between particular image locations and key surface positions on the species they represent. The correspondence sets are used to control the texture extraction where rigid surface models are fitted to the object instances by solving least square terms that approximate the projection parameters according to the correspondences of an individual. In this way the surface of the animal is approximated by a representation that provides a *dense mapping* from image locations to the species surface of interest.

Back-Projection. In a final step, this dense relationship is used to extract a pose-corrected texture by back-projection onto the model and re-rendering the model in a standard pose.

4.2 Grouping of Key Points and Association to Animal Instances

4.2.1 Interpreting Key Point Evidence

So far, the description of animals has been given as a component-like representation by key points that reflect the appearance evidence for particular surface points on animal coats.

However, scenes often contain multiple animals, so different instances of the *same* key point class (e.g. penguin top-centre chest) have to be associated to different animal instances. Figure 4.1 exemplifies such a multi-object scenario where the appearance-based key point evidence is visualised in the form of gradient-supported belief maps as discussed in the previous chapter.

(a) **(b)**

Figure 4.1: ***Key Point Evidence.*** **(a)** The frame shows a complex scene containing multiple penguins. **(b)** The visualisation illustrates the key point evidence extracted from the scene. All eight key point classes trained for penguins are depicted by superimposition of the eight gradient-supported belief maps $\mathcal{P}_A$ where each key point class is represented by a different colour (see model penguin). Detections of the most reliable 'master' key point, i.e. the centre of the top chest stripe, are visualised by blue points. This key point will be used as a root anchor to guide the exploration of further evidence in order to avoid an expensive extraction of the full evidence map for all key point classes over the entire image space. [wildlife images: I01]

This section will focus on searching for and grouping together key points that are spatially configured according to characteristic poses[1] typical of the species, thus, likely to represent one individual, Most essentially, a model structure is required to determine the degree of spatial fit between an encountered spatial configuration of key point evidence in the image, as exemplified in Figure 4.1(b), and key point configurations that represent valid species poses.

[1] As discussed in the previous chapter and illustrated in Figure 3.39, the employed statistical *multi-pose detectors* only *coarsely approximate* pose and are inherently restricted to the pose dimensions they resolve. Thus, a geometrical reconstruction from multiple key points is pursued.

Similar to pictoral structures [62], it is proposed to measure the cost for grouping a set of key points by considering 1) the matching costs for each key point based on appearance (as shown in Figure 4.1(b)), and 2) the deformation costs for fitting each key point with respect to its position in a spatial model. The extraction procedure for appearance fit has been described in the previous chapter (see Eq.(3.9)). A spatial model that explains valid configurations of sets of key points is left to be modelled.

4.2.2 Flexibly Linked Affine Domains and Tree Search

A number of different strategies have been suggested in the literature in order to explain a spatial configuration by *relations between object components* (represented here by the classes of key points). For instance, pictoral structures [60] define a network (usually a tree shape) composed of *pairwise relationships*[2] whilst other representations attempt to encode a location by considering *all* other locations in some neighbourhood [28].

For the task at hand it is proposed to consider only a subset of other locations and capture the structure of spatial component configurations of an object in a *chain of affine relationships* between the components. The chain is built from constraints. In each constraint three known key point locations $(\bar{\mathbf{x}}_m, \bar{\mathbf{x}}_n, \bar{\mathbf{x}}_o)$ are used to provide an affine reference system to describe the location of a 4^{th} point $\mathbf{x}_i$. Thus, given a set $\bar{\mathbf{X}}$ of key point locations, a constraint for the spatial position of another key point $\mathbf{x}_i \notin \bar{\mathbf{X}}$ is modelled by:

1) three key point identities $(\bar{\mathbf{x}}_m \in \bar{\mathbf{X}}, \bar{\mathbf{x}}_n \in \bar{\mathbf{X}}, \bar{\mathbf{x}}_o \in \bar{\mathbf{X}})$ that span an affine domain,
2) a position $\bar{\mathbf{a}}$ in this domain that predicts the position of $\mathbf{x}_i$, and
3) a Gaussian G that models the prediction uncertainty based on
the expected deviation of $\mathbf{x}_i$ from the position $\bar{\mathbf{a}}$.

The concept of building a chain of component relations by expanding the set $\bar{\mathbf{X}}$ of key points via iterative addition of predicted points $\mathbf{x}_i$ is schematically illustrated in Figure 4.2(a). The positional uncertainty (represented by ellipses) of predicting a key point $\mathbf{x}_i$ in an affine domain is given as the value of a Gaussian $\mathcal{P}_S^i = G(\bar{\mathbf{a}} - f_{(\bar{\mathbf{x}}_m, \bar{\mathbf{x}}_n, \bar{\mathbf{x}}_o)}(\mathbf{x}_i))$ where f is the projection of $\mathbf{x}_i$ into the affine system.

[2] Eq (2.15) describes the pairwise relationships in pictoral structures by cost measures d.

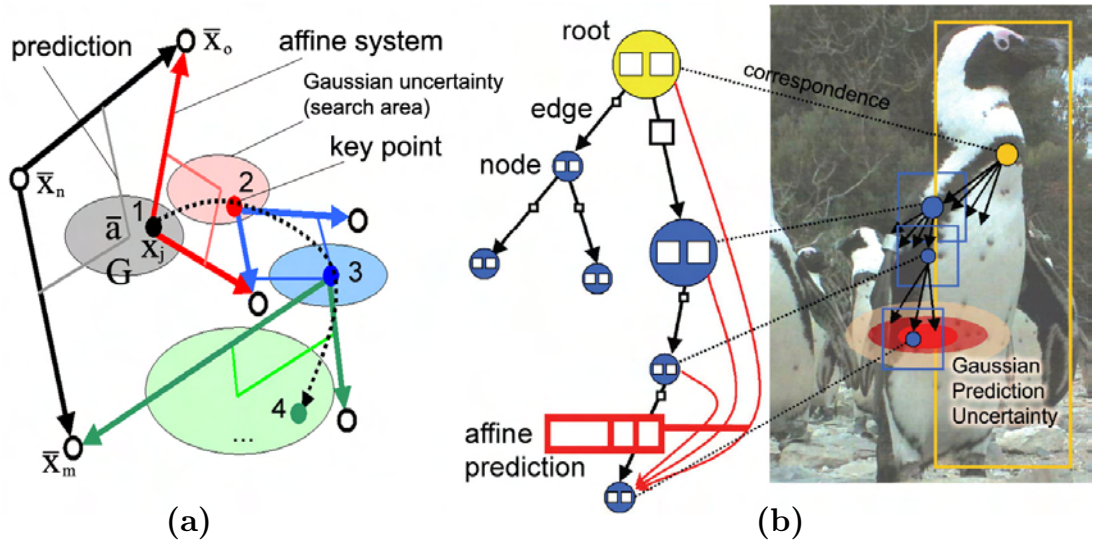


Figure 4.2: ***Structure of Feature Prediction Trees.*** **(a)** Illustration of the concept of object description by flexibly linking affine spaces. **(b)** Schematic illustration of a feature prediction tree. [wildlife image: I01]

For each suspected object instance a progressive search for further key points is modelled along the branches of a tree (see Figure 4.2(b)) starting at the root. Depth-first search with backtracking in this *feature prediction tree* (FPT) allows to explore the quality of fit for different configurations of key points along the most promising paths. Note that this strategy models the *process of image exploration* as a sequence of prediction steps: the tree structure is learned *offline* and encodes the most stable search passes through the key point set of an object, so that at each point during a later *online* search the tree can provide information on *which* key point to search for next and *where* to search for it.

4.2.3 Components of a Feature Prediction Tree

A particular key point – for penguins the centre of the top chest stripe, in general, the most reliable key point which can be seen in *all* poses of interest – forms the tree roots, initial *seed points* that are stipulated based on pure appearance information. The associated key point detector is executed over the entire image (exhaustive search) and all detected instances of this class are used as a root for a separate tree, each of which represents a potential animal detection.

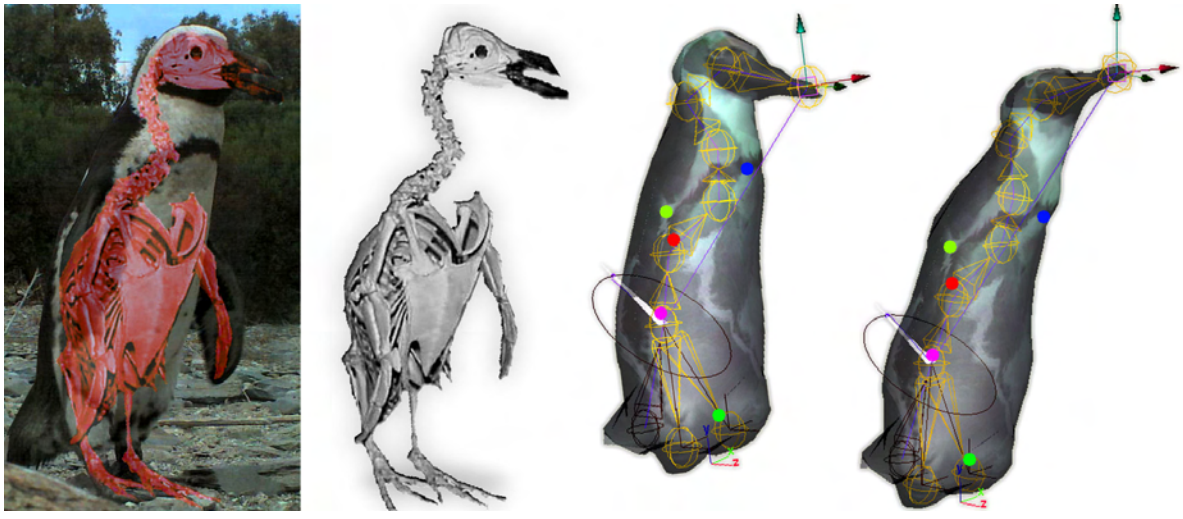
Starting from this root node, the tree models possible pathways for interpreting a key point configuration as an object instance in a flexible, sequential manner. The different components of an FPT are schematically depicted in Figure 4.2(b). While the tree structure (i.e. the edges) is learned offline the nodes hold data that is evaluated online:

The tree components can be understood as follows:

- **edge:** holds a key point *class* i to search for and an affine constraint with parameters $\bar{\mathbf{a}}$, Σ and (m, n, o) where the latter is a subset of feature classes in nodes on the path to the root;
- **root:** holds the position of an initial key point found via exhaustive search;
- **node:** holds (only if evaluated) the *predicted* location of a feature instance $\mathbf{x}_i$ of the node specific class and the model evidence L calculated for the path from node to root;
- **leaf:** represents a particular configuration of key points, which is defined by the key points along the path to the root;

4.2.4 Gathering Structural Sample Data from Animation

The tree structure is learned from sets of typical poses. In order to create necessary training information, that is valid configurations of key points, a deformable 3D model (see Figure 4.3) carrying $N = 8$ surface key points $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1, ..., \tilde{\mathbf{x}}_i \in \mathbb{R}^3, ..., \tilde{\mathbf{x}}_N]$ around the chest area was animated through T animation steps. The model was observed through $H = 3$ different virtual cameras placed in front of the model to register different view aspects in near-frontal poses as covered by the key point detectors.



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

Figure 4.3: ***3D Model used for Training.*** **(a)** an African penguin with its skeleton superimposed; **(b)** penguin skeleton, note the double 'S' of the spine; **(c)** 3D model used, the double 'S' is modelled using a virtual skeleton (yellow); **(d)** stretching the spine yields a significant neck extension.

During animation the cameras generated a set of $M = HT$ model projections, hence $M \times N$ projections of model key points $\tilde{\mathbf{x}}_i$ were registered. These observations were stacked into a matrix $\mathbf{X} \in \mathfrak{M}_{M \times N}$ where each row $[\mathbf{x}_{i1}, ..., \mathbf{x}_{ij} \in \mathbb{R}^2, ..., \mathbf{x}_{iM}]$ contained the measured 2D vertex coordinates of one of the N key points. Columns, indexed by j, described a particular configuration of key points which will be referred to as a *formation*.

To preserve information about the occlusion of key points by the object itself, a second matrix $\mathbf{V} \in \mathfrak{M}_{M \times N}$ with elements $v_{ij} \in \{0 = \text{invisible}, 1 = \text{visible}\}$ was saved holding values that indicate the visibility of features $\tilde{\mathbf{x}}_i$ in different formations.

4.2.5 Tree Construction

For each node of the tree let F describe the set of key point classes associated to the edges on the path from the node to the root and let $\bar{\mathbf{X}}$ hold their hypothesised locations. In order to expand the tree (initially just the root) the most predictable key point class i is found based on three measures:

First: Can the key point be seen?

The co-visibility $\mathcal{P}_{Vis}(i)$ of the class i with respect to F is calculated as the fraction of possible formations that contain a key point of class i:

(PROBABILITY OF CO-VISIBILITY)

$$\mathcal{P}_{Vis}(i) = \frac{|\overbrace{\{\ j \mid (v_{ij} = 1) \wedge (\forall_{k \in F} : (v_{kj} = 1))\ \}}^{\text{formations where all } k \in F \text{ and } i \text{ are visible}}|}{|\underbrace{\{\ j \mid \forall_{k \in F} : (v_{kj} = 1)\ \}}_{\text{formations where all } k \in F \text{ visible}}|} \tag{4.1}$$

where $|.|$ is the set cardinality. Rare key points (only visible in a few poses) are ranked low.

Second: **Can the key point be reliably found by its appearance?**

The relative success rate of the appearance classifier of the class i is of interest, calculated as:

(APPEARANCE STABILITY)

$$\mathcal{P}_{App}(i) = \frac{\mathsf{GAR}/\mathsf{FAR}_0}{\mathsf{GAR}/\mathsf{FAR}_i} = \frac{\mathsf{FAR}_0}{\mathsf{FAR}_i} \tag{4.2}$$

where FAR_0 is the false accept rate of the best performing key point detector, FAR_i is the false accept rate of detector associated to class i, and all FAR's are measured at a constant genuine accept rate.

Third: **Can the position of the key point be reliably predicted?**

The most stable affine system for a prediction of the key point class i given F is determined by calculating the spatial *precision* of the estimation in the training data.

Let $P = \{\mathbf{p}\} = \{(m, n, o)\}$ be the set of all triples from the set F. Let $V_i^{\mathbf{p}}$ be the set of indices of formations for which all key point classes m, n, o and i are visible. and let $\mathbf{a}_{ij}^{\mathbf{p}}$ be the location of $\mathbf{x}_{ij}$ projected into the affine system $(\mathbf{x}_{mj}, \mathbf{x}_{nj}, \mathbf{x}_{oj})$ given by the observed formation j. Then for each triple $\mathbf{p} = (m, n, o)$ the variance $\Sigma_i^{\mathbf{p}}$ over all predictions $\mathbf{a}_{ij}^{\mathbf{p}}$ observed during training is:

(PRECISION OF SPATIAL PREDICTION)

$$\Sigma_i^{\mathbf{p}} = \tfrac{1}{|V_i^{\mathbf{p}}|} \underbrace{\sum_{j \in V_i^{\mathbf{p}}}}_{\text{formations}} (\mathbf{a}_{ij}^{\mathbf{p}} - \mu_i^{\mathbf{p}})^2 \qquad \text{where} \qquad \mu_i^{\mathbf{p}} = \tfrac{1}{|V_i^{\mathbf{p}}|} \underbrace{\sum_{j \in V_i^{\mathbf{p}}}}_{\text{formations}} \mathbf{a}_{ij}. \tag{4.3}$$

The affine system $\mathbf{p}$ with the smallest variance shows the highest precision of spatial prediction. This variance is chosen to represent the affine localisation potential:

(LOCALISATION PRECISION)

$$\mathcal{P}_{Loc}(i) = \min_{\mathbf{p}}(\Sigma_i^{\mathbf{p}}) \tag{4.4}$$

The measure is associated to an affine constraint with the parameters $\hat{\mathbf{p}} = \arg\min_{\mathbf{p}}(\Sigma_i^{\mathbf{p}})$, $\bar{\mathbf{a}} = \mu_i^{\hat{\mathbf{p}}}$, and $\Sigma_i^{\hat{\mathbf{p}}}$.

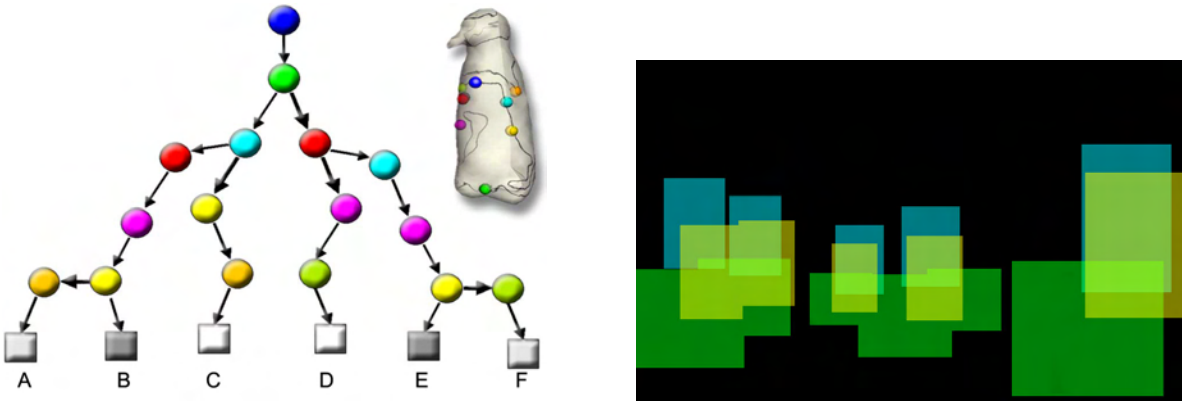
The 'predictive potential' for a feature class i given a set F of detected feature classes is finally assembled from the three factors:

(PREDICTIVE POTENTIAL)

$$\mathcal{P}_{Pot}(i) \;=\; \underbrace{\mathcal{P}_{Vis}(i)}_{\text{visibility}} \quad \underbrace{\mathcal{P}_{App}(i)}_{\text{recognisability}} \quad \underbrace{\mathcal{P}_{Loc}(i)}_{\text{localisability}} \tag{4.5}$$

By evaluating $\mathcal{P}_{Pot}$ for all classes $i \notin F$ the key point classes are ranked. Using this ranking, edges (i.e. constraints) are subsequently added for top ranking classes until these classes cover all formations that are possibly visible given F.

The procedure of adding edges is stopped if $\mathcal{P}_{Pot} = 0$ or all visible features are determined, in which case a leaf is added and the branch is closed. Figure 4.4(a) illustrates the FPT built for penguins in the form of a colour-coded tree where node colours are chosen according to the feature class they represent.



(a) **(b)**

Figure 4.4: ***Example of a Feature Prediction Tree.*** **(a)** Schematic view of the FPT built for penguins. Nodes are colour-coded and represent key point classes according to colourations of the model visualised. The strength of arrows indicate the ranking of the predictive potential of edges. Note that two leafs turned out to represent frontal views (dark grey) and the other four leafs represent different near-frontal (light grey) and side views (white). **(b)** Visualisation of the reduction in search area in image space due to a *guided* feature extraction for three different key points. Instead of full-size evaluation the appearance evidence is calculated in the predicted area only (bounding box around the Gaussian estimate within 2 standard deviations + overlap effects due to the area of influence of spatial priors) [wildlife images: I01]

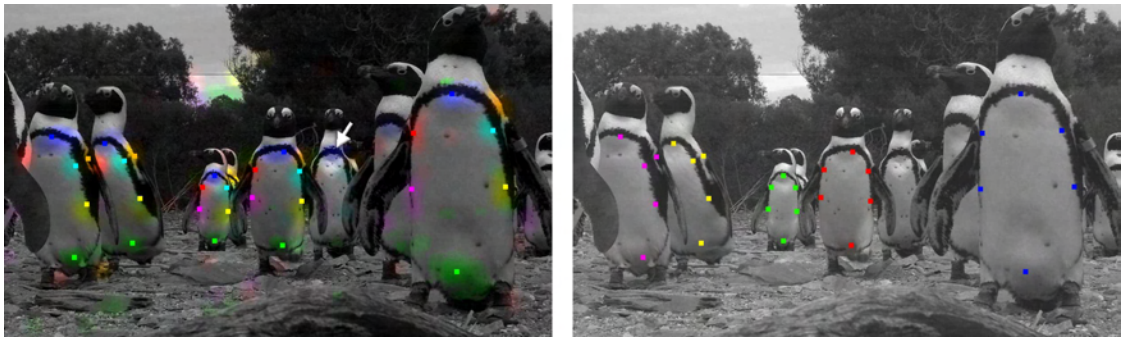
4.2.6 Greedy Detection: Depth First Search with Backtracking

During detection, the nodes of the graph are evaluated starting from the root in a depth-first search. The graph is navigated top-down to leafs along the edges with the highest predictive potential first. For each node the image is searched for the assigned feature i, i.e. $\mathcal{P}_A^i$ is locally evaluated in an area D_i around the predicted location according to the expected variance and the appearance structural evidence is calculated for the path from the root up to the node as:

(APPEARANCE-STRUCTURAL FIT)

$$L = \frac{1}{|F|} \underbrace{\sum_{i \in F}}_{\text{features on root-path}} \underbrace{\max_{\mathbf{x} \in D_i}}_{\text{search area}} \Big(\underbrace{\mathcal{P}_A^i(\mathbf{x})}_{\text{appearance evidence}} + \underbrace{\mathcal{P}_S^i(\mathbf{x})}_{\text{Gaussian prediction}} \Big) \tag{4.6}$$

The measure L is stored in the node. A path is discarded and backtracking is applied in the case that the evidence L of a node drops below a rejection threshold. The search is continued until all branches are searched or a leaf is reached with evidence above an acceptance threshold, indicating a detection. Note that dynamic programming aids the search since subproblems overlap, e.g. different paths share common features and, thus, $\mathcal{P}_A^i$ is not evaluated multiple times. Successful matching naturally yields a correspondence set for all the features in the nodes along the found path from root to leaf. Figure 4.5 visualises the correspondence set for a scene containing multiple penguins.



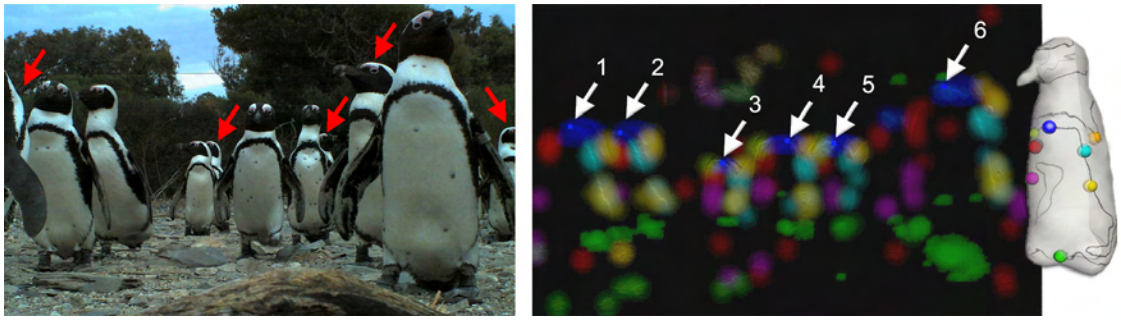
(a) **(b)**

Figure 4.5: ***Key Point Assignment using Feature Prediction Trees.*** **(a)** Key points confirmed by the FPT model. Key points (dots) are colour-coded by type superimposed on the the original luminance image and the key point belief maps. Note that one initialised FPT did not produce a detection, thus one master key points was dropped (see white arrow). **(b)** Confirmed points colour-coded so that two key points found by the same FPT carry the same colour. Since complete biometric entities (fully visible spot patterns) are required for the later individual identification, only frontal and near-frontal detections are forwarded to the biometric extraction components. For the above example only the individuals labelled in blue, red and green are used. [wildlife images: I01]

4.2.7 Case Study of Operation and Error Rates

This section briefly quantifies performance aspects of the approach for the case of detecting near-frontal instances of penguins in the image depicted in Figure 4.5.

The initial seeding of the most reliable key point (top chest) by appearance alone – as described in the previous chapter (see Eq.(3.9)) – leads to an initialisation of 6 detection trees. Figure 4.6(b) highlights the seed points detected whilst Figure 4.6(a) depicts the original image with annotations of the animals not found by the seeding process. Note that these 5 (mainly occluded) animal instances are not considered for any further detection, thus, in this particular image 45% of animals remain untested since no seed can be put in place.



(a) **(b)**

Figure 4.6: ***Seeding of the Detection.*** **(a)** Original image and **(b)** location of the six seed points (white arrows) highlighted in the appearance-based belief map calculated using Eq.(3.9). Note that not all animals in the scene spark a seed. The appearance information of a number of partly occluded animals (highlighted by red arrows in (a)) is insufficient to produce a seed point. [wildlife images: I01]

In order to interpret the further detection process, for each of the six seed points the tree based search was performed and the quality of fit L given in Eq. (4.6) was recorded at different stages of the detection process, starting with the initialisation at the root and ending with either a detection of an animal instance or the rejection of the seed point whenever no path in the tree yielded a successful match.

This process is exemplified on a single detection for the animal depicted as seed point 6 (right closeup penguin) in Figure 4.6(b). After seeding, further key points are subsequently searched for according to a next untested path of the tree (see Figure 4.7(c)). The image support from different key points along the path from the root was recorded and is visualised as a solid curve in Figure 4.7(a), the resulting development of the overall quality of fit L for the path was calculated after Eq. (4.6) and is depicted as a stroked line.

It can be seen that at tree depth 5 the image support turns out particularly low. As a result, the overall fit (stroked line) runs beneath the rejection threshold (horizontal grey line) and, thus, causes a rejection of the path altogether. As a result, backtracking to the last branching node (red) is performed and another path is tested.

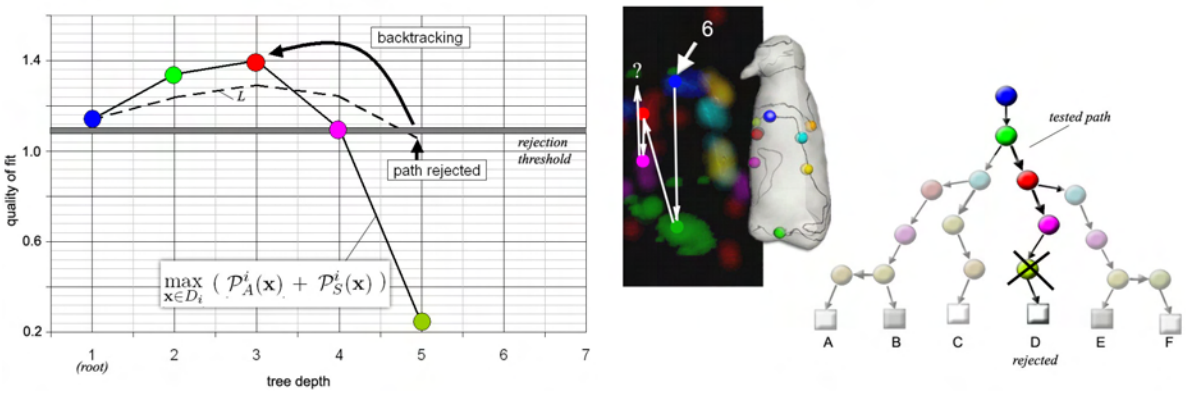


Figure 4.7: ***Path Testing Resulting in Rejection.*** **(a)** qualitative study of the contribution of nodes (solid line with colours indicating tested key point) along the first path probed and development of the quality of overall fit L (stroked line); **(b)** relevant section of the appearance-based belief map with superimposed spatial locations (coloured discs) of key points hypothesised during path evaluation; **(c)** visualisation of the first path tested in the tree;

A successful detection can finally be established with path E where a leaf is reached without breaking the rejection threshold at any point during the evaluation of the path. Figure 4.8 illustrates this case, again depicting the development of the quality of fit during detection.

Overall 5 out of the 11 animals present in the particular scene (i.e. approx. 45%) were successfully detected as shown earlier in Figure 4.5. Since the method is not robust with

regard to missing key points, i.e. it can not perform partial matching, occlusion constituted the main reason for failure to detect animal instances. However, all non-occluded animals were detected and no false positive detections were observed in this particular image.

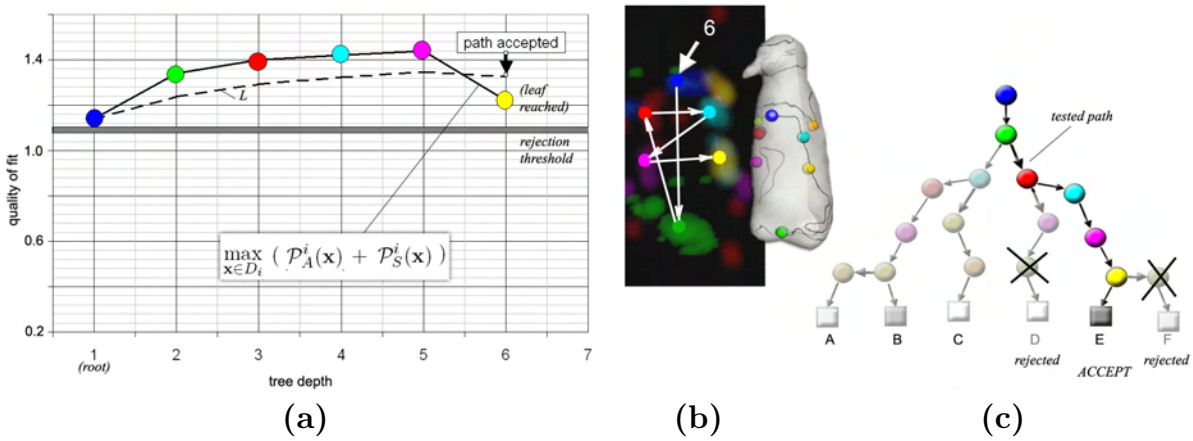


Figure 4.8: ***Path Testing Resulting in Acceptance.*** **(a)** qualitative study of the contribution of nodes (solid line with colours indicating tested key point) and development of the quality of overall fit L (stroked line); **(b)** relevant section of the appearance-based belief map with superimposed spatial locations (coloured discs) of key points detected; This configuration provides the detection result as shown in Figure 4.5(b) for the whole scene; **(c)** visualisation of the successfully tested path, in this particular case the animal is assigned path E which represents a frontal view;

In order to quantify the detection performance on a small image collection, 20 different images containing penguins in near frontal poses (exemplified in Figure 4.9) similar in content to the one depicted in Figure 4.6(a) were subjected to the detection algorithm.

Figure 4.9: ***Selection from the Sample Set.*** The figure illustrates three representative images from the test set used for testing the performance of the method.

The detections were then compared against a manually annotated ground truth. Subsequently, separate detection rates for 1) all animals present in an image, 2) only non-occluded animals, as well as 3) non-occluded and lightly occluded animals together were calculated. Figure 4.10 depicts the results. It can be seen that, while showing excellent

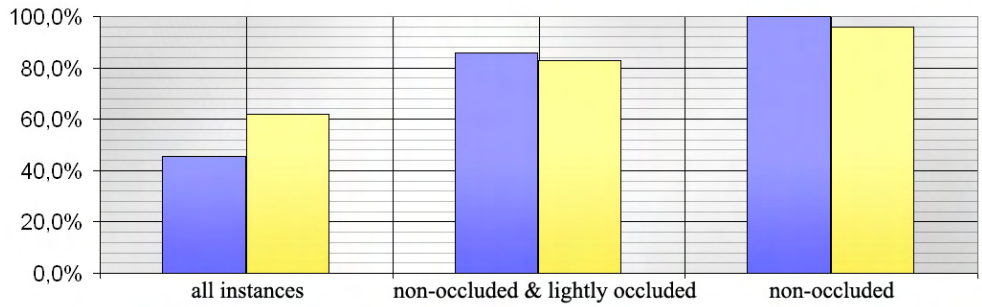


Figure 4.10: ***Detection Performance.*** The diagram visualises the detection rates (yellow bars) exhibited by the proposed method while tested on a small sample set of 20 images. The particular detection rates for the one sample image shown in Figure 4.6(a) are depicted by blue bars. Note that no false positive detections were observed. The diagram confirms that the method is well suited to detect non-occluded animal instances while occlusions lead to failure to detect.

performance in detecting non-occluded instances, the detection performance drops rapidly when occluded instances are taken into consideration.

4.2.8 Brief Discussion of Virtues and Drawbacks

It has been shown that the technique's robustness to occlusion is very limited. However, occluded instances are of little interest since biometric identification requires the *full spot pattern* to be visible. Thus, only non-occluded, frontal and near-frontal positions are forwarded to the biometric extraction components.

FPT's rapidly produce a landmark association to objects by greedy maximisation over the combined image support for appearance and structure. Note that the size of the tree is highly dependent on the object structure (e.g. the tree collapses to a single branch for a truly affine object). Also note that the approach predicts and thereby restricts the area of feature search (see Figure 4.4(b)), reducing the computational cost for the extraction of appearance evidence.

Although the model presented seems an 'overkill' for the association task (especially given the later use of rigid models for texture recovery), it offers a promising prospect for future extensions that may include the posing of flexibly adjustable 3D models on the basis of the sparse correspondences. Some ground breaking work by *Huang et al.* [97] on approximating 3D human face structures from 2D content based on 3D morphable models (after an initial, currently manual posing) could lead the way to an interesting future research direction. However, correspondences do not provide an explicit representation of a dense surface projection as needed for texture extraction – thus further modelling is required.

4.3 Texture Extraction

4.3.1 Fitting Affine Surface Models

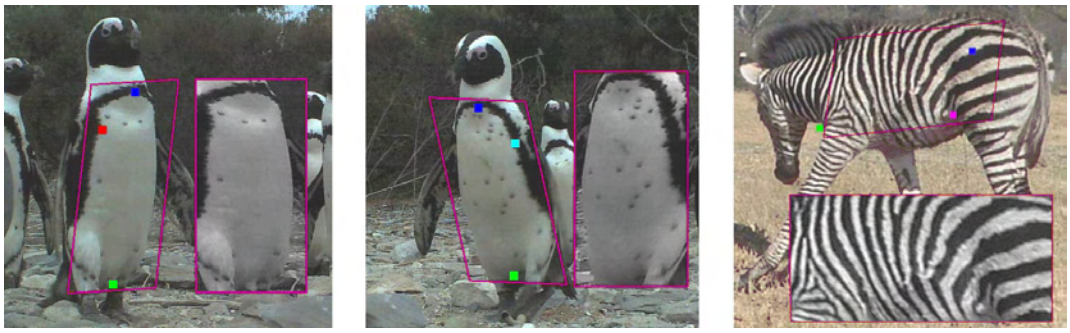
A globally affine texture description provides only a basic approximation of the transforms that an animal coat may undergo. Nevertheless, the current software prototype uses an affine representation since it is *quick and easy* to evaluate on client machines actively observing a colony. However, note that the subsequently applied biometric matching techniques – which are executed on a specially dedicated server – take skin deformations explicitly into account (see Chapter 6). Preliminary work on more complex surface models will be presented at the end of this section in order to illustrate potential system expansions.

An affine representation approximates a surface region as planar where the scale, rotation, shear and translation (six degrees of freedom) of the surface are considered. A model can then be mapped to an image instance according to a basic linear relationship:

$$\text{(AFFINE PROJECTION)} \qquad \mathbf{x} = \mathbf{A}\tilde{\mathbf{x}} + \mathbf{t} \tag{4.7}$$

where $\tilde{\mathbf{x}} \in \mathbb{R}^2$ is a model point, $\mathbf{x} \in \mathbb{R}^2$ is an image point, $\mathbf{A} \in \mathfrak{M}_{\mathbf{2}\times\mathbf{2}}$ is a linear transform and $\mathbf{t} \in \mathbb{R}^{\mathbf{2}}$ is a translation vector. Since three points (covering the six degrees of freedom) fully describe a 2D affine reference system, three correspondences $(\mathbf{x}_i, \tilde{\mathbf{x}}_i)$ between the affine model and an image instance are sufficient to provide an estimate for both $\mathbf{A}$ and $\mathbf{t}$.

Thus, the location of only three *sparse* key points can provide the information for an affine transform of *dense* textures. Warping according to this projection from the posed model into a *normalised pose* (standard scale at orthogonal view, no shear) can be used to implement the texture transform. Figure 4.11 illustrates the process on three examples.



(a) **(b)** **(c)**

Figure 4.11: ***Affine Texture Extraction.*** Three key point correspondences are sufficient to fit a basic, affine surface model. Warping the patch of interest of the underlying texture into a standard projection yields an affinely normalised texture patch. [wildlife images: I01, I02]

4.3.2 Affine Least Squares Fitting

Since only frontal and near-frontal images are permitted for further biometric analysis, at least *six key point correspondences* are available for *all* penguin instances of interest.

In order to increase the robustness of the model fitting by combining *all* of the observed correspondences instead of using merely three measurements, a standard least squares approximation is used to model a consensus over all points. (Note that coarse outliers are suppressed by the previously applied FPT fit.) Rewriting Eq. (4.7) in homogenous coordinates for model points yields:

(AFFINE PROJECTION)

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_1 \\ a_{21} & a_{22} & t_2 \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ 1 \end{bmatrix} = \mathbf{A}\tilde{\mathbf{x}} \tag{4.8}$$

For the purpose of retrieving the projection $\mathbf{A}$ by six correspondences $(\mathbf{x}_i, \tilde{\mathbf{x}}_i)$, the observations are first stacked into matrices $\mathbf{X} = [\ \mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_3\ \mathbf{x}_4\ \mathbf{x}_5\ \mathbf{x}_6\ ]$ and $\tilde{\mathbf{X}} = [\ \tilde{\mathbf{x}}_1\ \tilde{\mathbf{x}}_2\ \tilde{\mathbf{x}}_3\ \tilde{\mathbf{x}}_4\ \tilde{\mathbf{x}}_5\ \tilde{\mathbf{x}}_6\ ]$ where $\mathbf{X} \in \mathfrak{M}_{2\times 6}$ and $\tilde{\mathbf{X}} \in \mathfrak{M}_{3\times 6}$. The projection matrix $\mathbf{A}$ is then recovered as:

(PROJECTION RECOVERY)

$$\mathbf{A} = \mathbf{X}\tilde{\mathbf{X}}^T(\tilde{\mathbf{X}}\tilde{\mathbf{X}}^{\mathbf{T}})^{-1} \tag{4.9}$$

and the surface texture is retrieved by warping as explained before. Note that this affine model still describes a 'flat' geometry, i.e. the texture is modelled as merely sitting on a plane. Future extensions to the software prototype will address this problem.

4.3.3 Experiments with 3D Surface Models and Future Research Directions

This last section of this chapter now presents preliminary work on 3D model fitting and discusses the biometric benefits and prospects of using further advanced surface descriptions.

In general, true 3D models are needed in order to represent surface curvatures and account for occlusions. However, a weak perspective projection of these models appears to be sufficient since the animals are registered within a highly constrained range of distances. As a result, the encountered variance in depth is small. So neglecting the distortions of the content of interest due to depth variations, an orthographic projection scaled by a factor inversely proportional to an average depth Z can be used to approximate the mapping from

a 3D model space into the 2D image space. This weak perspective projection can formally be expressed by:

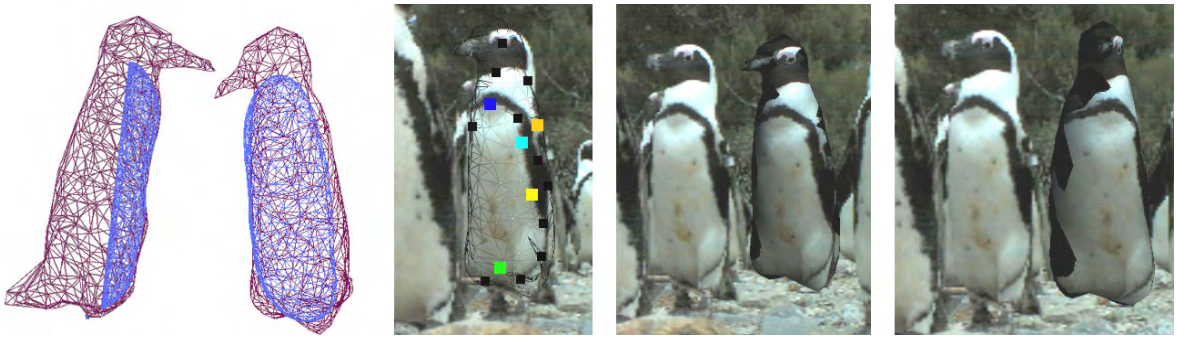
(WEAK PERSPECTIVE PROJECTION)

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} = \underbrace{\frac{1}{Z} \begin{bmatrix} f_x & 0 \\ 0 & f_y \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \end{bmatrix}}_{\text{projection matrix } \mathbf{A}} \begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \\ 1 \end{bmatrix} = \mathbf{A}\tilde{\mathbf{x}} \tag{4.10}$$

where a world point $\tilde{\mathbf{x}} \in \mathbb{R}^4$ (homogeneous coordinates) is projected onto an image point $\mathbf{x} \in \mathbb{R}^2$, f_x and f_y are the magnifications in the two image directions, Z reflects the average scene depth, t_1 and t_2 determine the translation, the coefficients r_{ij} encode rotations, and the matrix $\mathbf{A} \in \mathfrak{M}_{2\times 4}$ encodes the entire projection as a linear map.

In order to fit a 3D model according to (at least four) correspondences $(\mathbf{x}_i, \tilde{\mathbf{x}}_i)$, the projection has to be approximated. First the key point observations are stacked into matrices $\mathbf{X} = [\ \mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_i\ \ldots\ \mathbf{x}_n\ ]$ and $\tilde{\mathbf{X}} = [\ \tilde{\mathbf{x}}_1\ \tilde{\mathbf{x}}_2\ \ldots\ \tilde{\mathbf{x}}_i\ \ldots\ \tilde{\mathbf{x}}_n\ ]$ where $\mathbf{X} \in \mathfrak{M}_{2\times n}$ and $\tilde{\mathbf{X}} \in \mathfrak{M}_{4\times n}$. Using the same concept as in the affine case, the projection matrix $\mathbf{A}$ is recovered using, again, a standard least squares solution of the underlying linear system as outlined earlier in Eq. (4.9).

The projection $\mathbf{A}$ is used to extract a pose-corrected texture by first back-projection onto the model and then re-rendering in a standard pose. Figure 4.12 illustrates 1) the 3D model used for the fitting process, and 2) a preliminary result of the process of pose normalisation.



(a) **(b)** **(c)** **(b)**

Figure 4.12: ***Texture Correction by Posing a 3D Model.*** **(a)** Wire-frame visualisation of the 3D model employed; **(b)** The 3D model fitted based on 5 key point correspondences; Note the inaccuracies of fit. **(c)** Back-projection, textured model (right) next to the original; **(d)** Rendering the model in a standard view produces a pose-corrected texture. Note the occluded texture (black).

An integration of the 3D surface fitting technique into the software prototype could open the path 1) to a biometric identification of subpatterns by considering the area of spatial overlap of patterns, and 2) to a further improvement of the distortion normalisation by accounting for curvature and (weak) perspective distortion effects.

As already indicated, a rigidly posed model could be used as the initialisation stage of a 3D morphable model [97] in order to recover even more detailed descriptions of a surface. However, in order to build and apply morphable models, the 3D parameter space for valid 'penguin surfaces' and 'penguin textures' must be stipulated by collecting significant numbers of real-world 3D sample scans of the object of interest. This poses a number of practical as well as ethical questions on how to produce a database of wild animal scans.

4.4 Chapter Summary and Outlook

In this chapter it has been illustrated how the sparse, component-like surface representation via key points can be exploited to extract dense and pose-corrected surface textures of biometric interest.

A spatially flexible model for spatial model representation and detection termed *feature prediction trees* has been used to achieve an assignment of key point instances to animal instances where the flexible configuration of landmarks has been modelled by several affine domains linked by Gaussians.

The resulting correspondence sets have then been exploited for fitting geometrical surface descriptions that allowed for the extraction of *pose-corrected texture maps*. It was outlined that the current prototype uses an affine correction model whilst future expansions of the system will focus on fitting 3D models in order to account for curvature, perspective distortions and partial occlusion.

The next and final content chapter of this thesis will discuss the utilisation of the extracted texture patches for biometric identification, that is to isolate unique components from these animal textures and associate them with individual identities. The recognition of individual members of the species' population will complete the proposed 'locate-pose-identify' methodology and open a path to the practical identification of individual animals. ■

Chapter 5

INDIVIDUAL IDENTIFICATION BY COAT PATTERN

'Methodological individualism and functionalism, both tend to be reductionistic.' [158]

(Karl R. Popper, 1902 - 1994)

5.1 Chapter Overview

The following chapter culminates the theme of the thesis. It demonstrates how features of Turing-like coat textures can be utilised to achieve an identification of *individual animals*.

In particular, the chapter presents two biometrically motivated algorithms which are applicable to animal populations. That is 1) an image processing technique that extracts sparse sets of *individually unique landmarks* by combining a bandpass filter with a curl detector that operates on the gradient direction field, and 2) a representation and comparison framework that expands on *shape contexts* [20] and uses the *earth mover's distance* [172] as well as the *Hungarian method* [115] for associating detected landmarks with animal identities.

The techniques are then evaluated based on experiments conducted in a colony of African penguins. Finally, the chapter concludes by discussing the influence of random pattern correspondence on the system's identification performance and reflects on the prospects and limitations of visual animal monitoring using coat textures as biometric entities.

5.2 Sparse Landmark Signatures from Coat Patterns

5.2.1 Generality of Spatial Phase Singularities

A multitude of animals exhibit prominent pattern elements in the form of spots, line endings and/or bifurcations on their coats. Figure 5.1 illustrates that these features are commonplace in coat markings across the animal kingdom. For Turing patterns, in particular, the features represent visual evidence for underlying phase singularities, i.e. locations of un-

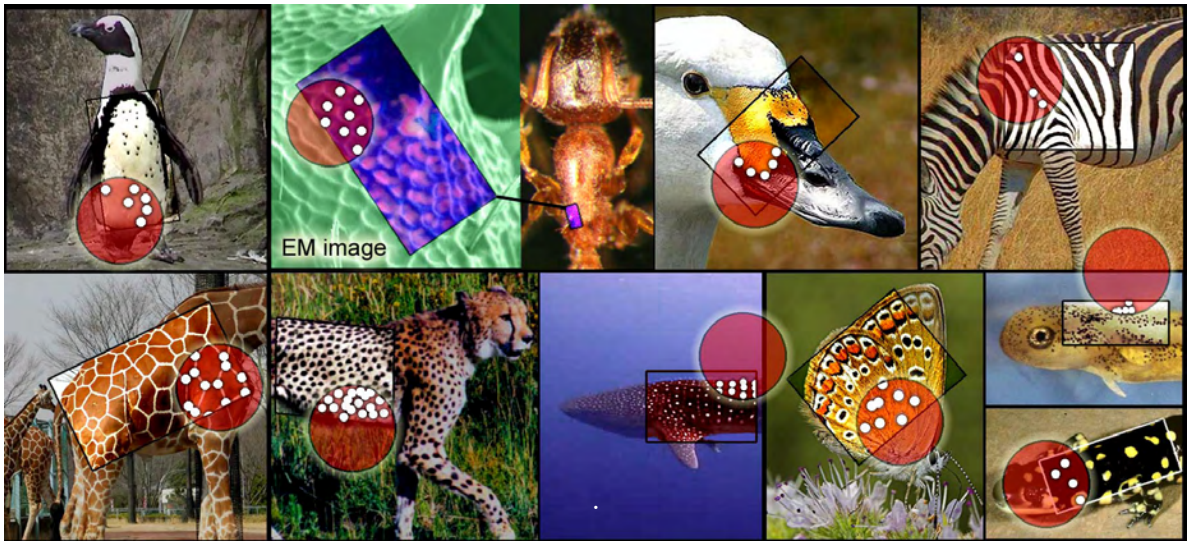


Figure 5.1: ***Spatial Phase Singularities in Coat Patterns across the Animal Kingdom.*** The images show examples of phase singularities (highlighted by white spots) in coat patterns occurring in diverse animal species.

defined gradient orientation in the system of pattern-inducing substance concentrations[1]. These specific landmarks embody characteristics of a particular system evolution and, thus, carry information on the *identity of individual animal*s.

In order to utilise the *geometrical configuration* and *topological typing* of singularity landmarks for animal authentication and identification, the temporal persistence of the physiological landmark features must be guaranteed. This property will now be briefly discussed for the two sample species, that is for plains zebras and African penguins.

5.2.2 Fixation and Permanence of Landmarks

As a matter of fact, all zebras are born with a finalised, *topologically permanent* and *geometrically robust*[2] stripe pattern. African penguins, in contrast, are subjected to several moults before developing a persistent adult plumage. According to careful observations on several individuals, chest patterns only remain constant after a bird has undergone the second moult which coincides with the development of a prominent neck stripe[3].

[1]Spatial phase singularities are discussed in detail in Section 2.5. Note that there exist biologically relevant Turing systems such as (rotating) spiral structures [117] which do not exhibit phase singularities.

[2]The landmark geometry is *not* truly rigid: breathing, locomotion, pregnancy etc. – all alter the geometry.

[3]The temporal coincidence of the development of the chest stripe and the permanent spot pattern allows the species detector to select only stable adult patterns. However, the strict additive nature of the lay-down process may permit for future work on a partial spot pattern recognition scheme.

Figure 5.2 illustrates this pattern evolution on the chest of an individual African penguin at the relevant phase of its plumage maturation.

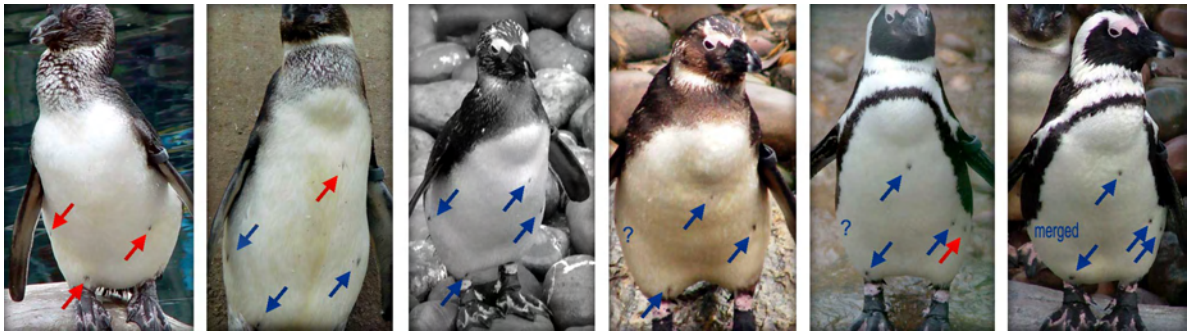


Figure 5.2: ***Spot Pattern Development in an African Penguin.*** The time line of photographs visualises the final stages of the incremental lay-down of an African penguin's chest pattern. Newly occurring, initially very faded spots are highlighted by red arrows whilst blue arrows confirm previously existing spots. Notice that the pattern only remains constant in the final two images where the the chest stripe also appears matured. Note that, despite the possibility of spots merging with the developing stripe, the maturation process is *additive* in its nature. [photos courtesy of R Sherley: I01]

Once the adult coat pattern is fully developed the spot configuration is topologically persistent, that is the specific colouration of feathers is regrowing after moults between seasons.

Figure 5.3 provides visual evidence for this phenomenon, showing photographic material of African penguins taken at particular nesting sites over substantial periods of time. The images confirm the reoccurrence of the same (highly unusual) pattern in different years.

Figure 5.3: ***Permanence of Spot Patterns in African Penguins.*** The images exemplify the persistence of chest patterns. The two individuals shown carry particularly abnormal patterns that remained stable over a period of several years. (As a side note: these two specific birds are not identified as African penguins by the species detector due to their strong pattern teratology.) [photos courtesy of PJ Barham: I01]

5.2.3 Detection of Sparse Points in Dense Coat Textures

Spatial phase singularities can be interpreted as a special class of interest points where the biometric focus lies on a high *repeatability* of their detection on coat textures and *robustness* against imposter features. To achieve this, spectral as well as directional properties will be analysed in order to infer the locations and types of landmarks in coat textures.
A prototypical spatial phase singularity in a Turing pattern $\mathbf{I}$ has the following properties:

1. **Extrema.** The relevant localisation information is locked around dominant, species-typical frequencies $f_{\mathbf{x}}$, i.e. the equivalents of the anticipated local line or spot wavelengths $\lambda_{\mathbf{x}}$ – as illustrated in Figure 5.4(a)-(b) – where the function $f_{\mathbf{x}}$ assigns a frequency of interest to each location $\mathbf{x}$. At this value spatial phase singularities represent extreme points of the image function as shown on a spot feature in Figure 5.4(c).

2. **Local Gradient Curls.** A landmark is surrounded by (at least partial) curls of the spatial gradient direction[4] Θ_f as exemplified in Figure 5.4(d). Thus, in a digital representation, at the tip of the curl (i.e. at the actual singularity) pixels exhibit high values of the spatial phase gradient $\nabla\Theta_f$.

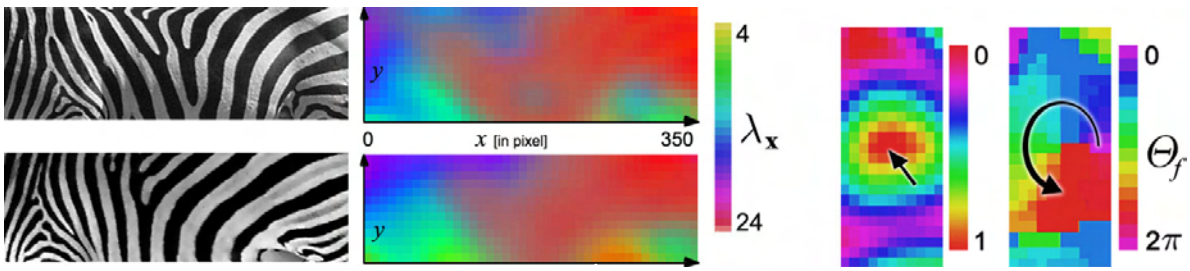


Figure 5.4: ***Dominant Frequency Field and Gradient Direction Field.*** **(a)** two coat patterns of plains zebras; **(b)** the dominant frequency field of the two patterns illustrates the strongest local frequency $f_{\mathbf{x}}$ at different locations $\mathbf{x}$ of the texture (where the wavelength $\lambda_{\mathbf{x}} = 1/f_{\mathbf{x}}$ is shown in pixels). Note that this field $f_{\mathbf{x}}$ is homogenous in African penguins. **(c)** a spatial phase singularity (arrow) marks the extreme point at the centre of an isle feature; **(d)** the surrounding curl of gradient directions Θ_f is indicated by an arrow. [zebra images used: I02]

In real animal textures the two properties are often 'quantitatively compromised' due to noise induced by fur or feather coverage, faded patterns, shadows and/or pattern degenerations. However, it will be shown that – used in conjunction – the two properties provide

[4]The curling property of the spatial phase gradient has been used before in the *Bray-Wikswo algorithm* [26] for identifying phase singularities in cardiac signals. Note that, although all phase singularities represent corners, not all corner points are spatial phase singularities (counterexample: corner of a large rectangle).

a set of characteristics that offers both a spectral constraint for an effective noise suppression and a structural constraint for an accurate localisation. For the practical extraction of landmarks, a sequence of four image processing steps is proposed:

1. **Bandpass Filtering.** First, the relevant frequencies around $f_{\mathbf{x}}$ are isolated into an image $\mathbf{I}_f$ by convolution using a *Difference of Gaussians* (DoG) to approximate a Laplacian of Gaussian (LoG) bandpass kernel[5], i.e. $\mathbf{I}_f \approx \nabla^2 G_f * \mathbf{I}$. Figure 5.5(a)-(d) illustrates the creation of the bandpass-filtered signal $\mathbf{I}_f$.

2. **Extraction of Gradient Directions.** Second, the *gradient direction field* $\Theta_f = \arctan\left(\frac{\partial G_f * \mathbf{I}}{\partial y}\left(\frac{\partial G_f * \mathbf{I}}{\partial x}\right)^{-1}\right)$ is calculated. Note that filtering by a low pass is essential to control the sensitivity of Θ_f with regard to noise as shown in Figure 5.5(e)-(f).

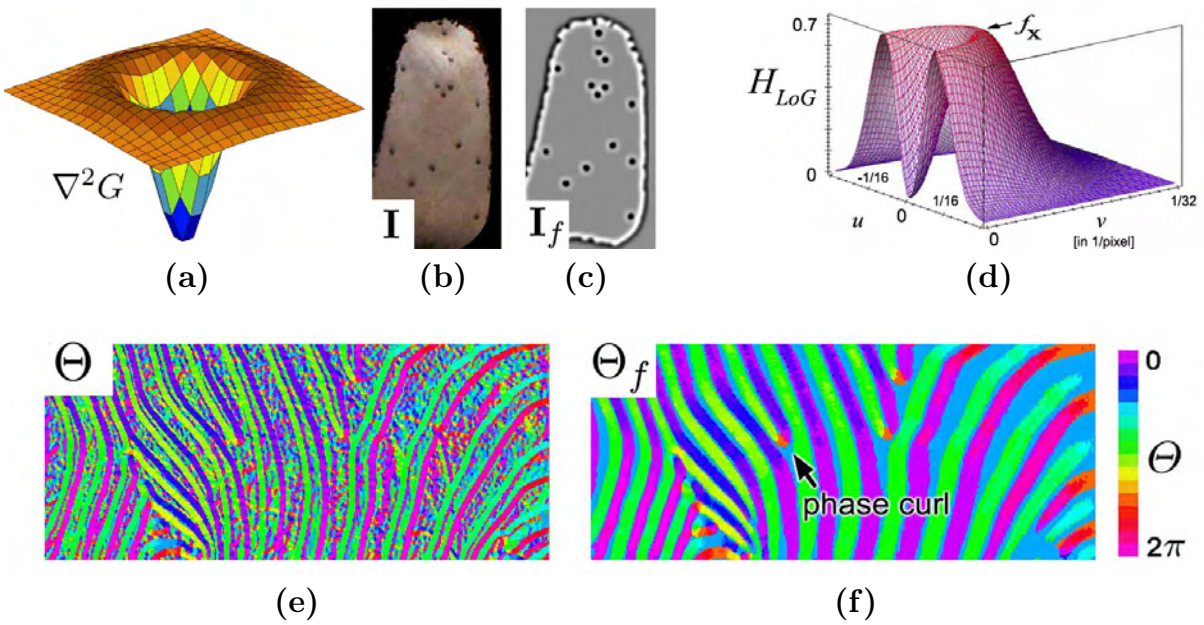


Figure 5.5: ***Spectral Confinement.*** **(a)** Laplacian of Gaussian (LoG) or 'Mexican Hat' kernel; **(b)** texture map $\mathbf{I}$ of a penguin chest; **(c)** the bandpass-filtered map $\mathbf{I}_f$ – the spots are clearly amplified since **(d)** the transfer function of the LoG kernel peaks at $f_{\mathbf{x}}$; **(e)** the gradient direction field Θ of a zebra texture $\mathbf{I}$; **(f)** The gradient direction field Θ_f of the bandpass-filtered image $\mathbf{I}_f$ highlights the information-bearing gradient directions: relevant curls of the spatial phase gradient (one is indicated by an arrow) are amplified and noise is effectively suppressed. [original animal images used: I01, I02]

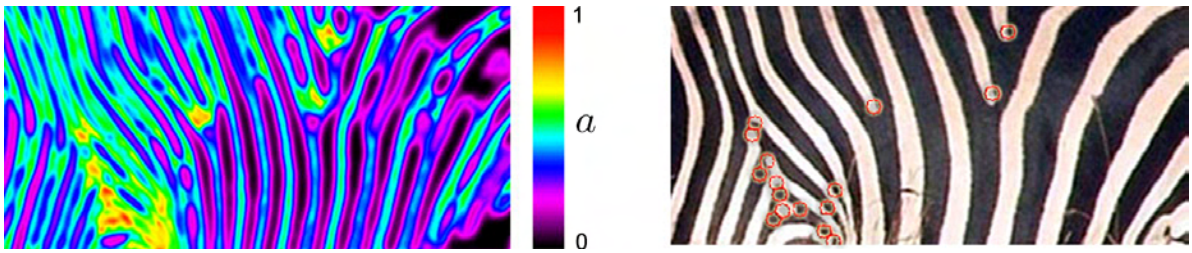
[5]The DoG is used to approximate a Laplacian of Gaussian (LoG) $\nabla^2 G_f = -\left(\frac{x^2+y^2-\sigma^2}{\sigma^4}\right) \mathrm{e}^{\frac{x^2+y^2}{2\sigma^2}}$ kernel which is depicted in Figure 5.5(a). As shown by Koenderink [110], Gaussians are highly suitable for scale space analyses. In fact, a multitude of blob detectors exist which use Gaussian kernels. Some of them have been shown to provide minute performance improvements compared to LoG such as the Determinant of the Hessian (DoH) [16].

3. **Curl Detection.** In order to probe the neighbourhood of a location $\mathbf{x}$ for a phase curl, a set of m phase histograms $h_{\mathbf{x}}$ is built by counting phase directions in disc-shaped neighbourhoods of radii 1 to $\lambda_{\mathbf{x}}/2$. The L_2-distance of each histogram from an even distribution[6] is then added in a weighted form into an accumulator array $a(\mathbf{x})$ (see Figure 5.6(a)). Formally, the calculation of $a(\mathbf{x})$ can be summarised as follows:

(ACCUMULATOR ARRAY FOR CURL DISCOVERY)

$$a(\mathbf{x}) = 1 - \underbrace{M}_{\text{normalisation}} \overbrace{\sum_{j=1}^{m}}^{m \text{ histograms}} \left(\underbrace{\frac{(1+m-j)}{N}}_{\text{importance weights}} \sqrt{\overbrace{\sum_{i=1}^{n}}^{n \text{ bins}} \underbrace{(h^j_{\mathbf{x}}(i) - \overbrace{\frac{|D^j_{\mathbf{x}}|}{n}}^{\text{mean}})^2}_{\text{bin residual from mean}}} \right) \tag{5.1}$$

where the $h^j_{\mathbf{x}}(i) = \sum_{\mathbf{d} \in D^j_{\mathbf{x}}} \begin{cases} 1 & \text{if } \left(i = 1 + \lfloor \frac{n\ \Theta_f(\mathbf{d})}{2\pi} \rfloor\right) \\ 0 & \text{elsewise} \end{cases}$ represent the bin values of the n-bin phase histograms of mean bin value $\frac{|D^j_{\mathbf{x}}|}{n} = \frac{1}{n} \sum_{i=1}^{n} h^j_{\mathbf{x}}(i)$, $D^j_{\mathbf{x}}$ is a disc-shaped neighbourhood around $\mathbf{x}$ of radius $\frac{j\lambda_{\mathbf{x}}}{2m}$, Θ_f is the gradient direction, $|.|$ represents the set cardinality, $\mathbf{I}_f = G_{f_{\mathbf{x}}} * \mathbf{I}$ is the filtered texture image, $G_{f_{\mathbf{x}}}$ are Gaussian kernels, and M and N are normalisation terms[7] which ensure that $a(\mathbf{x}) \in (0, 1)$.



(a) **(b)**

Figure 5.6: ***Curl Detection by Histogramming.*** **(a)** The image shows a visualisation of the accumulator array $a(\mathbf{x})$ built from **(b)** the underlying, original zebra texture $\mathbf{I}(\mathbf{x})$ with strong curl maxima superimposed. Note that a number of bifurcation features depart in their properties from the dominant frequency assumption, i.e. the width of contributing stripes varies greatly. As a result, features are missing or misplaced. For instance, the three black bifurcations at the body centre are not detected whilst their counterparts (i.e. white stripe terminations) are found. Topological considerations will help with overcoming the problem of degenerated features. [original zebra image: I02]

[6] For a location to be the centre of a complete ($= 360°$) curl of the phase gradient in $\mathbf{I}_f$, *all* gradient directions must be present in *equal* quantities over a disc-shaped neighbourhood $D_{\mathbf{x}}$ of diameter $\lambda_{\mathbf{x}}$.

[7] The normalisation constant $M = \frac{2}{m^2+m}$ balances the importance weights whereas $N = |D^j_{\mathbf{x}}| \sqrt{\frac{(n-1)}{n}}$ describes the maximal cumulative residual of a histogram (in the case all gradients in $D^j_{\mathbf{x}}$ aim at the same direction). For the implementation, the user-defined parameters were chosen to be $m = \lambda_{\mathbf{x}}/2$ (no. of different neighbourhoods) and $n = 8$ (no. of different histogram bins to resolve phase). Note that the importance weight in Eq. (5.1) decreases with the size of the neighbourhood, i.e. adjacent gradient curls are given higher relevance. As a result, the accumulator array $a(\mathbf{x})$ shows distinctive peaks.

4. **Fusion of Evidence and Localisation of Landmarks.** The described curl detector $a(\mathbf{x})$ registers both types (i.e. L-type=black=-1 and H-type=white=1) of landmarks as distinct maxima whilst the DoG filter separates them as either minima or maxima. By multiplying the two measures, a specific type T of landmarks (e.g. all L-type features) can be amplified yielding a detector function $g_T(\mathbf{x})$ (see Figure 5.7):

$$\text{(SPATIAL PHASE SINGULARITY DETECTOR)}$$
$$g_T(\mathbf{x}) = \underbrace{(\frac{T+1}{2} - T\mathbf{I}_f(\mathbf{x}))}_{\text{DoG blob evidence}} \underbrace{a(\mathbf{x})}_{\text{phase curl evidence}} \tag{5.2}$$

where $T \in \{1, -1\}$ determines the type and the landmark positions are detected based on this measure by non-maxima suppression in a neighbourhood of radius $\lambda_{\mathbf{x}}$ and subsequent thresholding.

The combined approach fixes some of the shortcomings of the single detectors: the curl detector $a(\mathbf{x})$ on its own is intensity-blind whilst the DoG has somewhat less pronounced peaks (see Figure 5.7(a)) and disambiguates poorly between line endings and line segments due to the radial symmetry of the kernel (see Figure 5.7(b)). These specific, directionally sensitive features are, on the other hand, strongly amplified by the histogram technique which probes for the *specific change of gradient directions* in the neighbourhood.

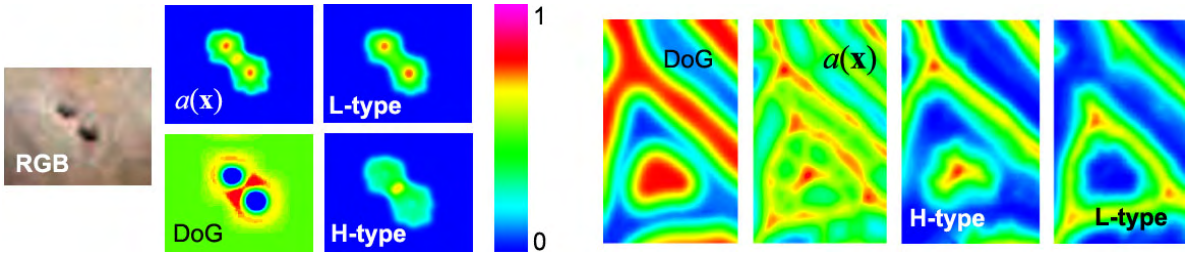


Figure 5.7: ***Type Identification using the Combined Detector.*** **(a)** The images illustrate the process of landmark localisation and typing on two penguin spots. The phase histogram measure $a(\mathbf{x})$ clearly shows two maxima and one less pronounced maximum whilst the DoG blob detector exhibits two minima. Note that using the DoG the saddle point (i.e. the middle point in $a(\mathbf{x})$) is not pronounced at all. Multiplication according to Eq.(5.2) leads to two images that each amplify landmarks of one specific type. **(b)** The same four functions as shown for penguins are illustrated on a zebra texture (scapular detail of the coat pattern shown in Figure 5.6(b)). Note the improvement in localisation and the suppression of spurious maxima in the combined detectors. [original animal images used: I01, I02]

The described approach can be intuitively understood as a search over a confined subspace of scale for locations that are surrounded by gradient curls[8].

An application of the L-type detector to synthetically altered penguin textures (depicted in Figure 5.8) demonstrates the technique's selectiveness with respect to scale, its robustness to JPEG-compression artefacts, its tolerance to some degree of distortion[9] and changes in lighting, as well as its applicability in the presence of noise as quantified in Table 5.1.

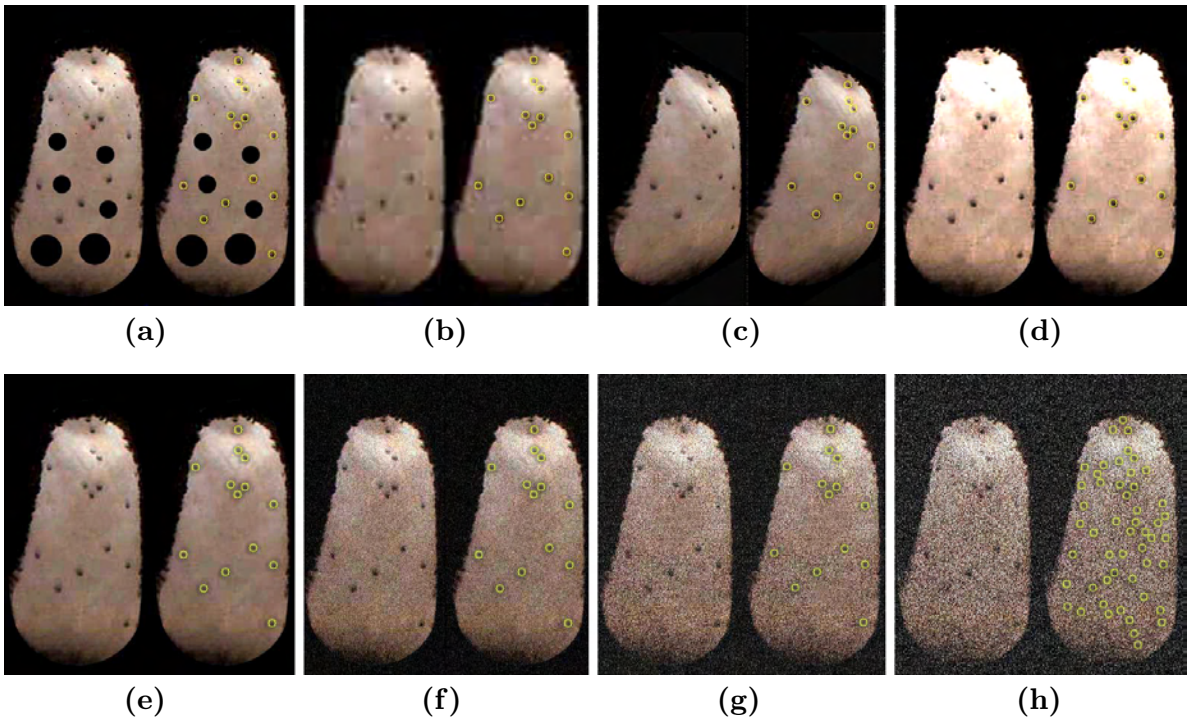


Figure 5.8: ***Illustration of Synthetic Robustness Tests on Penguin Patterns.*** **(a)** the detector's selectiveness of scale is illustrated on a number of synthetically generated spots of different sizes (detections are indicated by yellow circles); **(b)** robustness to JPEG artefacts: shown is a spot detection in an image resolved at 80×160 pixels at maximum JPEG compression; **(c)** spot detection after an perspective transform of the chest texture; **(d)** the differential nature of the DoG filter and intensity-blindness of the phase detector $a(\mathbf{x})$ results in a tolerance towards lighting changes. **(e)-(h)** spot detection in images where 0%(original), 10%, 20% and 30% Gaussian noise was added. The detector starts to produce erroneous results at about 20% added Gaussian noise. [original texture: I01]

[8]The main differences of the proposed method to the *Bray-Wikswo algorithm* [26] are 1) the use of DoG for noise control and the spectral selectivity, and 2) the use of local histogramming instead of circular integration (which rigidly assumes that $\oint \nabla\Theta = 2\pi$ around a phase singularity). Note that neither method can account for noise events *at* the dominant frequency (possibly deep shadows of leafs, branches etc.).

[9]Note that fully affine-invariant interest point detectors such as the one by *Mikolajczyk and Schmid* [141] perform a local, affine normalisation (e.g. a Hessian analysis) *before* determining the location (e.g. LoG).

Noise Level	**Repeatability Rate** [%]	**Imposter Rate** [%]
original images	100.0	0.0
+10% Gaussian noise	100.0	0.0
+15% Gaussian noise	100.0	0.0
+20% Gaussian noise	100.0	2.2
+25% Gaussian noise	97.4	18.5
+30% Gaussian noise	92.3	231.2

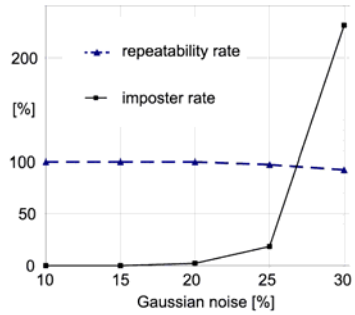
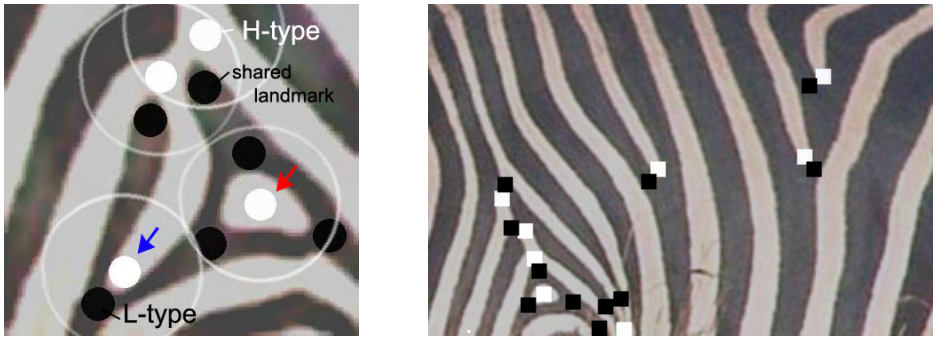


Table 5.1: **Repeatability of Spot Detection under Noise.** The table shows experimental results conducted on 50 penguin chest patterns filmed in good lighting conditions (comparable to Figure 5.8(e)). The spot detector was found to operate without failure on the original images (manually checked). In order to determine the robustness to noise, Gaussian noise was then added at different levels to the originals and the percentage of accurately detected spots (repeatability rate) as well as the percentage of falsely identified features (imposter rate) were measured at the working area of the detector. It can be concluded that, for good quality penguin patterns, the detector is a reliable means for spot identification below a level of 20% added Gaussian noise.

5.2.4 Topological Supplement

Landmarks in densely structured Turing patterns (e.g. in zebra patterns) are *always* associated to at least one (potentially shared) counterpart of the opposite feature type[10] situated in their adjacent $\lambda_{\mathbf{x}}$-neighbourhood. For instance, as illustrated in Figure 5.9(a), an H-type termination may be coupled with an L-type bifurcation (blue arrow) and an H-type isle may be coupled with multiple L-type bifurcations (red arrow). By using the combined probability of pairs for detection, a coupling criteria can be enforced, i.e. landmarks are identified by probing for the presence of two, inversely typed and neighbouring landmarks, where the strongest available pairing is used for a landmark to represent its certainty.



(a) **(b)**

Figure 5.9: ***Topological Considerations.*** **(a)** The image shows a configuration of scapular zebra stripes. It illustrates the pairing of landmarks of opposite types where white circles represent neighbourhood regions of white (H-type) landmarks. **(b)** Type-sensitive and topologically supplemented landmark detection in a zebra texture using Eq.(5.3). [original animal images used: I02]

[10] As before, the L-type is associated to black features whilst the H-type describes white features.

This technique of coupled landmark detection can be formally expressed as:

$$\begin{aligned} &\text{(PAIRING-AWARE DETECTOR)} \\ &t_{_T}(\mathbf{x}) \;=\; g_{_T}(\mathbf{x}) \;\max_{\bar{\mathbf{x}}\in D_{\mathbf{x}}} \left(g_{_{-T}}(\mathbf{x})\right) \end{aligned} \tag{5.3}$$

where $D_{\mathbf{x}}$ is a disc-shaped neighbourhood around $\mathbf{x}$ of radius $(1+\epsilon)\lambda_{\mathbf{x}}$. The parameter ϵ is empirically set to 0.5; it embodies the degree of natural deviation of pattern elements from the locally dominant frequency. Figure 5.9(b) depicts a sample application of the coupled landmark detection to a zebra pattern. Note the accurate identification of pairings in comparison to Figure 5.6(b).

5.2.5 Representation of Landmark Sets

The spatial configuration of singularity features is captured by a 2D point cloud $\{\mathbf{x}_1, ..., \mathbf{x}_j, ...\}$ which contains the positions $\mathbf{x}_j \in \mathbb{R}^2$ of features (e.g. spot centres, bifurcation points) expressed in the reference system of the animal texture $\mathbf{I}$. For illustration, Figure 5.10 shows a selection of spot clouds extracted on chest of different African penguins.

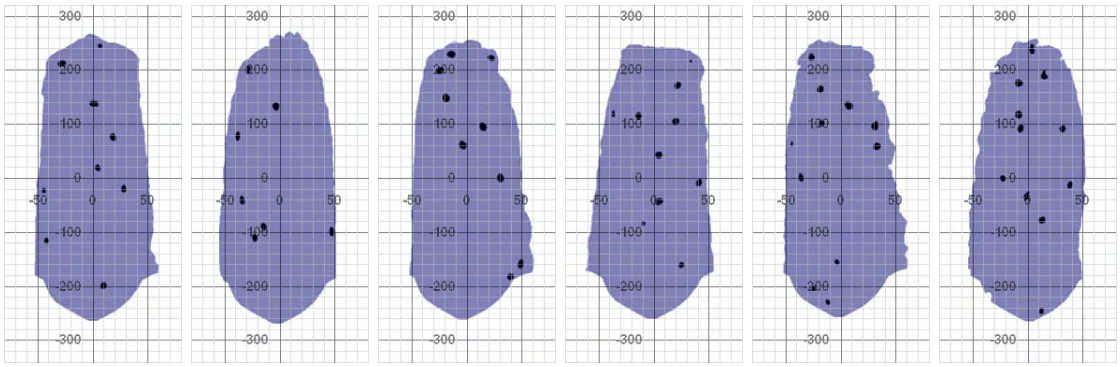


Figure 5.10: ***Chest Patterns of Penguins.*** The images visualise pose-normalised frontal views of penguin chests (blue) and detections of spot features on them (black dots). [data from field session: I01]

However, not all landmarks are equally likely to be valid detections. In scenes taken under difficult acquisition conditions (low resolution, specular reflections etc.) spots *are* missed by the detector as illustrated in Figure 5.11(a).

The probability of missing features was found to cluster in particular regions of the coat pattern, e.g. at skin folds, deeply shaded regions and at the border of the segmented pattern region (due to inaccurate segmentation, merging features and/or perspective distortions). To incorporate this a-priori knowledge, each species pattern is assigned a confidence mask that approximates the local confidence in the detection outcome. Figure 5.11(b)-(c) illustrates the masks used for penguins and zebras.

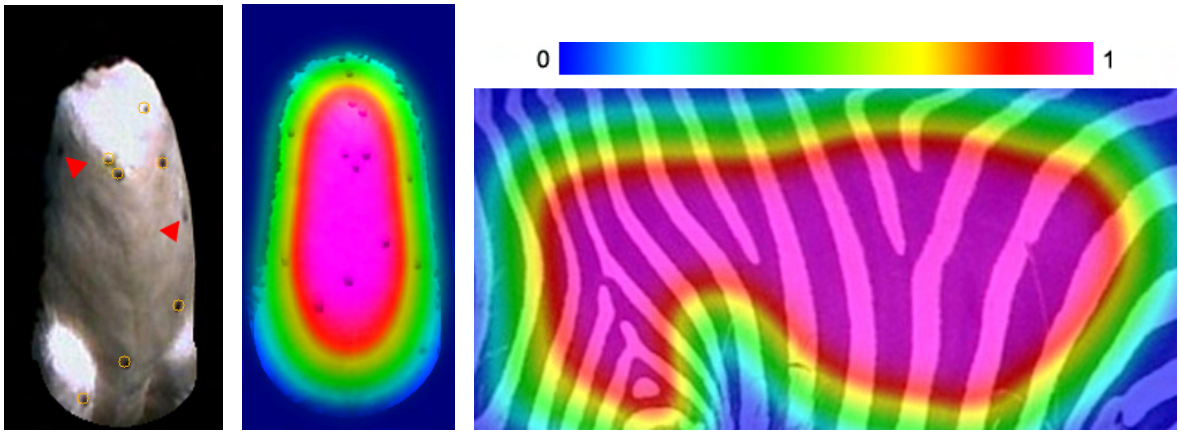


Figure 5.11: ***Modelling Detection Confidence.*** **(a)** spot detections (yellow circles) and missed spots (marked by red triangles) in a penguin chest filmed under critical lighting conditions; **(b)** confidence mask overlayed as a colour map onto a representative penguin chest pattern; **(c)** confidence mask used for zebra patterns; [original animal images used: I01, I02]

Employing both the confidence masks as well as the landmark type information, each detected feature location $\mathbf{x}_j$ is decorated with a confidence vector $\mathbf{t}_j \in [0,1]^\zeta$ that carries entries reflecting the confidence in having detected an instance of a specific feature class at $\mathbf{x}_j$ where each entry in $\mathbf{t}_j$ covers one of the disambiguated landmark classes[11].

Pairs $[\mathbf{x}_j, \mathbf{t}_j]$ of spatial data and confidence annotations form *typed points*[12]. Clouds of typed points are used as the basic model for representing a landmark configuration:

$$\text{(TYPED POINT CLOUD)} \quad \{[\mathbf{x}_0, \mathbf{t}_0], [\mathbf{x}_1, \mathbf{t}_1], ..., [\mathbf{x}_j, \mathbf{t}_j], ...\} \tag{5.4}$$

Note that, despite the global normalisation achieved by the pose model, the point cloud is susceptible to non-linear variance due to *local skin distortion*.

In order to produce a model that captures translation with respect to *all* the individual landmarks, each of the locations $\mathbf{x}_i = [x_i, y_i]^T$ is used once as the origin of a polar reference system to represent all other landmarks in a 'radar view'[13].

[11] For Turing patterns the length varies within $0 < \zeta \leq 6$. Penguins, for instance, carry a single feature class only (black spots) while zebras exhibit the full range of six different classes. The detection approach presented for zebras disambiguates, however, just types (L-type or H-type). See Section 2.5 for a full discussion of the feature classes occurring in Turing patterns.

[12] One can clearly find an analogy to the paradigm of 'key points' discussed earlier in this thesis: both concepts combine spatial and semantic features for identification. However, whilst key points are used to recognise object poses (a structurally confined space), typed points operate on arbitrary configurations.

[13] Anchor-referenced translation invariance is qualitatively stronger than global translation invariance: while the latter accounts for overall rigid translations only by using a 2-dimensional parameter space, i-anchor invariance spans an $2i$-dimensional space that captures intra-pattern translations between points.

This transform yields an entire set $\mathfrak{X} = \{\mathfrak{x}^1, \mathfrak{x}^2, ..., \mathfrak{x}^i, ...\}$ of origin-referenced point clouds $\mathfrak{x}^i$:

$$\begin{aligned} &\text{(TYPED, ORIGIN-REFERENCED POINT CLOUD)} \\ &\mathfrak{x}^i = \{[\mathbf{x}_0^i, \mathbf{t}_0], [\mathbf{x}_1^i, \mathbf{t}_1], ..., [\mathbf{x}_j^i, \mathbf{t}_j], ...\} \end{aligned} \tag{5.5}$$

where the upper index i indicates an encoding with respect to the origin $\mathbf{x}_i$. The new landmark coordinates $\mathbf{x}_j^i$ are represented in polar form using the standard transform:

$$\begin{aligned} &\text{(POLAR RE-REPRESENTATION)} \\ &\mathbf{x}_j^i = [\ r_j^i \in \mathbb{R}^+, \gamma_j^i \in [0, 2\pi]\ ]^T = [\ ||\mathbf{x}_j - \mathbf{x}_i||, \arctan(\tfrac{x_j - x_i}{y_j - y_i})\ ]^T \end{aligned} \tag{5.6}$$

where $||.||$ is the vector norm. Figure 5.12 illustrates the richness of this representation.

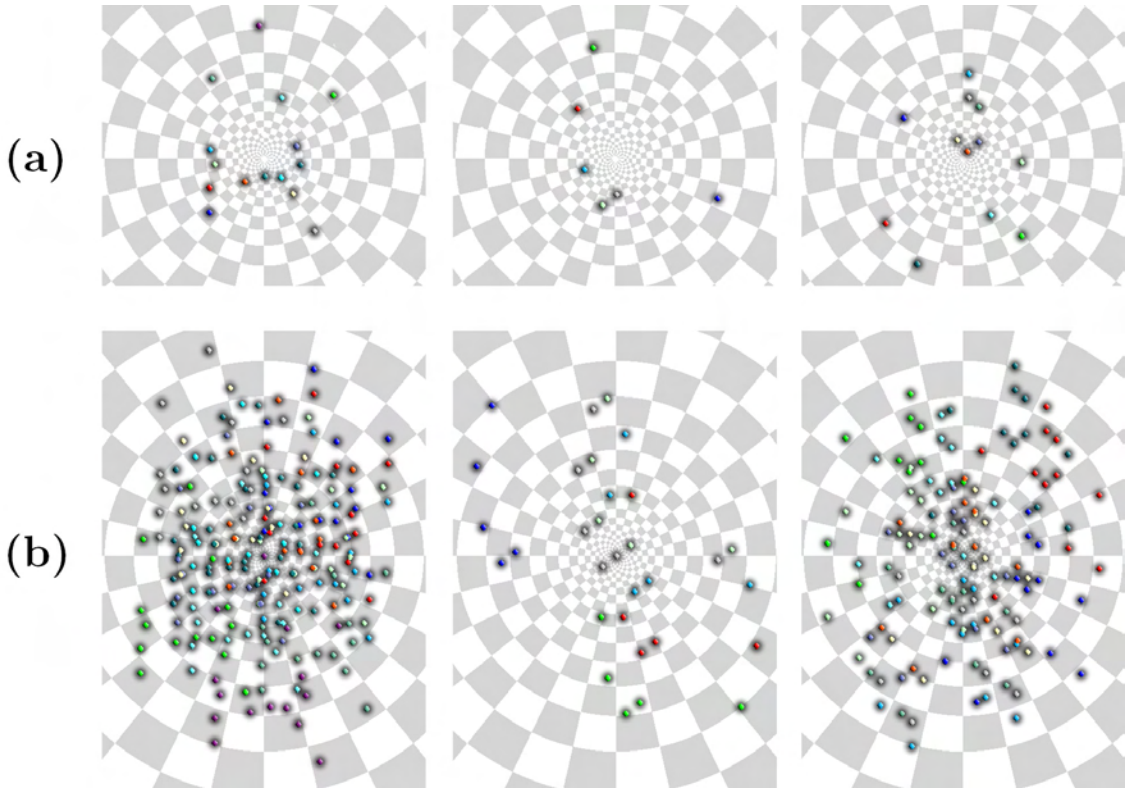


Figure 5.12: ***Origin-referenced Point Clouds.*** **(a)** Parts of patterns of African penguins shown in a $\log^2$-polar system. Note that landmarks are colour-coded. **(b)** Origin-referenced point cloud sets $\mathfrak{X}$ are constructed by multiplexing, using each point as origin once where the landmarks carry the colour code of their origin landmark. The images reveal both the richness of the representation and the symmetry $((r_j^i, \gamma_j^i), \mathbf{t}_j) \in \bar{\mathfrak{x}} \Leftrightarrow ((r_i^j, \pi + \gamma_i^j), \mathbf{t}_i) \in \bar{\mathfrak{x}}$ of the accumulative cloud $\bar{\mathfrak{x}} = \bigcup \mathfrak{x}_i$.

Note that all clouds $\mathfrak{x}^i$ encode exactly the same landmark configuration. However, as a result of the transform, each cloud is now implicitly linked to one specific landmark $(\mathbf{x}_i, \mathbf{t}_i)$, that is its origin.

Clearly, the model is highly redundant, yet it compensates for translational variance *without using statistically derived measures*[14].

Thus, outliers (e.g. false detections, missed landmarks etc.) have *no* impact on the fine localisation of other landmarks, that is a similarity of sub-patterns is preserved with respect to the reference system. Figure 5.13 demonstrates this preservation in a scenario where landmarks are progressively removed from a pattern.

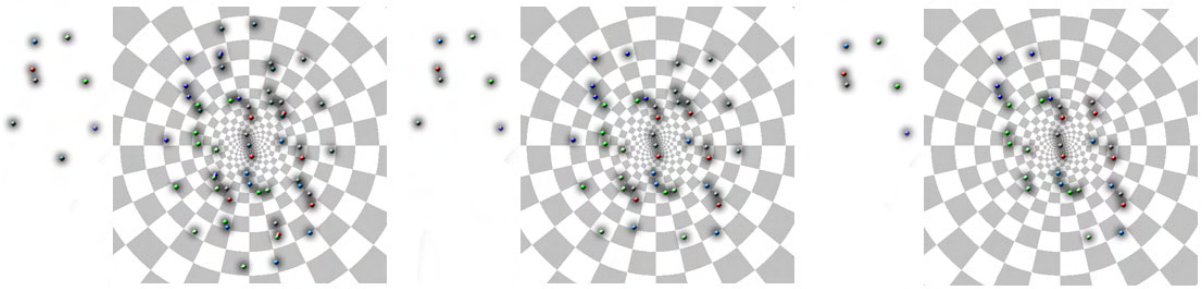


Figure 5.13: ***Structural Consistency of Sub-Patterns.*** **(a)** penguin pattern of 8 landmarks (shown left) and its origin-referenced point clouds; **(b)** removing a landmark produces clouds that form a sub-pattern of the original point clouds, thus, consistently capturing some of the original characteristics of the pattern; **(c)** clouds after removing another landmark;

5.3 Deformation-Robust Matching

5.3.1 Construction of Isotropic Shape Contexts

A robust comparison of two sets of point clouds essentially requires 1) identifying paired (and missing) landmarks between the patterns, before 2) creating a distance measure for comparison, a measure that interprets the differences found in a domain-specific manner.

The first subtask poses a correspondence problem. As known from other areas of vision, e.g. stereopsis [104] or object classification [171], *rich descriptors* generally promote solutions of lower computational complexity for this problem class.

In order to exploit both the richness as well as the invariance features inherent to origin-referenced point clouds, it is proposed to apply spatial histogramming as described by Belongie in the approach of *shape contexts* [20]. In essence, the technique performs a spatial density estimation by capturing the regional densities of each of the point clouds $\mathfrak{x}^i$ (which may encode an arbitrary numbers of landmarks) in polar histograms. Note that all the

[14]Section 2.3 provides a more in-depth discussion.

histograms used have the same layout which is constructed by dividing the polar plane into regularly sub-sectioned segments of concentric rings.

Figure 5.14 illustrates a design where the number of segments per ring is represented by a segment count s and the bin radii are specified by $\mathfrak{r}_i$ (inner radius of a bin) and $\mathfrak{r}_{i+1}$ (outer radius of a bin).

In order to guarantee the structural consistency of the histogram so that all bins are equally proportioned, a polar layout *must be* linear in log-polar$^{\mathbf{2}}$ space (see Appendix B.2 for details). Consequently, the radii of any neighbouring pair of rings are related by an overall *constant factor* a, so that $\mathfrak{r}_{i+1} = a\mathfrak{r}_i$ holds.

In addition, histograms should exhibit a *minimal anisotropy*, i.e. elongated bins[15] are avoided. This feature is modelled by enforcing an equal length of a bin's ring width $w = \mathfrak{r}_{i+1} - \mathfrak{r}_i$ and its median radial height $h = \frac{2\pi}{s}\left(\frac{\mathfrak{r}_{i+1}-\mathfrak{r}_i}{2} + \mathfrak{r}_i\right)$. The equivalence yields a constraint that explicitly defines the factor a as a function of the number s of radial divisions:

(MINIMAL ANISOTROPY)

$$w = h \;\Leftrightarrow\; \mathfrak{r}_{i+1} - \mathfrak{r}_i = \frac{2\pi}{s}\left(\frac{\mathfrak{r}_{i+1}-\mathfrak{r}_i}{2} + \mathfrak{r}_i\right) \;\Leftrightarrow\; \mathfrak{r}_{i+1} = \frac{s+\pi}{s-\pi}\,\mathfrak{r}_i \;\Leftrightarrow\; a = \frac{s+\pi}{s-\pi} \tag{5.7}$$

Considering the recursive application of the constraint, a fixation of the segment number s and *one* ring radius $\mathfrak{r}$ exhaustively defines the structure of a polar histogram with minimal bin anisotropy. Figure 5.14 visualises polar layouts with different isotropic properties.

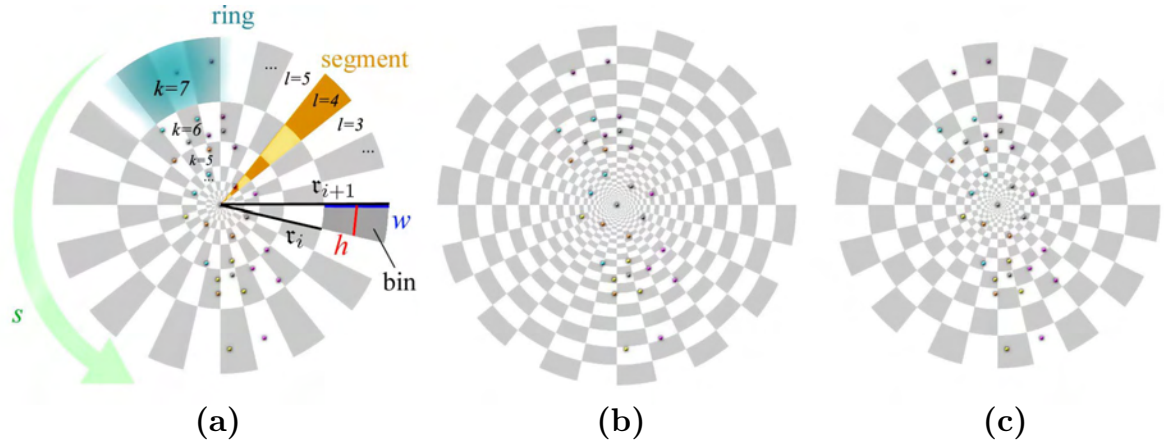


Figure 5.14: ***Histogram Layout and Bin Isotropy.*** **(a)** The image shows point clouds in a sub-sectioned polar domain, which is constructed by intersection of $s = 28$ segments with concentric rings. It is structurally unbalanced since bins resolve azimuth higher than distance ($w \gg h$). **(b)** histogram where distance is quantised higher; **(c)** This image shows an isotropically balanced histogram where, given $s = 28$ segments, the ring radii are stipulated according to $\mathfrak{r}_{i+1} = \frac{28+\pi}{28-\pi}\mathfrak{r}_i = 1.253\mathfrak{r}_i$.

[15] Belongie does not consider this feature in his work on Shape Contexts [20]. A strong anisotropy of histogram bins, however, leads to an uncertainty bias, favouring distortions along only some spatial directions. The application domain does not indicate that such preferences reflect the actual uncertainty.

A histogram $\mathbf{B}^i = (\mathbf{b}^i_{1,1}, \mathbf{b}^i_{1,2}, .., \mathbf{b}^i_{k,l}, ..., \mathbf{b}^i_{m,s})$ captures the characteristics of a landmark $(\mathbf{x}_i, \mathbf{t}_i)$ by encoding the landmark densities in its 'radar view'. Bin vectors $\mathbf{b}^i_{k,l} \in \mathbb{R}^\zeta$ are calculated by weighted integration, that is by collecting evidence $\mathbf{t}_j$ over each bin's region:

(SHAPE CONTEXT CONSTRUCTION)

$$\mathbf{b}^i_{k,l} = \sum_{j=0}^{n} \begin{cases} \mathbf{t}_j & \text{if } \left(k = \left\lfloor \log_{\frac{s+\pi}{s-\pi}}(r^i_j - r) \right\rfloor\right) \wedge \left(l = \left\lfloor 2\pi s/\gamma^i_j \right\rfloor\right) \\ 0 & \text{elsewise} \end{cases} \tag{5.8}$$

where k is the ring index, l is the segment index, s is the overall number of segments, n is the number of landmarks $((r^i_j, \gamma^i_j), \mathbf{t}_j)$ in $\mathfrak{x}^i$, and r is the outer radius of the innermost disc.

While single histograms form a landmark descriptor, the set $\mathcal{B} = \{\mathbf{B}^1, \mathbf{B}^2, ..., \mathbf{B}^n\}$ of all n histograms taken in conjunction, i.e. the *shape context*, describe the full pattern. Figures 5.15 illustrates this technique for encoding a penguin's chest pattern.

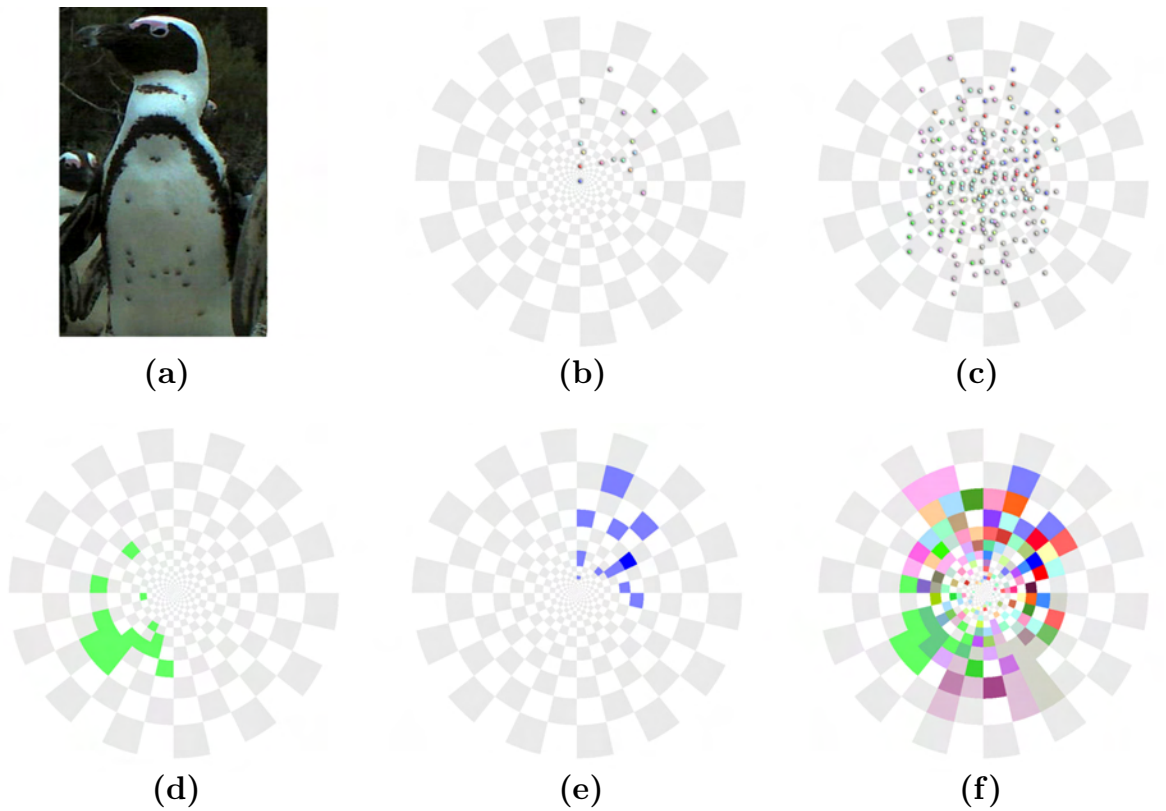
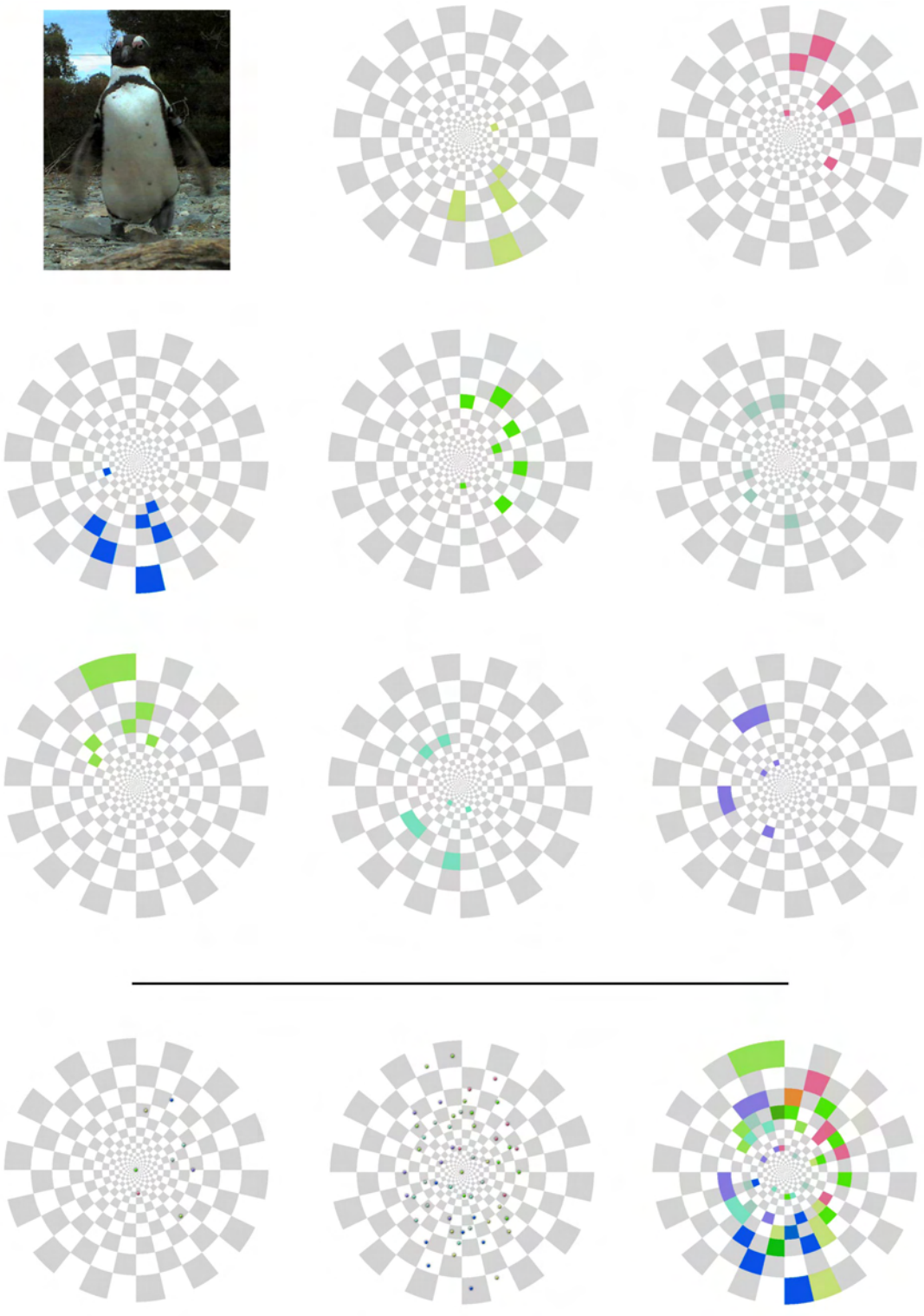


Figure 5.15: ***Application of Shape Contexts to a Penguin's Spot Pattern.*** **(a)** photograph of an African penguin; **(b)** one of the resulting point clouds $\mathfrak{x}^1$ in polar form after normalisation, feature extraction and origin transform; **(c)** the full set $\mathfrak{X}$ of origin referenced point clouds; **(d)-(e)** Histograms $\mathbf{B}^2$ and $\mathbf{B}^1$ form descriptors (colour-coded according to their origin) for two of the landmarks of the pattern; **(f)** The full shape context $\mathcal{B}$ of the pattern visualised by RGB-additive superimposition of all histograms $\mathbf{B}^i$. The image reveals the full complexity of the descriptor. (For reasons of simplicity all $\mathbf{t}^i_j = 1$.) [photographic source of African penguin: I01]



(g)

(g) The images illustrate the representation of an 8-spot pattern (top left) as a shape context, that is a set of 8 polar histograms. Each histogram is colour-coded according to the landmark used as the origin. Landmarks are shown at the bottom left next to the accumulative, origin-centred point cloud (bottom middle) and an RGB superimposition of all 8 histograms (bottom right). [original penguin image source: I01]

5.3.2 Statistics of Biological Deformations

Isotropic shape contexts quantise patterns linearly in $\log^2$-polar space, that is the histogram structure models the uncertainty of landmark positions as linearly increasing per dimension when moving away from the origin [20]. Yet, in how far is this model supported by the actual chest deformation occurring in African penguins?

In order to quantify the extent of pattern deformations (which have not been accounted for by pose correction), a video of standing and walking penguins containing approximately 1500 frames was used for an analysis where pairs of chest spots were tracked in the sequence. After normalisation for the global body pose (as described in Chapter 5), the remaining variance $\bar{\sigma}^2$ in bilateral spot distances was calculated from 22 continuously tracked intervals of 50 frames length. The results are visualised in Figure 5.16. It can be seen that there is a strong positive correlation[16] between the bilateral mean distance (abscissa) and the standard deviation from this mean (ordinate).

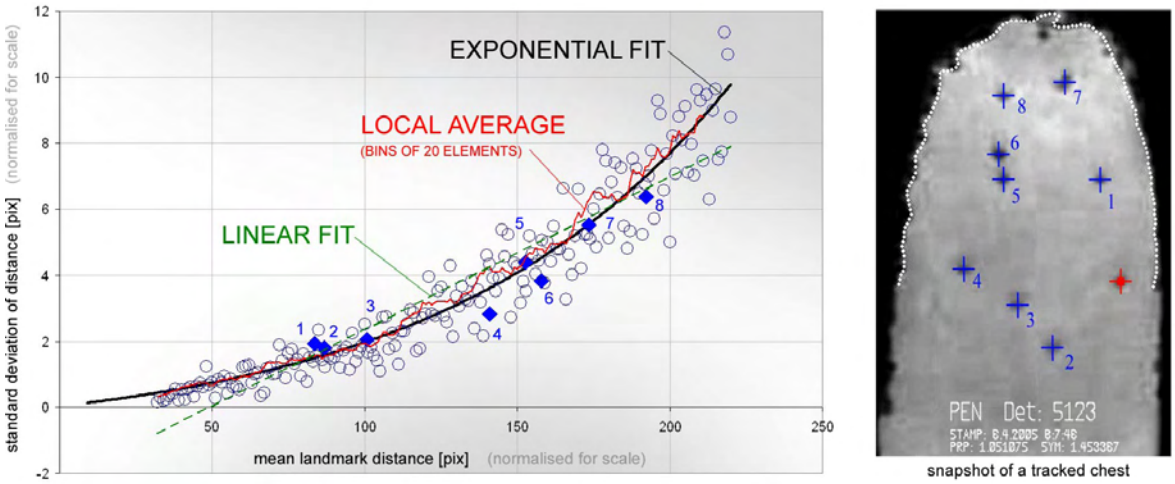


Figure 5.16: **Statistics of Unaccounted Deformation in Penguin Chests.** **(a)** The graph plots the mean distance between landmarks against the positional spread calculated over tracking sequences of landmark pairs. Every 10^{th} measurement is visualised as a circle. The dotted green line shows a linear approximation of the data. The red curve visualises a moving average in buckets over the 20 closest data points in mean distance. The black curve visualises the suggested exponential approximation of the data as given in Eq. (5.9). **(b)** The image shows a snapshot of a tracked chest pattern. The measured positional variance for the landmarks of this specific pattern – taken with respect to the reference (red) – is highlighted in the graph by blue data points.

[16]The variance of distance between landmarks on animal coats is, in general, positively correlated with their mean distance: close landmarks show a more stable, more rigid spatial link than distant features due to cumulative effects of skin elasticity and joint-induced deformation during motion.

Although a linear approximation describes the major characteristics of the data (green line), a more accurate fit can be achieved by modelling the relationship using a slowly growing exponential function:

$$\begin{aligned}&\text{(MODEL OF PRECISION: DISTANCE-VARIANCE RELATIONSHIP)}\\&\bar{\sigma}(d) = \mathrm{e}^{\alpha d} - \beta\end{aligned} \tag{5.9}$$

where d is the mean distance between landmarks and $\bar{\sigma}$ is the estimated standard deviation from the mean distance. The fitting parameters were found to be $\alpha = 1.07 \cdot 10^{-2}$ and $\beta = 0.97$. A precision measure, that is the expected uncertainty about the distance of two spots (captured by $\bar{\sigma}$), can be estimated with the help of Eq. (5.9) given the measured (mean) distance d of the spot pair.

5.3.3 Representation of Landmark Contexts by Distributions

Shape contexts approximate the distance-variance relationship by a linear component, the exponential nature of the term cannot be represented. Shape contexts also suffer from 'feature jumps' between bins particularly when probing sparse penguin spot patterns[17]. As an alternative to rigid histogramming of landmarks, modelling their positions as Gaussians G_j^i provides a smooth representation that also takes the found precision model into account:

$$\begin{aligned}&\text{(GAUSSIAN LANDMARK REPRESENTATION)}\\&G_j^i(\mathbf{x}) = \frac{\mathbf{t}_j}{2\pi\sigma^2}\ \mathrm{e}^{-\frac{1}{2\sigma^2}(\mathbf{x}-\mu)^T(\mathbf{x}-\mu)}\end{aligned} \tag{5.10}$$

where the mean is stipulated at the landmark centre $\mu = \mathbf{x}_j^i$, the scaling is controlled by the landmark's confidence weight $\mathbf{t}_j$, and the standard deviation is set to the expected precision measure depending on the specific radius to the origin, that is $\sigma = \bar{\sigma}(r_j^i)$.

The notion of considering uncertainty distributions addresses the problem of bin jumps by smoothly shifting weights over bins. It also compensates for the non-linearities of the precision measure $\bar{\sigma}$ which are not reflected by the histogram layout. Figure 5.17(a)-(b) visualises the context of Gaussian landmark distributions built from a penguin pattern.

However, instead of applying a computationally expensive transform of the landmarks into dense Gaussians, a distance measure will be used for histogram comparison that models the distributions of landmark uncertainty as cost factors in the matching technique.

[17] Since the shape context technique is of a statistical nature, it is designed for large landmark sets. The technique was introduced by *Belongie* [20] to capture statistics of very rich point clouds, e.g. the shape of characters represented as point clouds.

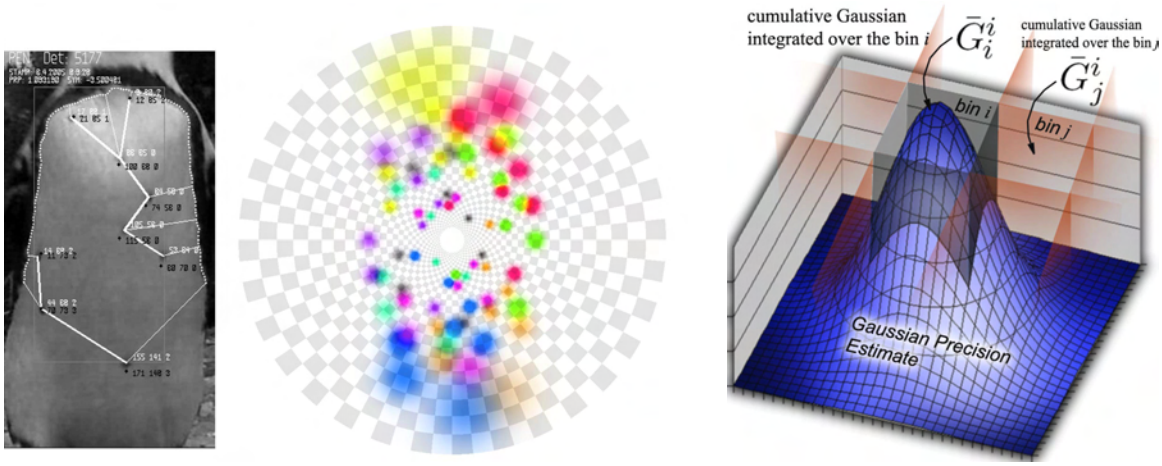


Figure 5.17: ***Distribution Contexts.*** **(a)** detected spot pattern of an African penguin (star constellation view); **(b)** distribution context created by replacing landmarks $\mathbf{x}_j^i$ with Gaussians $G_j^i(\mathbf{x})$ that reflect their expected measurement precision; (Note that the depicted histogram layout of 64 sectors × 40 rings (innermost 32 shown) is used in the actual system.) **(c)** cumulative Gaussians $\bar{G}_j^i$ evaluated over bin areas are used to assign ground matching costs d_{ij} between bins;

5.3.4 Encoding Context Similarity using the Earth Movers Distance

In the pursued methodology landmarks are modelled by their context, that is by distributions of surrounding landmark evidence. Thus, the similarity between any two landmarks can be quantified by comparing the similarity of their context representations.

Essentially, the process of matching two histograms $\mathbf{A}$ and $\mathbf{B}$ is interpreted as a *transportation problem*: the bins of one histogram $\mathbf{A}$ are defined to be the suppliers and the bins of $\mathbf{B}$ are assigned to be the consumers where the bin values $\mathbf{a}_i$ (shorthanded for the value $\mathbf{t}_i$ of a polar histogram bin $\mathbf{a}_{k,l}$) and $\mathbf{b}_j$ represent the supply or demand[18], respectively.

The base cost for transportation between supplier-consumer bin pairs is to be defined by a ground distance measure d_{ij}. Naturally, the estimated, symmetricalised landmark uncertainty is employed for forming this distance measure:

(GROUND DISTANCE BETWEEN BINS)

$$d_{ij} = 1 - \frac{\bar{G}_j^i + \bar{G}_i^j}{\bar{G}_i^i + \bar{G}_j^j} \tag{5.11}$$

where $\bar{G}_j^i = \int_{(\text{bin } j)} G_j^i$ is the cumulative Gaussian which is centred at bin i with $\sigma = \bar{\sigma}(\bar{d}_i)$ and accumulated[19] just over the area of bin j (see Figure 5.17(c)). The measure $\bar{d}_i$ is the Euclidean distance between the centre of bin i and the origin of the histogram.

[18] In the case at hand the 'entity of trade' can be interpreted as 'landmark evidence'.

[19] Note that the cumulative multivariate normal distribution is numerically precalculated for bin pairs.

Fundamentally, the model regulates the cost level d_{ij} by averaging the processes of matching bin i to bin j and vice versa. The cost is consistent with the Gaussian estimation of landmark uncertainty. Note that the measure $d_{ij} \in [0, 1)$ is symmetrical $d_{ij} = d_{ji}$, preserves the identity of indiscernibles $d_{ii} = 0$, and remains zero only at the identity $i \neq j \Rightarrow d_{ij} > 0$.

Operating on this ground cost, the earth mover's distance (EMD), detailed in *Rubner et al.* [172], is defined as the minimal overall cost of transporting a maximum amount of entities from the suppliers to the consumers. This cost is associated to some specific flow $\mathbf{F} = [f_{ij}]$ from $\mathbf{A}$ to $\mathbf{B}$ that proves to be most cost-efficient. A matrix entry f_{ij} in $\mathbf{F}$ reflects the flow from bin i in $\mathbf{A}$ to bin j in $\mathbf{B}$. Entities that are not transported due to different spot count in $\mathbf{A}$ and $\mathbf{B}$ are modelled by a residual U. This extended earth mover's distance[20] yields:

(MINIMAL TRANSPORTATION COST)

$$\mathrm{EMD}_{\mathbf{A},\mathbf{B}}(\mathbf{F}) = \min\left(\frac{1}{n+U}\left[\sum_i\sum_j\left(\underbrace{d_{ij}}_{\text{cost}}\underbrace{f_{ij}}_{\text{flow}}\right) + \underbrace{U}_{\text{residual}}\right]\right) \tag{5.12}$$

where 1) the flow must be strictly directional, that is $f_{ij} \geq 0$, and 2) supply and demand are bounded by the bin weights, that is $(\sum_j f_{ij} \leq \mathbf{a}_i) \wedge (\sum_i f_{ij} \leq \mathbf{b}_j)$, and 3) the maximally demanded supply must be shipped, thus satisfying $n = \sum_i \sum_j f_{ij} = \min(\sum_i \mathbf{a}_i, \sum_j \mathbf{b}_j)$, and the residual term $U = \max(\sum_i \mathbf{a}_i, \sum_j \mathbf{b}_j) - n$ establishes a cost for unmatched landmarks.

5.3.5 Pattern Authentication by Measuring Landmark Association Costs

In essence, the earth mover's distance assigns a measure $\mathrm{EMD}_{\mathbf{A}^i,\mathbf{B}^j}$ to a given pair of landmarks (i.e. histograms $\mathbf{A}$ and $\mathbf{B}$) that reflects their degree of structural similarity.

In order to establish 1) an assignment of landmarks between patterns and, 2) a distance between two entire histogram sets $\mathcal{A} = \{\mathbf{A}^1, ..., \mathbf{A}^i, ..., \mathbf{A}^m\}$ and $\mathcal{B} = \{\mathbf{B}^1, ..., \mathbf{B}^j, ..., \mathbf{B}^n\}$ on the basis of landmark distances $\mathrm{EMD}_{\mathbf{A}^i,\mathbf{B}^j}$, a *bipartite graph matching problem*[21] has to be solved: in the $m \times n$ matrix $\mathbf{M} = [\mathrm{EMD}_{\mathbf{A}^i,\mathbf{B}^j}]$ containing pairwise landmark distances, a 1-to-1 assignment vector $\mathbf{m} = [i, j]^{\min(m,n)}$ between rows i and columns j has to be found so that the overall assignment costs $C_{\mathcal{A},\mathcal{B}} = \sum_{(i,j)\in\mathbf{m}} \mathrm{EMD}_{\mathbf{A}^i,\mathbf{B}^j}$ are minimal[22]. As suggested

[20] The transportation simplex algorithm (detailed in *Hillier and Lieberman* [95]) is used to calculate the earth mover's distance where the algorithm by *Russell* [174] is used to initialise the simplex method.

[21] An undirected graph $G = (V, E)$ is bipartite if there exists a partition $V = \mathcal{A} \cup \mathcal{B}$ so that every edge in E is of the form $(\mathbf{A}^i \in \mathcal{A}, \mathbf{B}^j \in \mathcal{B})$. The histogram sets $\mathcal{A}$ and $\mathcal{B}$ are interpreted as the vertex sets of the two subgraphs whilst the distance matrix $\mathbf{M}$ describes the edges between them.

[22] Note the similarity of this assignment problem to the flow optimisation used in the earth mover's distance: both problems constitute assignment tasks in bipartite graphs, however, the EMD allows for splitting a flow

by *Belongie* [20], the Hungarian method by *Kuhn* [115] is used to solve the assignment problem. The resulting cost $C_{\mathcal{A},\mathcal{B}}$ is interpreted as a distance measure in pattern space for the task at hand. It is non-negative: $C_{\mathcal{A},\mathcal{A}} \geq 0$, symmetrical: $C_{\mathcal{A},\mathcal{B}} = C_{\mathcal{B},\mathcal{A}}$, and preserves the identity: $C_{\mathcal{A},\mathcal{A}} = 0$. For authentication, a threshold ρ is finally employed to interpret the distances in a binary fashion, that is to assign to pairs of patterns either the label '*different*' or '*authentic*' if $C_{\mathcal{A},\mathcal{B}} < \rho \cdot w_{m,n}$. In the latter case $\mathcal{A}$ and $\mathcal{B}$ are identified to represent the same individual. Note that $w_{m,n}$ is a weight (see Section 5.5) which depends on the spot counts m and n, and that the choice of threshold ρ defines the working point in ROC space. The next section now describes tests that estimate the performance of the technique.

5.4 Results from a Real-world Prototype

5.4.1 Setup in an African Penguin Colony

After preliminary tests had been completed at Bristol Zoo, the identification system was taken to a *native colony of African penguins* situated on Robben Island, South Africa (see Appendix B.3 for details on the study site) in order to study the performance in a natural habitat. For the experiments an environmental ethernet camera was commissioned at one of the major penguin walkways[23] and used to stream images to a laptop computer placed in a distant hide. Figure 5.4.1 depicts this totally non-invasive acquisition scenario.

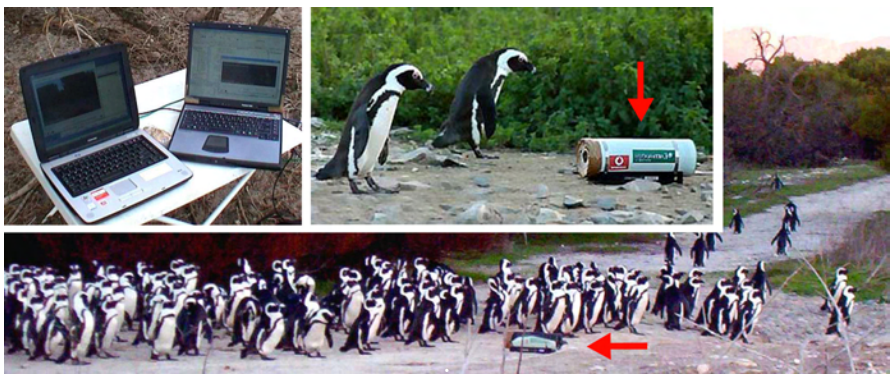
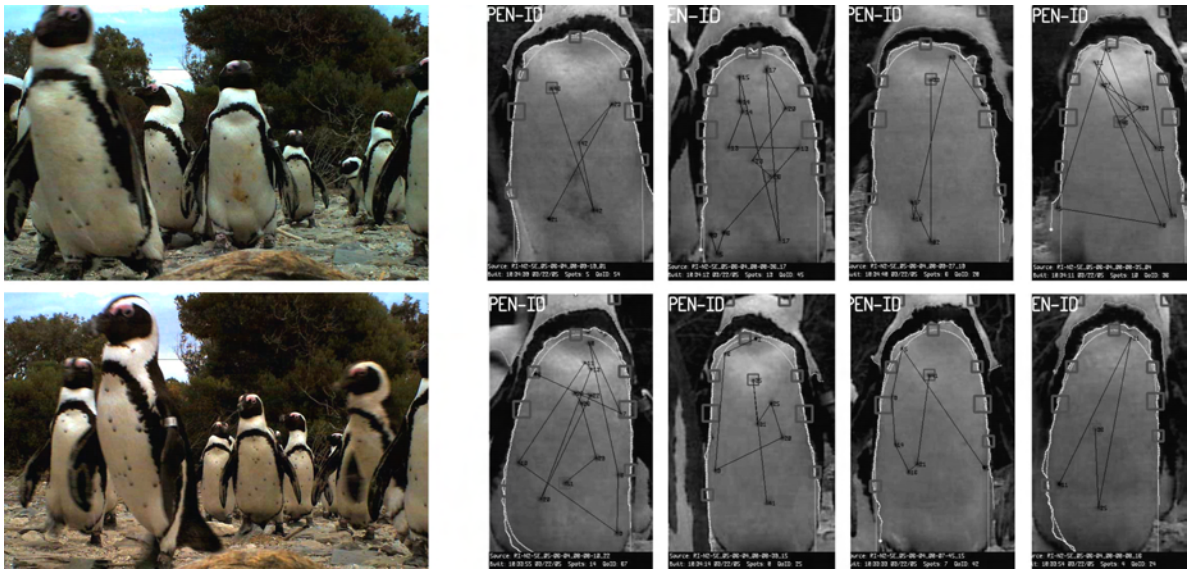


Figure 5.18: ***Monitoring Penguins in their Natural Habitat.*** The camera (see arrow) was placed at the 'penguin highway' N2 on Robben Island. The images show birds that arrive from the sea and spread out into different directions towards the colony. The animals accept the device as part of their natural environment and, thus a non-invasive observation without any human presence is possible. The captured images are then streamed to the control laptops in a hide. [photographs: I01]

to several targets (=consumers) whilst the landmark matching retrieves a binary, 1-to-1 assignment **m**.

[23] African penguins exhibit a specific behaviour: birds gather at particular paths and walk in groups to the sea for foraging trips; they later return to the colony using the same network of routes. These hubs constitute suitable locations for monitoring large fractions of the population on a regular basis.

The high resolution imagery produced by the camera[24] allowed for a sufficient spatial quantisation of spots in distances between approximately 1 and 5 meters. Two representative frames as taken by the field camera at the penguin highway are shown in Figure 5.19(a). Notice the presence of occlusions, motion blur and dirt on the bird's chests.



(a) **(b)**

Figure 5.19: ***Acquisition and Extraction of Patterns.*** **(a)** Two representative sample frames as submitted by the field camera. **(b)** The images depict a selection of extracted penguin patterns from the sample set of 1000 patterns used for the performance tests (and as masters profiles). In these visualisations the chest spots are projected onto the original luminance image patches. Spots are shown as the vertices of a polygon ordered by the distance to the chest outline. The snake lines used to segment the chest area are also shown. [original image sequences underpinning visualisations: I01]

In order to estimate the performance of the identification system in this natural environment, the species detector[25] was used to recognise image patches containing penguins in near-frontal pose where the full spot pattern is visible. Figure 5.19(b) shows a selection of these extracted chest patches with projections of the detected spot markings. A sample set of 1000 pose-corrected chest patches formed the input for the performance study.

A ground truth for the performance experiments was created for the identities of detected birds by manually associating the $n = 1000$ sample image patches with one of 114 individuals[26]. The landmark detector was finally applied to each of these patterns and,

[24]The camera produces an MJPEG stream at 6-8 fps resolved at 1280×856 pixels compressed at level 75%.

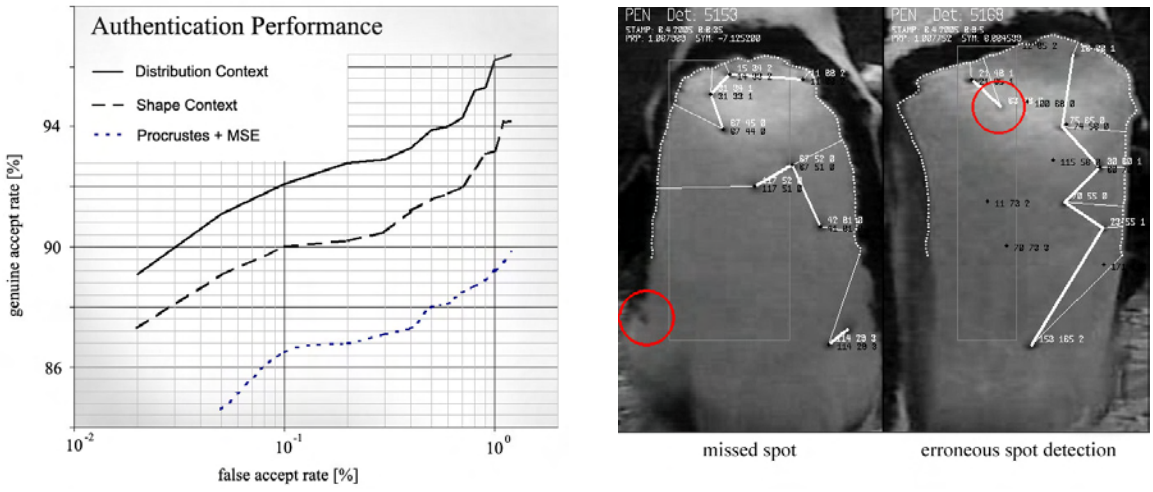
[25]Notice that in the particular sequence chosen for the trial the species detector did not produce any false positives, that is image patches that do not contain a penguin. However, lighting conditions were good (i.e. diffuse daylight, no rain or fog).

[26]Thus, on average about ten images were captured of each bird. Note that the 114 individuals that were detected in a suitable, near-frontal pose represent only $\approx 33\%$ of the 341 birds which passed by the

successively, the measured spot configurations were transformed into histogram sets. It needs mentioning that the process of spot extraction often produced imperfect results, that is landmarks were missed and erroneous detections occurred due to shadows, dirt patches, disordered plumage, image compression artefacts, motion blur, and/or proximity to the chest outline (Figure 5.20(b) depicts two examples of erroneous detections). However, even these inaccurate spot extractions were used in order to produce a more realistic performance measure.

5.4.2 Cross-Over Authentication

First, the **cross-over authentication** performance was estimated. Essentially, each test pattern was compared to all other patterns so that either the label '*authentic*' or '*different*' could be assigned to all $\frac{(n-1)n}{2} \approx 5 \times 10^5$ pattern pairs. Performing the calculations at different identity thresholds ρ yielded a receiver operating characteristic (ROC), which is depicted in Figure 5.20(a) as a solid curve in the region of interest in ROC space.



(a) **(b)**

Figure 5.20: ***Cross-Over Authentication Performance.*** **(a)** Receiver operating characteristic (ROC) curves of the identification performance of the three methods tested. It is suggested to operate this authentication system below a genuine accept rate of GAR≈ 92% to keep the false acceptance rate below 1 false positive in 1000 trials; **(b)** The images show failures (red) of the spot detector in form of a missed spot (left image) a erroneous detection (right image). [imagery used: I01]

In order to place the performance curve into context, the proposed matching system was

camera throughout the approx. 20 min test sequence (consisting of different segments of activity filmed over 5 hours). Notably, most birds were occluded by other conspecifics or walking by the camera sideways (see Figure 5.19(a)). However, the particular video sequence shows good acquisition conditions (e.g. fair lighting, little specular reflection on birds). In a truly arbitrary setting (sunset, rain, fog etc.), I believe the capture rate will be reduced to somewhere between 10% and 20% of the passing birds.

evaluated alongside two other techniques, that is 1) Belongie's original version of shape contexts [20] for landmark comparison (dashed curve), and 2) a rigid alignment matcher that uses the Procrustes algorithm [107, 108] for pose correction before calculating the mean squared error (MSE) over all closest landmark pairs (dotted curve).

It can be seen that at a false accept rate of $10^{-1}\%$ the proposed matching scheme performed at about 92% sensitivity, that is an increase of 2% with respect to shape contexts. This improvement persists over the tested spectrum of the characteristic.

Thus, explicitly modelling the spatial uncertainty of landmarks in a flexible matching scheme results in a measurable increase in performance for the task at hand. The rigid landmark matcher was clearly outperformed by both context matchers; however, it still achieved about 86% genuine acceptance rate at the comparison point.

Experiments with parameters of the proposed matching scheme, i.e. histogram granularity and pose correction method, yielded further benchmarks as summarised in Table 5.2.

Histogram Resolution	**GAR***
96 × 60 bins	92.09%
▶ 64 × 40 bins ◀	92.08%
48 × 30 bins	91.44%
32 × 20 bins	90.72%

(a)

Pose Correction Method	**GAR***
▶ least squares affine model (see Chapter 5) ◀	92.08%
3-key-point affine chest model (see Chapter 5)	90.25%
Procrustes normalisation of spot patterns	88.38%
none – comparison on the detected patch	67.93%

(b)

Table 5.2: **Cross-over Authentication Performance at the Working Point. (a)** Performance comparison of distribution contexts at the working point using different histogram resolutions defined by sectors×rings. All benchmarks for genuine accept rates (GAR) are retrieved at a constant false accept rates of FAR= $10^{-1}\%$. Note that at the top end a significant increase in histogram resolution has (almost) no effect on the performance. Thus, the lower resolution is chosen for operation (indicated by triangles). **(b)** The table shows the effect of using different pose correction schemes on the authentication performance. It can be seen that some form of normalisation proves highly important even in near-frontal poses. A rigid Euclidean normalisation (Procrustes algorithm) before using distribution contexts performs well but lacks the specific relevance to the domain provided by surface models of the chest. [penguin imagery used for the experiments: I01]

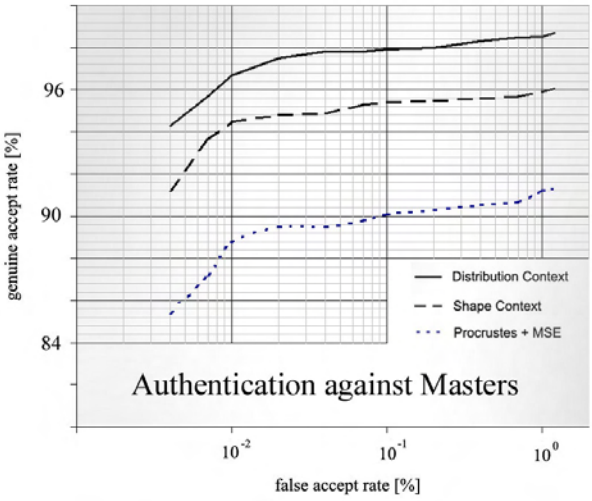
Basically, a sensitivity above 90% on the ROC curve must be traded for a false accept rate above 1 in about 8000 (see Figure 5.20(a)). This score sits above benchmarks for human voice and facial identification systems [155, 173]. However, coat patterns lack the richness of iris and retina entities (containing thousands of features) which are exploited [103, 173] by human biometrics to confidently authenticate up to *millions* of individuals (see Table 2.1). It will be shown now that, to enhance its performance, the coat pattern identification system can be applied in an administrated scenario, as is commonplace in biometric applications.

5.4.3 Estimation of the Administrated Identification Performance

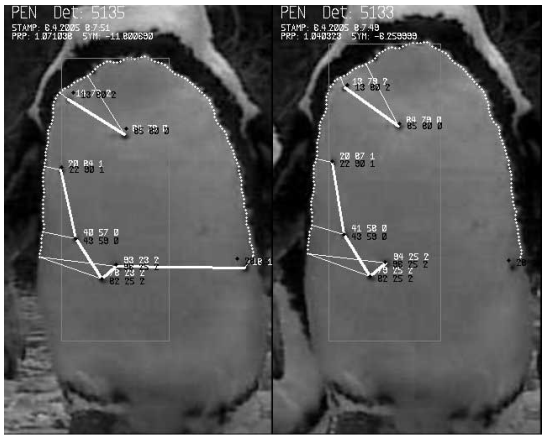
A careful observation of erroneous pattern pairings revealed one major cause for the failure of matching operations: authentication was often flawed in cases where spots were missed (or introduced) by the landmark detector in *both* of the compared patterns. The authentication algorithm then operated on two extended sub-patterns neither of which represented the true pattern. The problem can be addressed by ensuring the authenticity of one of the patterns.

Moving towards a simulation of a realistic identification system, $m = 80$ of the 114 test individuals were each represented by a *master profile*, that is one prototypical registration which was manually checked[27] to contain complete spot configurations. The entirety of master profiles then formed a (small) population database.

In order to estimate the **identification performance** in this scenario, it was required to authenticate each pattern against the entire population database. Thus, the remaining test patterns (non-masters) were compared to each of the master profiles yielding $m(n-m) \approx 7 \times 10^4$ comparisons. Again, three different matching techniques were investigated for authentication. The use of different accept thresholds ρ yielded the receiver operating characteristics which are depicted in Figure 5.21(a).



(a)



authenticated correctly despite missed chest spot
in a very sparse pattern

(b)

Figure 5.21: ***Administrated Performance Characteristic.*** **(a)** Receiver operating characteristic (ROC) curves of the identification performance of the three methods tested; **(b)** Two measurements (master left) where a spot was missed in the detection. This particular pair was matched correctly. The missed feature had a low confidence weight $\mathbf{t}_i$ (due to the peripheral spot location).

[27] Note that the 34 individuals not used as masters simulate unregistered animals, essentially acting as distractors. Also notice that the single supervision step during enrolment is common practice in fingerprint and iris biometrics since it is required only once and ensures a high quality master profile.

The administrated ROC curve clearly illustrates that the identification system proved robust showing, for instance, only about 1 in 30000 false authentications to master profiles at 96% genuine accept rate[28]. This performance permits for a recovery of identities beyond the human capabilities of correctly reading flipper band numbers (estimated 5% false readings and limited to only the banded population), which constitutes the currently applied technique to gather sighting data of penguins in most parts of Southern Africa.

5.4.4 Practical Use in a Future Application: Identification, Coverage and Drawbacks

Most essentially, the proposed technique can generate place- and time-stamped sighting information. A sample *identification diagram* is shown in in Figure 5.22. Individual identifications are plotted there against a time line where identifications (black markings) of 12 individuals (ordinate) are registered over a duration of circa five minutes (abscissa).

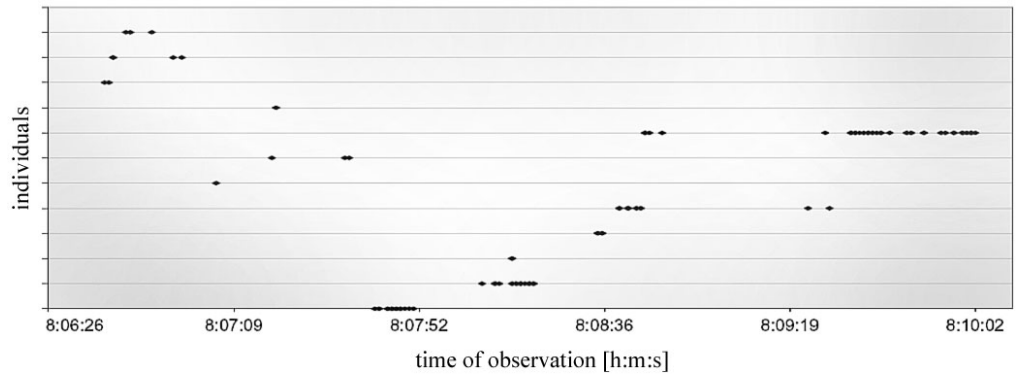


Figure 5.22: ***Example of an Identification Diagram.*** The plot shows a detection time line from the penguin highway N2 on Robben Island collected over about 5 min. [underlying image data: I01]

It can be seen that some penguins passed the camera quickly whilst others remained standing in front of the acquisition device for minutes. The majority of penguins, however, passed without being identified mainly due to occlusion. The resulting low rate of immediate cov-

[28]Effectively, the introduced supervision step allowed for a relaxation of the identity threshold ρ whilst achieving the same level of FAR. As a result, the sensitivity(=GAR) increased and the ROC curve shifted upwards, thus the performance increased. For instance, at FAR=10^{-2} the system now labelled only 7 of 72803 negatives as positives whilst 26 of the 797 positives were rejected, yielding a sensitivity of 96.7%. It must be noted though that the resulting measure built over the 73600 comparisons represents only a coarse approximation of the performance in arbitrary scenarios. The benchmark *will degenerate* to some degree if the master database grows by several orders of magnitude. A robust performance projection for truly large scale scenarios is still something of an art also in human biometrics since it is highly dependent on the specific test sets and acquisition conditions [103]. Only manually annotated databases containing millions of samples allow for producing a truly realistic benchmark. However, a continuation of the penguin recognition project on Robben Island will allow for building a large scale database that will help with stipulating benchmarks that reflect the performance on an entire population. In addition, a number of extensions are planned to allow further increase of the identification performance (see Section 6.3).

erage, i.e. the fraction c of birds identified in a stream of penguins, constitutes a major limitation of the identification methodology. Based on manual counts of passing and identified, it can be estimated that only 10% to 35% of passing birds can be identified by the system, depending on the environmental conditions. The coverage $f_c(t)$, i.e. the fraction of the population identified at least once from the trial start to time t, can be modelled as:

(POPULATION COVERAGE)

$$f_c(t) = c \sum_{i=1}^{t} (1 - f_c(i-1)) \tag{5.13}$$

where $c \in [0, 1]$ is the fraction of the population identified per day, t is the time in days after the trial start, $f_c(0) = 0$ so that the coverage is zero at the start of the trial, and birds are assumed to be filmed in a truly random fashion. Figure 5.23 illustrates the development of coverage over time in accordance to this model. Thus, a total of 99.9% of the daily filmed population could be identified within a month given an observation rate of $c = 20\%$. Consequently, given long term trials the system appears to have the potential to assist in visually capturing the entire subpopulation that regularly passes the field camera.

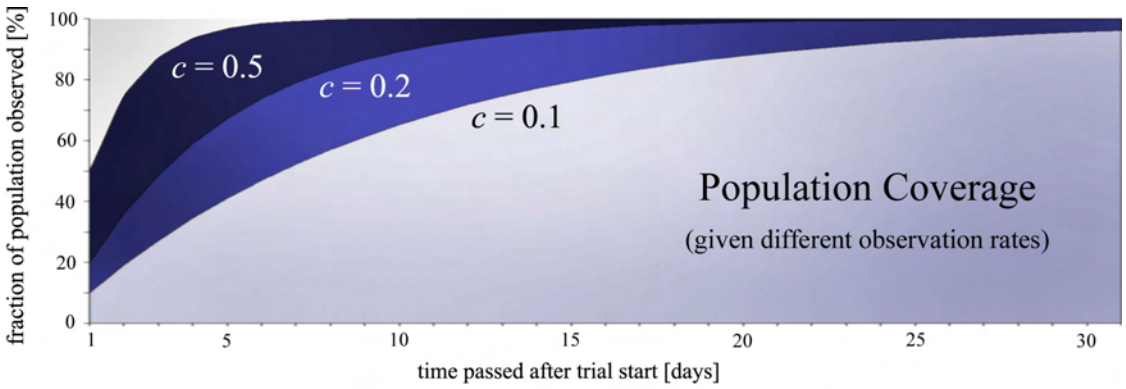


Figure 5.23: ***Population Coverage vs. Duration of Observation.*** The graph depicts the development of population coverage with an increasing duration of the trial period for different daily observation rates. The diagram, therefore, visualises the recursive model defined in Eq. (5.13).

One has to bear in mind that all the performance estimates discussed are based on video sequences taken under good acquisition conditions. An exposure to the full dynamics of environmental conditions in Africa (including the high dynamic range of lighting intensities) will reduce the performance of the recognition system, especially the one for the components that rely on appearance, i.e. the key point detector and the spot extractor.

5.4.5 Preliminary Performance Study on Plains Zebras

In order to showcase the applicability of the approach to striped animals, the developed identification system has also been tested on a *small set* of digital side photographs of plains zebras. Figure 5.24 illustrates a selection of this sample set.

Figure 5.24: ***Plains Zebra Image Collection.*** Depicted are 28 representative samples from the image collection showing side-views of plains zebra. [image source I02]

The collection covered 200 images of 47 individuals each depicted in various different shots. Therefrom, the biometrically relevant side pattern patches were extracted using affine projection. Subsequently the extracted textures were parsed for singularity points using the techniques described in Section 5.2. In order to simulate a realistic identification scenario, 30 out of the 47 individuals were chosen to create a small population database where one observed configuration of singularity points was stored as a prototypical profile for each of the individuals. All database entries were manually verified to ensure all singularity points were correctly detected. Figure 5.25 depicts samples from the database built.

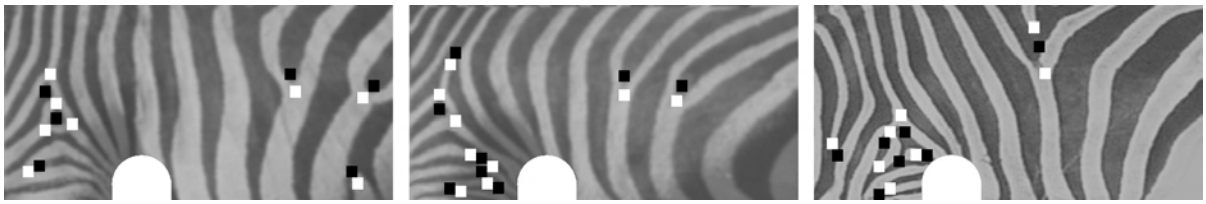


Figure 5.25: ***Samples from the Individual Database.*** Three representative samples from the individual database of plains zebra. As before, white blocks show H-type singularity points while L-type points are shown as black blocks. [original image source I02]

In order to measure identification performance, the remaining 170 test patterns (not used in the database) were compared to each of the 30 master profiles stored in the database. The technique discussed in Section 5.3 (shape context+earth movers distance) then yielded a matching decision for each of the $30 \times 170 = 5.100$ comparison operations.

A true positive was counted whenever a test pattern was successfully matched with a master profile in the database representing the correct individual. Failure to match with the correct master profile resulted in a false negative. Matching a test pattern against the master profile of a wrong individual counted as a false positive decision, all other cases represented true negatives.

The use of different accept thresholds ρ for the earth movers distance yielded a receiver operating characteristic that pinned down the system performance in ROC space. The resulting ROC curve is depicted in Figure 5.26.

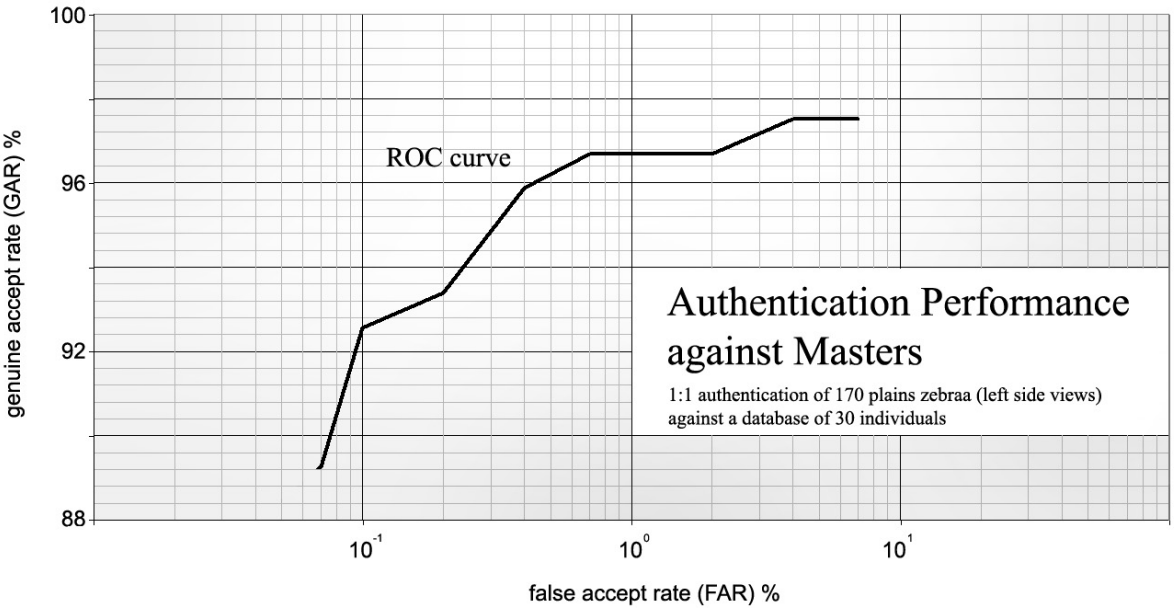


Figure 5.26: ***Zebra Authentication Performance.*** Receiver operating characteristic (ROC) curve of the identification performance measured for the application of the identification system to plains zebras. Note that the coarseness of the curve is due to the small size of the sample set.

It can be seen that the system shows a reliable authentication performance where, for instance, at 0.1% FAR (i.e. 1 in 1000 wrong authentications) the system still achieves about 92% coverage. As an example, Figure 5.27(a)-(b) illustrates a pair of matched images together with the extracted texture patch with the detected singularity points superimposed.

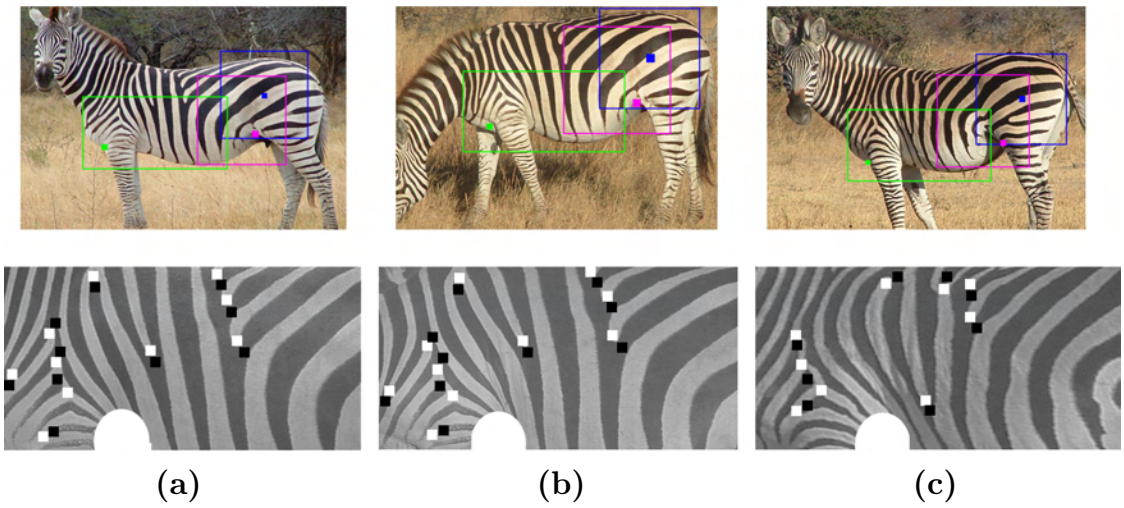


Figure 5.27: ***Successful Zebra Identification.*** The first row shows the original zebra images with key points used for affine back-projection superimposed. The second row depicts these back-projected texture patches together with the detected singularity points superimposed. **(a)** One of the zebra entries from the database used as master profile. **(b)** Example of a test image successfully matched with (a). Note the minor deformation of the pattern due to a different neck and leg pose. **(c)** Example of the closest negative rejected. The matching scheme successfully rejects this test pattern as a match for (a) despite some similarity of sub-patterns. [image source I02]

Authentication errors were mainly due to misdetections of singular points which were often caused by deep shadows or white bleeding of very bright image areas. Figure 5.28 depicts an example of shadow-induced failure of the singularity detector. It can be seen that a hard shadow causes 'phantom singularities' which are not distinguished from pattern-inherent singularity points by the detector built.

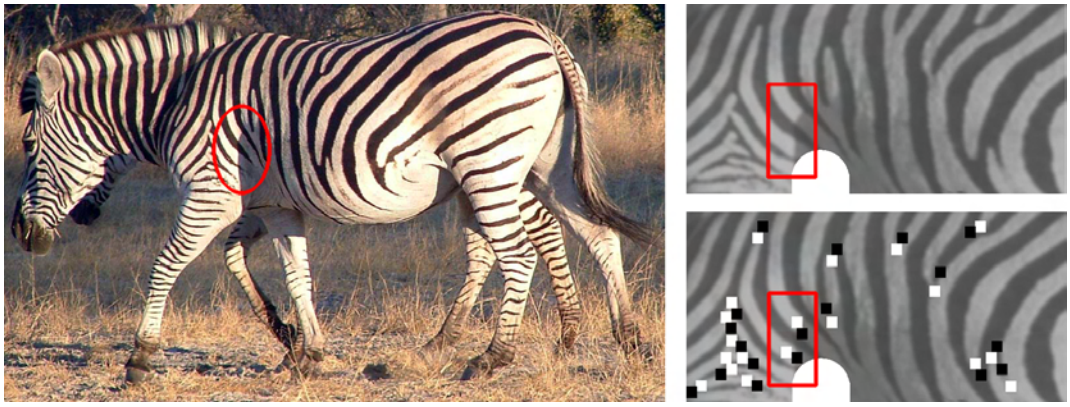


Figure 5.28: ***Shadow-induced Detector Failure.*** The left image shows a sample where the depicted zebra shows a hard shadow (red ellipse) behind the scapular area. The shadow (see red rectangle) is prominently visible in the luminance channel of the extracted texture patch (top right) and causes the erroneous detection of 'phantom singularities' (bottom right). [original image source I02]

It should be noted that the performance estimates discussed for zebras are based on only 200 images taken as full side views under relatively good acquisition conditions.

An application of the technique to photographic material that reflects the full dynamics of environmental conditions in Southern Africa will reduce the performance significantly, particularly the one of the key point detector and the singularity point extractor.

Section 6.3 on future work will focus on the problem classes that lie ahead for transforming the presented prototype system into a generic, environmental vision tool for monitoring animals.

5.5 Exploring a Theoretical Model of Coat Pattern Uniqueness

5.5.1 A Stochastic Model for Landmark Patterns

The discussed identification system exploits, due to its technical limitations, only a fraction of the uniqueness inherent to the landmark configurations in coat patterns.

Can the physically present individuality be modelled? Based on this model, how many individual animals could be confidently disambiguated? To this end, no work has been published for animal coats. In fact, the exact degree of uniqueness in human fingerprints is unknown – it has merely been approximated by stochastic models[29].

In conclusion of this chapter it will now be shown that the ideas of a state of the art uniqueness model in human fingerprint biometrics, i.e. an approach by *Prabhakar* [159], is applicable to coat patterns.

In particular, the focus will be on a stochastic model that estimates the uniqueness captured by landmarks of coat patterns. The model will be built over *landmarks only* which are approximated as *truly randomly configured* patterns with respect to the partitions of some surface area.

[29] More than a century ago, Galton [73] first addressed the problem of random biometric correspondence by estimating the number of distinguishable human fingerprints to be $\frac{1}{\gamma!\beta}\alpha^{(A/B)} = 1.45 \times 10^{-11}$ where he identified $\beta = 16$ types of fingerprints (e.g. arch, right loop, whorl etc.) and $\gamma = 256$ ridge categories. He divided the fingerprint area A into $24 = A/B$ bins each of which he could reconstruct with a chance of $\alpha = 0.5$. Most biometric individuality models developed in the 20^{th} century (including Gupta [86], Cummins [43], Stoney and Thornton [191]) followed Galton's concept of approximating pattern cardinalities by a basic exponential function of the form ax^y. A recent study can be found in *Prabhakar* [159].

The model is constructed over a domain of biometrical interest (e.g a penguin's chest) where $A \in \mathbb{R}_+$ represents the area of this domain (shown in red in Figure 5.29(a)) partitioned into bins of size B (shown in blue). Each bin may (or may not)

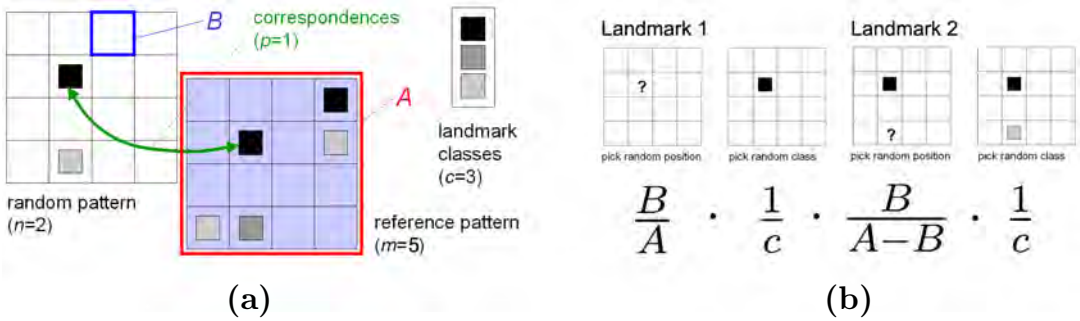


Figure 5.29: ***Basics of the Stochastic Model.*** The images explain the model by example: **(a)** A reference pattern with 5 landmarks is compared to a randomly created pattern with $n = 2$ landmarks, each of which is drawn from 3 available classes. Both patterns cover the same area A divided into bins of size B. One correspondence (green) exists between the two patterns. **(b)** A random pattern is generated by progressively picking classes and (unused) landmark positions. Terms below the images reflect the odds of each step. The sequence of picking the landmarks is not fixed and hence, there are $n!$ ways (permutations) of assembling the same pattern, that is $2! = 2$ ways for the pattern shown where either of the two landmarks could be picked first.

host one out of $c \in \mathbb{N}$ different landmark types (shown as blocks of different luminance). The stochastic process of synthesising an n-landmark pattern is modelled as 1) randomly selecting n different bins, and 2) randomly picking one of the c landmark classes for each of the chosen bins. As graphically illustrated in Figure 5.29(b) on a basic example, the chance $\mathcal{P}(\mathbf{X}_n)$ of creating a particular n-landmark pattern $\mathbf{X}_n$ is then given by:

$$\mathcal{P}(\mathbf{X}_n) = n! \underbrace{\left(\frac{B}{A}\frac{1}{c}\right)\left(\frac{B}{(A-B)}\frac{1}{c}\right)\ldots\left(\frac{B}{(A-(n-1)B)}\frac{1}{c}\right)}_{n \text{ factors}} \tag{5.14}$$

Substituting $A/B = R$ and simplifying the above reveals the probability function to be a inverse binomial coefficient scaled by a term that grows exponentially with n:

$$\mathcal{P}(\mathbf{X}_n) = n! \underbrace{\left(\frac{1}{cR}\right)\left(\frac{1}{c(R-1)}\right)\ldots\left(\frac{1}{c(R-n+1)}\right)}_{n \text{ factors}} = \frac{n!(R-n)!}{c^n R!} = c^{-n}\binom{R}{n}^{-1} \tag{5.15}$$

The model assumes that all specific n-landmark patterns (i.e. given the same c and R) are equally likely to occur. Therefore, the space of all possible n-landmark patterns follows to be of cardinality $C_n = \mathcal{P}(\mathbf{X}_n)^{-1} = c^n\binom{R}{n}$ satisfying $\sum_{\mathbf{X}_n} \mathcal{P}(\mathbf{X}_n) = C_n\mathcal{P}(\mathbf{X}_n) = 1$. Figure 5.30 visualises the cardinality C_n as a manifold in the parameter space.

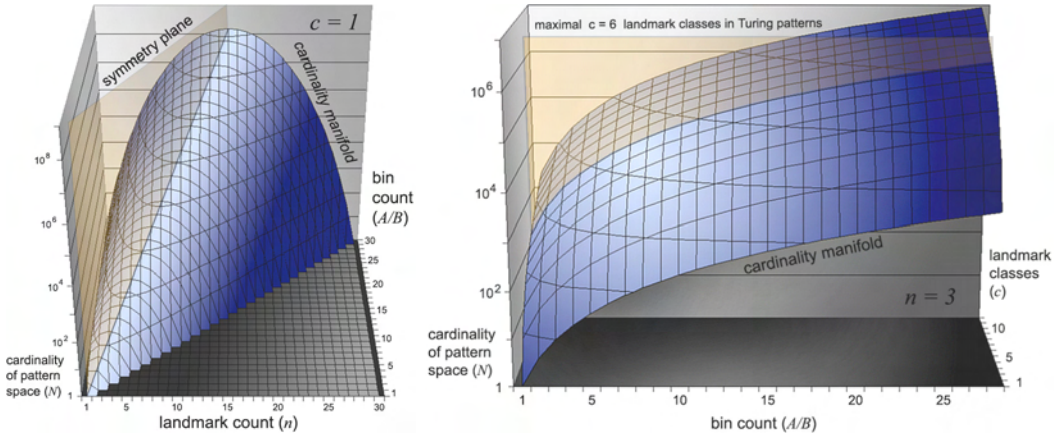


Figure 5.30: ***Cardinality of the Pattern Space.*** The images visualise the cardinality C_n. Naturally assuming $n \leq R$ (no more landmarks than bins), the cardinality C_n can be intuitively explained as the set of all combinations of the c feature classes over all n-element subsets found in the set of R bins. **(a)** Fixing c and plotting C_n against the bin count $R = A/B$ and the landmark count n reveals a symmetry of the cardinality function (when disregarding the special case $R = n$) indicated by a plane. The plane dissects the cardinality manifold along a curve that describes the landmark count n with maximal generative potential given a certain R. Individuals that carry this specific number of landmarks are, potentially, the most unique individuals amongst the group of animals with an equivalent number of landmarks. **(b)** Fixing the landmark count n and plotting the cardinality C_n against the bin count R and the number of landmark classes c shows the cardinality manifold as a family of scaled functions. The plane at $c = 6$ indicates the maximum number of landmark classes (i.e. different forms of phase singularities) a Turing pattern will exhibit (as detailed in Section 2.5).

Based on this model the number of different pattern configurations containing up to N landmarks is then given as the accumulation of all the contributing cardinalities C_n:

$$\sum_{n=1}^{n=N} C_n = \sum_{n=1}^{n=N} c^n \binom{R}{n} \tag{5.16}$$

where the measure reduces to $((c+1)^R - 1)$ if $N = R$. A conservative estimate of the domain parameters for the chest patterns of African penguins, say $R \approx 80$, $N \approx 30$ and $c = 1$, would yield a pattern domain containing about 2×10^{22} different configurations.

Although this measure of uniqueness is significantly smaller than the ones estimated for fingerprints (which range up to cardinalities of 10^{80} [191]), if exploited fully, it would allow for a disambiguation of the majority of forseeable generations of African penguins.

However, this theoretical configuration space cannot be harvested in practice since natural patterns are not drawn from this space as necessary, e.g. 7-spot patterns are more likely than 30-spot patterns although the latter can encode significantly more configurations.

5.5.2 Random Pattern Correspondence

Only a small fraction of the population is known to the system at any one time, so it will operate by authentication of pattern pairs. Focussing on authentication, the probability of *correctly matching* a given *reference profile* to a measured, *random configuration* will now be modelled in close proximity to the work by *Prabhakar* [159] on fingerprint uniqueness.

Assuming a single landmark type (i.e. $c = 1$), the probability of matching a randomly created pattern containing a *single* landmark (i.e. $n = 1$) to any *one* (i.e. $p = 1$) of m landmarks in a given reference profile is given by:

$$\mathcal{P}(A, B, p = 1, n = 1, m) = \frac{mB}{A} \tag{5.17}$$

The term reflects the chance of picking the right one bin out of m bins. The chance of matching exactly 1 out of 2 random landmarks in an m-landmark profile is composed out of two events, namely, matching the first landmark and not matching the second one and vice versa:

$$\mathcal{P}(A, B, p = 1, n = 2, m) = 2 \overbrace{\frac{mB}{A}}^{\text{match}} \overbrace{\frac{(A - mB)}{(A - B)}}^{\text{mismatch}} \tag{5.18}$$

The probability of matching exactly 1 out of n landmarks then requires factoring over the probabilities of matching one but none of the $(n - 1)$ remaining landmarks:

$$\mathcal{P}(A, B, p = 1, n, m) = n \, \frac{mB}{A} \underbrace{\frac{(A - mB)}{(A - B)} \frac{(A - (m - 1)B)}{(A - 2B)} \cdots \frac{(A - (m - n + 2)B)}{(A - (n - 1)B)}}_{(n-1) \text{ factors}} \tag{5.19}$$

$$= \binom{n}{p} \frac{mB}{A} \prod_{k=1}^{n-1} \frac{(A - (m - k + 1)B)}{(A - kB)} \tag{5.20}$$

Finally, for matching exactly p out of n landmarks in a random signal against m landmarks in the reference pattern the probability reads:

$$\mathcal{P}(A, B, p, n, m) = \binom{n}{p} \prod_{l=1}^{p} \frac{(m - l + 1)B}{(A - (l - 1)B)} \prod_{k=1}^{n-p} \frac{(A - (m - k + 1)B)}{(A - (k + p - 1)B)} \tag{5.21}$$

The term collapses to a group of factors which has been shown a hypergeometric distribution[30] [159, pp.70–71]:

$$\mathcal{P}(R, p, n, m) = \binom{n}{p} \binom{R - n}{m - p} \binom{R}{m}^{-1} \tag{5.22}$$

where $R = A/B$. Note that $c = 1$, so the distribution reflects a *spatial structure* measure.

[30] The hypergeometric distribution is a discrete probability distribution that reflects the number of successes p in a sequence of n draws from a finite population of $R = A/B$ without replacement where m elements of success exist in the population.

When projected into the m-n-space of the landmark counts in profile and measurement – as shown in Figure 5.31(a) – the function forms a saddle-like shape which is symmetrical

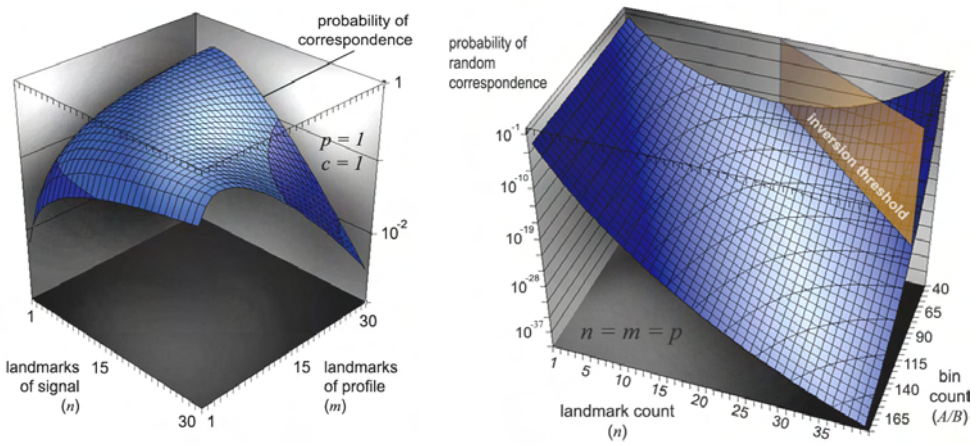


Figure 5.31: ***Random Correspondence.*** **(a)** The figure shows the probability density of random pattern correspondence for a fixed bin count of $R = 140$. **(b)** The probability surface shown (blue) visualises the chance of random correspondence (vertical axis) with respect to the bin count $R = A/B$ (depth axis) and the landmark count n (horizontal axis) for events of complete pattern match given $c = 1$. It can be seen that for a fixed bin count R there exists a landmark count n of minimal chance of random correspondence indicated by an inversion threshold hyperplane (shown in brown).

with regard to m and n, that is the two parameters are interchangeable. This confirms the intuitive observation that the process of matching two individuals is symmetrical in nature.

Finally, the probability of matching *at least* p landmarks between a pattern pair is:

$$\mathcal{P}(R,p,n,m) = \sum_{q=p}^{\min(m,n)} \binom{n}{q}\binom{R-n}{m-q}\binom{R}{m}^{-1} \tag{5.23}$$

This measure can be used coarsely to estimate the chance of random pattern correspondence for particular matching situations.

For instance in accordance with this model, $\mathcal{P}(80,3,4,4) \approx 2 \times 10^{-4}$ approximates the odds of dealing with a random correspondence after having matched 3 landmarks between two 4-landmark patterns on penguin chests.

5.5.3 Final Note on Landmark Density: Less Can Be More

The special case of matching all landmarks correctly, that is setting $n = m = p$, yields a term that directly relates the feature count n and the bin count R to random correspondence:

$$\mathcal{P}(R, n) = \binom{R}{n}^{-1} \tag{5.24}$$

When plotting the function $\mathcal{P}$ against the bin count R and the landmark count n (see Figure 5.31(b)), a property is revealed that opposes the idea of a positive correlation between landmark count and identification potential: a greater number of landmarks *does not* necessarily imply a better recognisability within a population.

Instead, there exists an **optimal landmark density** at the inversion threshold, which exhibits the best potential for successful pattern disambiguation with a minimal chance of a random correspondence[31]. For $c = 1$, the inversion threshold n_R sits at half the bin count[32]:

$$n_R = \min_n \binom{R}{n}^{-1} = \left\lceil \frac{R}{2} \right\rceil \tag{5.25}$$

Thus, comparing bin counts, a pattern with equal amounts of bins with and without a landmark (a salt'n'pepper pattern on the coarse scale of partitions B) can be disambiguated best from other patterns given a single landmark type. Biologists who seek to confidently disambiguate animals may especially look out for patterns of this particular kind...

5.6 Chapter Summary

During the course of this chapter an approach to exploiting the *visual uniqueness* of Turing-like coat textures has been discussed for the specific purpose of individual animal identification.

First, it has been suggested to use properties of Turing patterns 1) to identify sparse sets of landmarks in animal coats, and 2) to employ their configurations and landmark types compactly to encapsulate individually unique information contained in coat textures.

[31] Note that, naturally, the probability of finding a perfect random match, as modelled in Eq. (5.24), is identical to the probability of randomly creating the pattern from scratch as given in Eq. (5.15).

[32] The relation $n_R = \lceil R/2 \rceil$ can be intuitively understood as representing the centre line of maxima in the Pascal Triangle where R is interpreted as the row count.

Second, it has been illustrated how spatial histograms in the form of shape contexts can be used to encode the extracted coat pattern landmarks based on the configuration of surrounding landmarks. The uncertainty of landmark locations within this model has been approximated by Gaussians whose parameters were chosen to reflect real distortion data.

Third, the earth movers distance has been proposed as a means to calculate landmark dissimilarities by comparing histogram pairs. A variable base cost measure between bins was introduced to reflect the derived Gaussian uncertainty. By solving the assignment problem of optimal landmark association between patterns, the best fit and associated matching costs could be retrieved. These costs were then interpreted as a measure of pattern dissimilarity. Weighted thresholding of the measure subsequently allowed for authenticating pattern pairs.

Authentication and identification benchmarks have then been presented based on test results from a prototype operating in an African penguin colony and on a photo collection of plains zebras.

The results indicate an applicability of the presented technique to aid population monitoring of the sample species. Finally, a stochastic model of pattern uniqueness has been discussed. It has been argued that, in the case that animal patterns would be truly random and could be extracted accurately, there would exist sufficient uniqueness to disambiguate very large populations.

Nature has implemented a reductionistic scheme for generating the epiphenomenon of 'coat texture' from basic chemical reactions and diffusion. It appears that the study of the properties of these basic constituents can be used to engineer identification systems that actually disambiguate individuals.

When Turing began his modelling of auto-generative patterns in the early 1950's, he started by considering the problem of morphogenesis, that is the transformation of a *fertilised egg* into a *complex organism.* It now emerges that there is a prospect for his results being applicable not only to morphogenesis, but also as a tool to better understand and conserve some of these *complex organisms.* ■

Chapter 6

CONCLUSION AND FUTURE WORK

'If nature were not beautiful, it would not be worth knowing [...]'

(quote attributed to Jules H. Poincarė[1], 1854 - 1912)

6.1 Thesis Summary

In this thesis it has been demonstrated that, given fair environmental acquisition conditions, *animals can be detected and – in the case that they carry Turing patters – individually identified based on visual material acquired in their natural habitat.* To provide a proof of concept for the feasibility of the claim of individual identification, a real-world prototype has been evaluated in a colony of African penguins and on a small scale zebra image collection where the presented results indicate a performance that is sufficient to aid the current population monitoring of the species.

In its broad structure, the proposed system has been designed along a *recognition pipeline*[2] which separates the coarse species detection, the posing/texture extraction and the deformation robust identification into different modules. The different parts of the thesis were organised to reflect this subdivision:

Species Detection. First, an algorithmic framework for species localisation has been discussed in Chapter 3. It has been shown that the appearance context of *key points* on coat patterns contains a species-specific component which, using *boosted point-surround classifiers*, can be extracted and utilised as a local species descriptor and detector.

[1] Jules Henri Poincarė became the first scientist to describe the chaotic behaviour of deterministic systems. His ideas laid the foundations of modern chaos theory, paving the way to the study of systems such as the Turing system used to explain the formation of naturally occurring patterns, e.g. animal coats.

[2] The subdivision into compact components renders the system flexible for execution in networked environments where, for instance, clients detect species members and servers perform the identification.

A number of detector properties have been demonstrated, including the robustness to changes in textural detail and illumination of the coats, the operability in environmental clutter, the close-to-real-time performance on high-resolution imagery as well as the particular efficiency of the employed Haar-like features in *encoding* Turing-like coat patterns. However, a number of detector limitations have been disclosed, including susceptibility to deep shadows, cryptic resemblance, occlusion and specular reflection. Both the detection performance as well as the pose coverage were then enhanced by employing detector arrays.

Texture Extraction. In Chapter 4 it has been illustrated how the component-like key point representation can be exploited to extract surface textures of biometric interest. A spatially flexible model for spatial representation and detection termed *feature prediction trees* has been used to achieve an association of key point instances to animal instances. The resulting correspondence sets have then been exploited for fitting geometrical surface models that allowed for the extraction of *pose-corrected texture maps*.

Coat-Pattern Biometrics. In Chapter 5 these maps have been shown to contain *individually-characteristic features* which – using an extension of *shape contexts* – can be represented as deformation-robust sets of histograms. The *earth movers distance* has been used for comparing these sets with a population database to retrieve animal identities. Finally, results from an animal colony have been shown to indicate a system performance that allows for disambiguating individuals in a colony of thousands.

6.2 Claims and Contributions

In this thesis it has been shown that *Turing-patterned animals can be individually identified based on visual material acquired in their natural habitat.* To the best of the author's knowledge, fully automatic identification of individual animals by coat pattern has not been attempted before. Apart from the major claim and the novelty of the particular application class, a summary of vision-related claims and contributions is listed below:

- **Statistical Detection:**
  - an application of the *Viola-Jones detector* [212] as a point-surround descriptor
  - a method for estimating the detector resolution for encoding Turing patterns
  - the generation of a real-valued detection output by modelling detection priors
  - a gradient-supported extension that achieves improved localisation

  - perspective constraints for a significant detection speedup in restricted scenarios

- **Spatial Models and Biometrics:**
  - a model termed *feature prediction tree* for object formation from key points that allows to encode the structure of spatially flexible landmark sets by multiple, affine relationships
  - a technique for the extraction of individually characteristic visual features from Turing-like coat textures using *gradient direction curls* and *bandpass filtering*
  - an extension to *shape contexts* [20] that explicitly models landmark uncertainty
  - a framework for comparing shape contexts that uses the *earth movers distance* [172] and the *Hungarian method* [115] for the construction of a distance measure
  - an application of *Prabhakar's uniqueness model* [159] to coat patterns
  - an evaluation of the *prototype system* in a colony of African penguins

6.3 Future Work

To consolidate, enhance and expand the present work, a selection of additional avenues of research and practical implementations that could be pursued are briefly discussed now.

First, a selection of ideas on how the techniques proposed in this thesis could be specifically extend and further integrated are listed below:

- **Detection:**
  - automating the selection of key points based on a database of 3D texture scans of animals (similar to the one built by *Huang et al.* [97] for human faces)
  - extension of the detector to a full tracker and feeding results of the biometric identification back to the detector in order to resolve ambiguities caused by occlusion-by-conspecifics

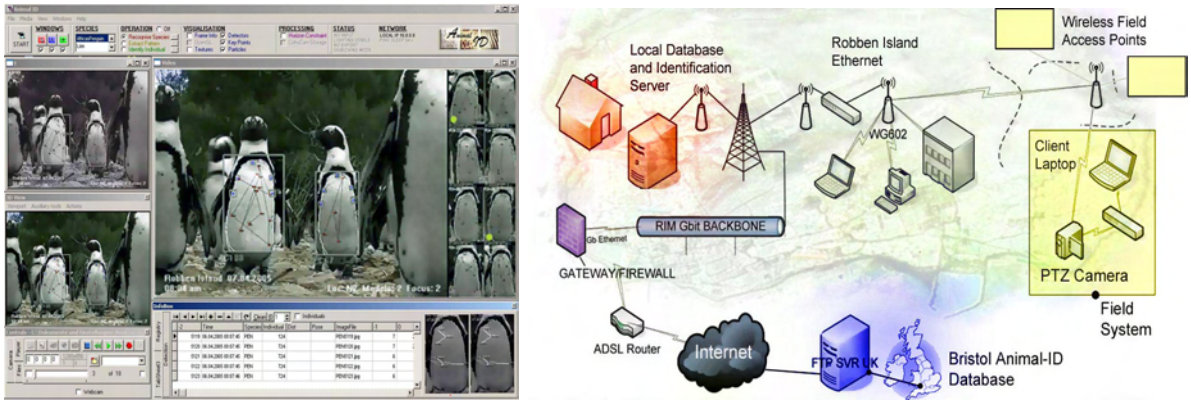
- **Biometric Matching:**
  - replacing the prototypical master profiles in the database by mean master patterns which are constructed as an average of multiple measurements (as recently suggested by *Ross et al.* [170] for fingerprints)
  - or replacing the one master profile used to represent an individual in the database by a non-parametric distribution built from multiple, confirmed measurements
  - building a truly physical species model to further compensate for occurring distortions

– find out about the true randomness intrinsic to coat patterns by analysing truly large numbers of spot patterns

In broader terms, the future success of vision as an enabling technology for an automatic monitoring of animal populations will crucially depend on the ability of systems:

- to cope with uncontrollable environmental factors such as high dynamics in lighting or the occurrence of specular reflectance, deeps shadows, fog, rain etc.
- to adequately compensate (via sensor multiplicity or smart acquisition) for animal behaviours that conceal the biometric entity of interest, that is activities such as grouping (resulting in occlusions) or rapid motion (resulting in motion blur)
- to improve the detection of non-linear body deformations and intra-population variances of coat textures by more efficient and effective models that demand less labeling/manual efforts.

Finally, the work discussed in this thesis has sparked an interdisciplinary research project between computer scientists, biologists and physicists. The project aims to develop and commission the world's first automated biometric vision system that autonomously monitors an entire animal colony (see Figure 6.1).



(a) **(b)**

Figure 6.1: ***Planned Monitoring System for Penguins on Robben Island.*** **(a)** Screenshot of the software prototype `Animal`*`ID`* currently under development. It is intended to be a distributed, network-based application. Its aim is to provide user-friendly access to an IT infrastructure that implements the algorithms discussed in this thesis, and thus enables biologists to utilise the vision technology for the benefit of their research and the conservation of the investigated species. **(b)** Diagram of the planned hardware infrastructure on Robben Island, South Africa. Field camera systems (yellow) detect species members in suitable poses and send coarsely cutout image patches to local servers (red) which then extract the surface textures of interest. Therefrom, the servers build biometric profiles which are finally compared to a local (red) or global (blue) population database.

6.4 Concluding Remark

Visual biometric identification by Turing-like patterns is – in theory – applicable to a broad variety of species ranging from *ants* to *whales*. Following from the results and indications of this thesis, the approach seems to offer one avenue towards non-invasive tools that can be applied to help in reconstructing the dynamics of real animal populations.

I believe that vision engineers should play a role in facilitating the understanding and conservation of endangered animals, first of all, for ethical reasons. However, not only the animal patterns we see but vision itself is a sophisticated pattern that has emerged with life and that represents the result of millions of years of evolution. Thus, an extinction of a species may also destroy an opportunity to understand the concepts inherent to their visual systems whose capabilities are, in many aspects and often by far, superior to artificially engineered solutions. ■

BIBLIOGRAPHY

[1] S. Abney. Bootstrapping. In *Annual Meeting of the Association for Computational Linguistics*, pages 360–367, 2002.

[2] D. G. Ainley, R. E. LeResche, and W. J. L. Sladen. Breeding biology of the adelie penguin. *The Quarterly Review of Biology*, 59(4):509, 1984.

[3] H. Akaike. A new look at statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, 1974.

[4] A. A. Amini, T. E. Weymouth, and R. C. Jain. Using dynamic programming for solving variational problems in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(9):855–867, 1990.

[5] B. N. Araabi, N. Kehtarnavaz, M. Yeary, G. Hillman, and B. Wursig. Locating an affine/projective invariant identifier patch on an image. In *IEEE Symposium on Image Analysis and Interpretation*, pages 121–125, 2002.

[6] Z. Arzoumanian, J. Holmberg, and B. Norman. An astronomical pattern matching algorithm for computer aided identification of whalesharks rhincodon typus. *Applied Ecology*, 42:999–1011, 2005.

[7] A. Aubel and D. Thalmann. Realistic deformation of human body shapes. In *Computer Animation and Simulation*, pages 125–135, 2000.

[8] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler. A system for traffic sign detection, tracking, and recognition using color, shape, and motion information. In *IEEE Intelligent Vehicles Symposium*, 2005.
[weblink: http://lmb.informatik.uni-freiburg.de/people/bahlmann/data/ba_zh_ra_pe_ko_iv2005.pdf].

[9] D. H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1981.

[10] A. L. C. Barczak, M. J. Johnson, and C. H. Messom. Real-time computation of haar-like features at generic angles for detection algorithms. *Res. Lett. Inf. Math. Sci.*, 9:98–111, 2006.

[11] C. J. Barnard. Ethical regulation and animal science: why animal behaviour is special. *Animal Behaviour*, 74(1):5–13, 2007.

[12] R. Basri. Recognition by prototypes. *International Journal of Computer Vision*, 19(2):147–167, 1996.

[13] R. Basri and Y. Moses. When is it possible to identify 3d objects from single images using class constraints? *International Journal of Computer Vision*, 33(2):95–116, 1999.

[14] H. W. Bates. Contributions to an insect fauna of the amazon valley. *Transactions of the Linnean Society of London*, 23:495–515, 1862.

[15] P. Bateson. Ethics and behavioural biology. *Advances in the Study of Behavior*, 35:211–233, 2005.

[16] H. Bay, T. Tuytelaars, and L. van Gool. Surf: Speeded up robust features. In *European Conference on Computer Vision*, volume 1, pages 404–417, 2006.

[17] T. Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions*, pages 370–418, 53.
[weblink: http://www.stat.ucla.edu/history/essay.pdf].

[18] M. Begon. *Investigating Animal Abundance: Capture-Recapture for Biologists*, page 97ff. University Press, Baltimore, Maryland, 1979.

[19] J. S. Beis and D. G. Lowe. Indexing without invariants in 3d object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(10):1000–1015, 1999.

[20] S. Belongie, J. Malik, and J. Puzicha. Shape context: A new descriptor for shape matching and object recognition. In *Neural Information Processing Systems*, pages 831–837, 2000.

[21] A. C. Berg, T. L. Berg, and J. Malik. Shape matching and object recognition using low distortion correspondences. In *Computer Vision and Pattern Recognition*, volume 1, pages 26–33, 2005.

[22] C. A. Berg and J. Malik. Geometric blur for template matching. In *Computer Vision and Pattern Recognition*, pages 607–614, 2001.

[23] P. J. Besl and N. McKay. A method for registration of 3-d shapes. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 14, pages 239–256, 1992.

[24] G. Bidloo. *Anatomy Humani Corporis (The Anatomy of the Human Body)*. Amsterdam, 1685. [weblink: http://special.lib.gla.ac.uk/anatomy/bidloo.html].

[25] E. Block. *Fingerprinting - Magic Weapon Against Crime*, page 7ff. David McKay, New York, 1969. ISBN 0679501932.

[26] M. A. Bray and J. P. Wikswo. Considerations in phase plane analysis for nonstationary reentrant cardiac behavior. *Physical Review E*, 65(5):051902, 2002.

[27] K. P. Burnham and Anderson D. R. *Model Selection and Multimodal Inference: A Practical Information-Theoretic Approach*, page 488ff. Springer-Verlag, 2 edition, 2002.

[28] G. Carneiro and D. Lowe. Sparse flexible models of local features. In *European Conference on Computer Vision*, number 3, pages 29–43, 2006.

[29] V. Castets, E. Dulos, J. Boissonade, and P. De Kepper. Experimental evidence of a sustained turing-type nonequilibrium pattern. *Physical Review Letter*, 64(24):2953–2955, 1990.

[30] A. L. F. Castro and R. S. Rosa. Use of natural marks on population estimates of the nurse shark (ginglymostoma cirratum) at atol das rocas biological reserve, brazil. *Environmental Biology of Fishes*, 72:213–221, 2005.

[31] M. Cavani and M. Farkas. Bifurcations in a predator-prey model with memory and diffusion ii: Turing bifurcation. *Acta Mathematica Hungarica*, 63(4):375–393, 1994.

[32] Y. Chen, S. Dass, A. Ross, and A. Jain. Fingerprint deformation models using minutiae locations and orientations. In *IEEE Workshop on Applications of Computer Vision*, number 150–156, 2005.

[33] G. E. Christensen, R. D. Rabbitt, and M. I. Miller. Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, 5(10):1435–1447, 1996.

[34] J. Clarke and K. Kerry. Implanted transponders in penguins: implantation, reliability, and long-term effects. *Journal of Field Ornithology*, 69:149–159, 1998.

[35] J. Cooper and P. Morant. The design of stainless steel flipper-bands for penguins. *Ostrich*, 52:119–123, 1981.

[36] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *European Conference of Computer Vision*, volume 2, pages 484–498, 1998.

[37] Intel Corporation. *Open Computer Vision Library (OpenCV) Reference*, 2007. [weblink: http://www.cs.unc.edu/Research/stc/FAQs/OpenCV/OpenCVReferenceManual.pdf].

[38] S. Cotin, H. Delingette, and N. Ayache. Real-time elastic deformations of soft tissues for surgery simulation. *IEEE Transactions on Visualization and Computer Graphics*, 5:62–73, 1999.

[39] G. Cox, G. De Jager, and B. Warner. A new method of rotation, scale and translation invariant point pattern matching applied to the target acquisition and guiding of an automatic telescope. In *South African Workshop on Pattern Recognition*, pages 167–172, 1991.

[40] B. M. Culik and R. P. Wilson. Swimming energetics and performance of instrumented adelie penguins, pygoscelis adeliae. *Journal of Experimental Biology*, 158:355–368, 1991.

[41] B. M. Culik, R. P. Wilson, and R. Bannasch. Flipperbands on penguins: what is the cost of a life-long commitment? *Marine Ecology Progress Series*, (98):209–214, 1993.

[42] H. Cummins. Ancient finger prints in clay. *Journal of Criminal Law and Criminology*, 32(4):468–481, 1941.

[43] H. Cummins. *Finger Prints, Palms and Soles - An Introduction To Dermatoglyphics*, page page 11. Dover Publications Inc., 1976. ISBN 0486207781.

[44] D. Cunado, M. S. Nixon, and J. N. Carter. Automatic extraction and description of human gait models for recognition purposes. *Computer Vision and Image Understanding*, 90(1):1–41, 2003.

[45] I. C. Cuthill. Field experiments in animal behaviour: methods and ethics. *Animal Behaviour*, 42:1007–1014, 1991.

[46] I. C. Cuthill. Ethical regulation and animal science: why animal behaviour is not so special. *Animal Behaviour*, 74(1):15–22, 2007.

[47] I. C. Cuthill, M. Stevens, J. Sheppard, T. Maddocks, C. A. Parraga, and T. S. Troscianko. Disruptive coloration and background pattern matching. *Nature*, 434:72–74, 2005.

[48] J. Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30, 2004.

[49] G. Donato and S. Belongie. Approximation methods for thin plate spline mappings and principal warps. In *European Conference on Computer Vision*, volume 3, pages 21–31, 2002.

[50] J. S. Doody. A photographic mark-recapture method for patterned amphibians. *Herpetological Review*, 26(1):19–21, 1995.

[51] C. Dorai, N. Ratha, and R. M. Bolle. Detecting dynamic behavior in compressed fingerprint videos: distortion. In *Computer Vision and Pattern Recognition*, pages 320–326, 2000.

[52] B. Duc, S. Fischer, and J. Bigun. Face authentication with gabor information on deformable graphs. *IEEE Transactions on Image Processing*, 8(4):504–516, 1999.

[53] K. M. Dugger, G. Ballard, D. G. Ainley, and K. J Barton. Effects of flipper bands on foraging behaviour and survival of adelie penguins (pygoscelis adeliae). *The Auk*, 123(3):858–869, 2006.

[54] J. A. Endler. A predator's view of animal color patterns. *Evolutionary Biology*, 11:319–364, 1978.

[55] J. A. Endler. Frequency-dependent predation, crypsis and aposematic coloration. *Philosophical Transactions of the Royal Society of London*, 319:505–522, 1988.

[56] M. Everingham and A. Zisserman. Identifying individuals in video by combining generative and discriminative head models. In *International Conference on Computer Vision*, pages 1103–1110, October 2005.

[57] L. Fei Fei, R. Fergus, and P. Perona. One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):594–611, 2006.

[58] P. F. Felzenszwalb. Representation and detection of deformable shapes. In *Computer Vision and Pattern Recognition*, volume 1, pages 102–108, 2003.

[59] P. F. Felzenszwalb. Representation and detection of deformable shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2):208–220, 2005.

[60] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient matching of pictorial structures. In *Computer Vision and Pattern Recognition*, volume 2, pages 66–73, 2000.

[61] A. Fick. *Philosophical Magazine*, 10(30–39), 1855.

[62] M. A. Fischler and R. A. Elschlager. The representation and matching of pictorial structures. *IEEE Transactions on Computers*, 22:67–92, 1973.

[63] J. Forcada and A. Aguilar. Use of photographic identification in capture-recapture studies of mediterranean monk seals. *Marine Mammal Science*, 16(4):767–793, 2000.

[64] W. Förstner. A framework for low level feature extraction. In *European Conference on Computer Vision*, pages 383–394, 1994.

[65] D. A. Forsyth and J. Ponce. *Computer Vision - A Modern Approach*. Prentice Hall, 2003. ISBN 0-13-085198-1.

[66] G. Foster, H. Krijger, and S. Bangay. Zebra fingerprints: towards a computer-aided identification system for individual zebra. *African Journal of Ecology*, 45(2):225–227, 2007.

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

[68] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

[69] N. Friday, T. D. Smith, and P. T. Stevick. Measurement of photographic quality and individual distinctiveness for the photographic identification of humpback whales (megaptera novaeangliae). *Marine Mammal Science*, 16:355–374, 2000.

[70] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: A statistical view of boosting. *The Annals of Statistics*, 38(2):337–374, 2000.

[71] G. Froget, M. Gauthier-Clerc, Y. Lehaho, and Y. Handrich. Is penguin banding harmless? *Polar Biology*, 20:409–413, 1998.

[72] D. Gabor. Theory of communication. *Journal of the Institute of Electrical Engineering*, 93(3):429–457, 1946.

[73] F. Galton. *Fingerprints*. MacMillan and Co., 1892. (Unabridged republication by DaCapo Press in 1965: [library weblink: http://medcat.wustl.edu/catflat/BFI/B463936.html]).

[74] G. Gamberale and B. Tullberg. Evidence for a peak-shift in predator generalization among aposematic prey. In *Biological Sciences*, volume 263, pages 1329–1334, 1996.

[75] H. Gamboa and A. Fred. An identity authentication system based on human computer interaction behaviour. In *Workshop on Pattern Recognition in Information Systems*, pages 46–55, 2003.

[76] S. Garfinkel. *Database Nation*. O'Reilly, 2001. ISBN 0596001053.

[77] M. Gauthier-Clerc, J. P. Gendner, C. A. Ribic, W. R. Fraser, E. J. Woehler, S. Descamps, C. Gilly, C. Le Bohec, and Y. Le Maho. Long-term effects of flipper bands on penguins. In *Royal Society of London*, B, pages 423–426, 2004.

[78] A. Gierer and H. Meinhardt. A theory of biological pattern formation. *Kybernetik [Cybernetics]*, 12:pp. 30–39, 1972.

[79] D. E. Gill. The metapopulation ecology of the red-spotted newt, notophthalmus viridescens (rafinesque). *Ecological Monographs*, 48:145–166, 1978.

[80] S. Gowans and H. Whitehead. Photographic identification of northern bottlenose whales (hyperoodon ampullatus): Sources of heterogeneity from natural marks. *Marine Mammal Science*, 17:76–93, 2001.

[81] P. Gray and S. K. Scott. Sustained oscillations and other exotic patterns of behavior in isothermal reactions. *Journal of Physical Chemistry*, 89:22, 1985.

[82] C. J. Green, P. N. Trathan, and M. Preston. A new automated logging gateway to study the demographics of macaroni penguins (eudyptes chrysolophus) at bird island, south georgia: testing the reliability of the system using radio telemetry. *Polar Biology*, 29:1003–1010, 2006.

[83] P. M. Griffin and C. Alexopoulos. Point pattern matching using centroid bounding. *IEEE Transactions on Systems, Man and Cybernetics*, 19(5):1274–1276, 1989.

[84] E. J. Groth. A pattern-matching algorithm for two-dimensional coordinate lists. *Astronomical Journal*, 91:1244–1248, 1986.

[85] P. D. Grünwald, I. J. Myung, and M. A. Pitt. *Advances in Minimum Description Length - Theory and Applications*. MIT Press, 2005. ISBN: 978-0-262-07262-5, page 7ff.

[86] S. R. Gupta. Statistical survey of ridge characteristics. *International Criminal Police Review*, 218(130), 1968.

[87] A. Haar. *Zur Theorie der orthogonalen Funktionssyteme [On the Theory of Orthogonal Functional Systems]*. PhD thesis, University of Göttingen, Germany, 1909.

[88] T. Hagstroem. Identification of newt specimens (urodela, triturus) by recording the belly pattern and a description of photographic equipment for such registrations. *British Journal of Herpetology*, 4:321–326, 1973.

[89] W. D. Hamilton. The genetical theory of social behavior. *Journal of Theoretical Biology*, 7:1–52, 1964.

[90] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–151, 1988.

[91] P. S. Heckbert. Filtering by repeated integration. *Computer Graphics*, 20(4):315–321, 1986.

[92] M. Heidegger. *Vom Wesen der Wahrheit [On the nature of truth]*. paragraph 35. Klostermann, 2nd edition, 1997. ISBN 978-3465029458, reprint of the 1930 original.

[93] L Hiby and P. Lovell. Computer-aided matchings of natural markings: a prototype for grey seals. Technical Report 57–61, Report of the International Whaling Commission, 1990.

[94] P. R. Hill, C.N. Canagarajah, and D. R. Bull. Statistical wavelet subband modelling for texture classification. In *IEEE International Conference on Image Processing*, volume 1, pages 165–168, 2001.

[95] F. S. Hillier and G. J. Lieberman. *Introduction to Mathematical Programming*. McGraw-Hill, New York, 1990.

[96] C. Huang, H. Ai, Y. Li, and S. Lao. Vector boosting for rotation invariant multi-view face detection. In *International Conference of Computer Vision*, volume 1, pages 446–453, 2005.

[97] J. Huang, B. Heisele, and V. Blanz. Component based face recognition with 3d morphable models. In *Audio-Video-Based Biometric Person Authentication*, pages 27–34, 2003.

[98] A. Hüseyin. Digitized and digital signatures for identification. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2003.
[weblink: http://anadolu.sdsu.edu/abut/dllecture-may_2003c.pdf].

[99] M. Isard and A. Blake. Condensation - conditional density propagation for visual tracking. *International Journal of Computer Vision*, 28(1):5–28, 1998.

[100] A. K. Jain. Biometric recognition: how do i know who you are? In *Signal Processing and Communications Applications Conference*, pages 3–5, 2004. ISBN 0-7803-8318-4.

[101] A. K. Jain, R. P. W. Duin, and Jiangchang Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, 2000.

[102] A. K. Jain and L. Hong. On-line fingerprint verification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:302–313, 1997.

[103] A. K. Jain, S. Pankanti, S. Prabhakar, L. Hong, A. Ross, and J. L. Wayman. Biometrics: A grand challenge. In *International Conference of Pattern Recognition*, volume 2, pages 935–942, 2004.

[104] D. Jones and J. Malik. Computational framework to determining stereo correspondence from a set of linear spatial filters. *Image and Vision Computing*, 10(10):699–708, 1992.

[105] M. Kearns and L. G. Valiant. Learning boolean formulae or finite automata is as hard as factoring. Technical Report TR-14-88, Harvard University Aiken Computation Laboratory, 1988.

[106] M. J. Kelly. Computer-aided photograph matching in studies using individual identification: an example from serengeti cheetahs. *Journal of Mammalogy*, 82:440–449, 2001.

[107] D. G. Kendall. Shape manifolds: Procrustean metrics and complex projective spaces. *London Mathematical Society*, 16:81–121, 1984.

[108] D. G. Kendall. A survey of the statistical theory of shape. *Statistical Science*, 4(2):87–99, 1989.

[109] H. Knutsson and C. F. Westin. Normalized and differential convolution: Methods for interpolation and filtering of incomplete and uncertain data. In *Computer Vision and Pattern Recognition*, pages 515–523, 1993.

[110] J. Koenderink. The structure of images. *Biological Cybernetics*, 50:363–370, 1984.

[111] M. Kölsch and M. Turk. Analysis of rotational robustness of hand detection with a viola-jones detector. In *International Conference of Pattern Recognition*, volume 3, pages 107–110, 2004.

[112] S. Kondo and R. Asai. The viable turing wave on the skin of p. imperator. *Nature*, 376:765–768, 1995.

[113] Z. Korotkaya. Biometric person authentication: Odor. Technical report, Department of Information Technology, Laboratory of Applied Mathematics, Lappeenranta University, 2004. [weblink: http://www.it.lut.fi/kurssit/03-04/010970000/seminars/Korotkaya.pdf] .

[114] H. Korves, B. Ulery L. Nadel, and D. Masi. Multi-biometric fusion: From research to operations. *Sigma*, pages 39–48, 2005.

[115] H. W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistic Quarterly 2*, pages pp. 83–97, 1955.

[116] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[117] Y. Kuramoto and S. I. Shima. Rotating spirals without phase singularity in reaction-diffusion systems. *Progress of Theoretical Physics*, (150):115–125, 2003.

[118] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. Malsburg, R. P. Wurtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computing*, 42:300–311, 1993.

[119] D. W. Lake. Ubiquitous biometrics: Fulfilling the promise at last - getting the picture. *Advanced Imaging*, 18(1):22–24, 2003.

[120] H. K. Lammi. Ear biometrics. Technical report, Lappeenranta University of Technology, Department of Information Technology, Laboratory of Information Processing, 2003. [weblink: http://www.it.lut.fi/kurssit/03-04/010970000/seminars/Lammi.pdf].

[121] J. D. Lebreton, K. P. Burnham, J. Clobert, and D. R. Anderson. Modeling survival and testing biological hypothesis using marked animals: A unified approach with case studies. *Ecological Monographs*, 62:67–118, 1992.

[122] H. C. Lee and R. E. Gaensslen, editors. *Advances in Fingerprint Technology*. CRC-Press, 1st edition, 1992. ISBN 0849395135.

[123] V. Lepetit, P. Lagger, and P. Fua. Randomized trees for real-time keypoint recognition. In *Computer Vision and Pattern Recognition*, volume 2, pages 775–781, 2005.

[124] D. Lewis. Extrinsic properties. *Philosophical Studies*, 44:196–200, 1983.

[125] S. Li, L. Zhu, Z. Zhang, A. Blake, H. J. Zhang, and H. Shum. Statistical learning of multi-view face detection. In *European Conference on Computer Vision*, volume 4, pages 67–81, 2002.

[126] R. Lienhart, A. Kuranov, and V. Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In *DAGM Pattern Recognition Symposium*, pages 297–304, 2003.

[127] L. Ljung. *System Identification - Theory for the User*. Prentice Hall, 2nd edition, 1999. ISBN 0-13-656695-2.

[128] P. Loafman. Identifying individual spotted salamanders by spot patterns. *Herpetological Review*, 22(3):91–92, 1991.

[129] E. N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, 1963.

[130] D. G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, pages 1150–1157, 1999.

[131] X. Lu and A. K. Jain. Deformation analysis for 3d face matching. In *Workshop on Applications of Computer Vision*, volume 1, pages 99–104, 2005.

[132] B. Luo and E. R. Hancock. Matching point-sets using procrustes alignment and the em algorithm. In *British Machine Vision Conference*, pages 43–52, 1999.

[133] M. Malpighius. *De externo tactus organo (On the external organs concerned with the sense of feeling)*. London, 1686. Online republication of a scanned original.
[web: www.illustratedgarden.org/mobot/rarebooks/page.asp?relation=QK41M3471687&identifier=0645].

[134] D. Marr. *Vision: A computational investigation into the human representation and processing of visual information*. W. H. Freeman, 1983. ISBN 9780716715672.

[135] H. Meinhardt. *The Algorithmic Beauty of Sea Shells*. Springer, Heidelberg, New York, 3rd edition edition, 2003. 9783540639190.

[136] H. Meinhardt. Different strategies for midline formation in bilaterians. *Nature Reviews Neuroscience*, 5:502–510, 2004.

[137] D. A. Melton. Pattern formation during animal development. *Science*, 252(5003):234–241, 1991.

[138] P. Menezes, J. C. Barreto, and J. Dias. Face tracking based on haar-like features and eigenfaces. In *Syposium on Intelligent Autonomous Vehicles*, 2004.
[online republication: http://www.isr.uc.pt/~paulo/PUBS/iav04.pdf].

[139] S. Merilaita and J. Lind. Background-matching and disruptive coloration, and the evolution of cryptic coloration. In *Royal Society*, volume 272 of *B*, pages 665–670, 2005.

[140] K. Messer et al. Face verification competition on the xm2vts database. In *International Conference on Audio and Video Based Biometric Person Authentication*, pages 964–974, 2003.

[141] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–66, 2004.

[142] A. Mohan, C. Papageorgiou, and T. Poggio. Example-based object detection in images by components. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(4):349–361, 2001.

[143] F. Mueller. Bemerkenswerthe fälle erworbener aehnlichkeit bei schmetterlingen [outstanding cases of a developed similarity in butterfly]. *Kosmos, Year 5 Issue*, 1881.
[weblink to an English review: http://www.wku.edu/~smithch/wallace/S353.htm].

[144] T.N. Mundhenk and L. Itti. Computational modeling and exploration of contour integration for visual saliency. *Biological Cybernetics*, 93(3):188–212, 2005.

[145] J. D. Murray. How the leopard gets its spots. *Scientific American*, page 20 ff., 1988.

[146] J. D. Murray. *Mathematical Biology 1 - An Introduction*. Springer-Verlag Berlin Heidelberg, 3rd edition, 2002. ISBN 0-387-95223-3.

[147] J. D. Murray. *Mathematical Biology 2 - Spatial Models and Biomedical Applications*. Springer-Verlag Berlin Heidelberg, 3rd edition, 2003. ISBN 0-387-95228-4.

[148] John von Neumann. *Collected Works by John von Neumann: Method in the Physical Sciences*, volume 6. Pergamon Books Ltd., 1962.

[149] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio. Pedestrian detection using wavelet templates. In *International Conference of Computer Vision and Pattern Recognition*, pages 193–199, 1997.

[150] M. Ozuysal, V. Lepetit, F. Fleuret, and P. Fua. Feature harvesting for tracking-by-detection. In *European Conference on Computer Vision*, pages 592–605, 2006.

[151] S. Pankanti, A. K. Jain, and S. Prabhakar. Learning fingerprint minutiae location and type. *Pattern Recognition*, 36(8):1847–1857, 2003.

[152] S. L. Petersen, G. M. Branch, D. G. Ainley, P. D. Boersma, J. Cooper, and E. J. Woehler. Is flipper banding of penguins a problem? *Marine Ornithology*, 33:75–79, 2005.

[153] J. C. B. Peterson. An identification system for zebra (equus burchelli, gray). *East African Wildlife Journal*, 10:59–63, 1972.

[154] A. Pfitzmann. Biometrie - wie einsetzen und wie keinesfalls? [biometrics - how to use and how not to?]. In *Security and Privacy in Information Society*, page 9, 2005.

[155] P. J. Phillips, P. Grother, R. J. Michaels, D. M. Blackburn, E. Tabassi, and M. Bone. Evaluation report. Technical report, Facial Recognition Vendor Test 2002, 2003.

[156] K. H. Pollack, J. D. Nicols, C. Browne, and J. E. Hines. Statistical inferences for capture-recapture experiments. *Wildlife Monographs*, 107:1–97, 1990.

[157] A. Pope and D. G. Lowe. Probabilistic models of appearance for 3-d object recognition. *International Journal of Computer Vision*, 40(2):149–167, 2000.

[158] K. Popper and J. C. Eccles. *Das Ich und sein Gehirn [The self and its brain]*. Piper, 1989. ISBN 9783492210966.

[159] S. Prabhakar. *Fingerprint Classification and Matching Using a Filterbank*. Doctoral theses, Computer Science Engineering Department, Michigan State University, 2001.

[160] I. Prigogine and R. Lefevre. Symmetry breaking instabilities in dissipative systems. *Journal of Chemical Physics*, 48:1665–1700, 1968.

[161] M. Pupilli and A. Calway. Real-time visual slam with resilience to erratic motion. In *Computer Vision and Pattern Recognition*, pages 1244–1249, 2006.

[162] J. E. Purkinje. *Commentatio de Examine Physiologico Organi Visus et Systematis Cutanei (Physiological Examination of the Visual Organ and of the Cutaneous System)*. Breslau: Vratisaviae Typis Universitatis, 1823. (Translated into English by H. Cummins and R. W. Kennedy, American Journal of Criminal Law, Criminology vol 31, pp. 343-356, 1940).

[163] A. Rangarajan, H. Chui, and F. L. Bookstein. The softassign procrustes matching algorithm. *Information Processing in Medical Imaging*, pages 29–42, 1997.

[164] E. Ranguelova, M. Huiskes, and E. J. Pauwels. Towards computer-assisted photo-identification of humpback whales. In *International Conference on Image Processing*, volume 3, pages 1727–1730, 2004.

[165] N. Ratha, K. Karu, S. Chen, and A. K. Jain. A real-time matching system for large fingerprint databases. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):799–813, 1996.

[166] S. Ravela and L. R. Gamble. On recognizing individual salamanders. In *Asian Conference on Computer Vision*, volume 2, pages 741–747, 2004.

[167] I. Rigoutsos. *Massively Parallel Bayesian Object Recognition.* PhD thesis, Computer Science Department, Courant Institute of Mathematical Sciences, New York University, 1992.

[168] J. Rodriguez. South atlantic crossings: Fingerprints, science, and the state in turn-of-the-century argentina. *The American Historical Review*, 109(2), 2004. (Review based on: Origen del Vucetichismo by Almandos, L. R., 1909).

[169] A. Ross, S. Dass, and A. Jain. Estimating fingerprint deformation. In *International Conference on Biometric Authentication*, pages 249–255, 2004.

[170] A. Ross, S. Dass, and A. K. Jain. Fingerprint warping using ridge curve correspondences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):19–30, 2006.

[171] F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce. 3d object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraints. *International Journal on Computer Vision*, 66(3):231–259, 2006.

[172] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision-*, 40(2):99–121, 2000.

[173] T. G. Ruggles. Biometric technical assessment. Technical report, Biometric Tech. Inc., 2002. [weblink: http://bio-tech-inc.com/Bio_Tech_Assessment.html].

[174] E. J. Russell. Extension of dantzigs algorithm to finding an initial near-optimal basis for the transportation problem. *Operational Research*, 17:187–191, 1969.

[175] G. Ruxton, T. Sherratt, and M. Speed. *Avoiding Attack: The Evolutionary Ecology of Crypsis, Warning Signals and Mimicry.* Oxford University Press, 2004. ISBN 0198528604.

[176] M. Sallaberry and J. Valencia. Wounds due to flipper-bands on penguins. *Journal of Field Ornithology*, 56:275–277, 1985.

[177] A. R. Sanderson, M. Kirby, C. R. Johnson, and L. Yang. Advanced reaction-diffusion models for texture synthesis. *Journal of Graphics Tools*, 11(3):47–71, 2006.

[178] R. E. Schapire. The strength of weak learnability. *Machine Learning*, 5(2):197–227, 1990.

[179] R. E. Schapire and Y. Singer. Improved boosting using confidence-rated predictions. *Machine Learning*, 37(3):297–336, 1999.

[180] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172, 2000.

[181] B. Schmidt. *Beeinflussung von Turingstrukturen in der Chlordioxid-Iod-Malonsure Reaktion mit elektrischen Feldern [Manipulation of Turing Patterns in Chlordioxid-Iod-Malonacid Reactions using Elektrical Fields].* PhD thesis, Otto-von-Guericke-Universität Magdeburg, Germany, 2005.

[182] J. Schnakenberg. Simple chemical reaction systems with limit cycle behavior. *Journal of Theoretical Biology*, 81:389–400, 1979.

[183] E. Schrödinger. *What is life?* Cambridge University Press, 1944. [weblink to lecture script: http://home.att.net/~p.caimi/Life.doc].

[184] D. K. Scott. Identification of individual bewicks swans by bill patterns. In B. Stonehouse, editor, *Animal marking: recognition marking of animals in research*. MacMillan, London, 160–168.

[185] A. Senior and R. Bolle. Improved fingerprint matching by distortion removal. *Transactions on Information and Systems*, 8(7):825–831, 2001.

[186] C. E. Shannon. Communication in the presence of noise. In *IEEE Proceedings*, volume 86, pages 447–457, 1998. [Reprint from Proceedings IRE, Vol. 37(1), 1949].

[187] J. Shi and C. Tomasi. Good features to track. *Conference on Computer Vision and Pattern Recognition*, pages 593–600, 1994.

[188] P. Y. Simard, L. Bottou, P. Haffner, and Y. L. Cun. Boxlets: A fast convolution algorithm for signal processing and neural networks. *Advances in Neural Information*, 11:571–577, 1999.

[189] F. Smeraldi and J. Bigun. Retinal vision applied to facial features detection and face authentication. *Pattern Recognition Letters*, 23:463–475, 2002.

[190] C. W. Speed, M. G. Meekan, and C. J. A. Bradshaw. Spot the match - wildlife photo-identification using information theory. *Frontiers in Zoology*, 4(2):1–11, 2007.

[191] D. A. Stoney and J. I. Thornton. A critical analysis of quantitative fingerprint individuality models. *Journal of Forensic Sciences*, 31(4):1187–1216, 1986.

[192] M. Stubbs and M. Edmunds. Defence in animals. *The Journal of Animal Ecology*, 45(2):607, 1974.

[193] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[194] G. H. Thayer. *Concealing-coloration in the animal kingdom*. Macmillan, New York, 1909.

[195] The Human Genome Project. Dna forensics. Technical report, 2006.
[weblink: http://www.ornl.gov/sci/techresources/Human_Genome/elsi/forensics.shtml].

[196] D. Thomas. Artificial enzyme membranes transport, memory and oscillatory phenomena. In *Analysis and Control of Immobilized Enzyme Systems*, pages 115–147, 1976.

[197] C. Tisse, L. Martin, L. Torres, and M. Robert. Person identification technique using human iris recognition. In *Vision Interface*, pages 294–299, 2002.

[198] B. Tordoff, W. W. Mayol, T. E. Campos, and D. W. Murray. Head pose estimation for wearable robot control. In *British Machine Vision Conference*, pages 807–816, 2002.
[weblink: http://www.bmva.ac.uk/bmvc/2002/papers/167/full_167.pdf].

[199] F. C. D. Tsai. *A Probabilistic Approach to Geometric Hashing using Line Features*. PhD thesis, Department of Computer Science, New York University, 1993.

[200] Z. Tu, X. Chen, A.L. Yuille, and S.C. Zhu. Image parsing: Unifying segmentation, detection, and recognition. In *International Conference on Computer Vision*, pages 18–25, 2003.

[201] A. M. Turing. The chemical basis of morphogenesis. In *Philosophical Transactions of the Royal Society of London*, volume 237 of *B*. The Royal Society, 1952.

[202] G. Turk. Generating textures on arbitrary surfaces using reaction-diffusion. In *ACM Siggraph*, volume 25, pages 289–298, 1991.

[203] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[204] J. Tyson. The belousov-zhabotinsky reaction. *Lecture Notes in Biomathematics*, 10:128ff., 1976.

[205] United States National Centre for State Courts. Biometric comparison chart. In *Electronic Courts Conference*, 2006.
[weblink: http://ctl.ncsc.dni.us/biometweb/BMCompare.html].

[206] United States NSTC Subcommitee on Biometrics. Biometrics history. Technical report, National Science and Technology Council, 2006.
[weblink: http://www.biometrics.gov/docs/biohistory.pdf].

[207] U.S. Department of Defense. Biometric standards under development as of 30 november 2005. [weblink: http://www.biometrics.dod.mil/documents/standards/Developing_Biometric_Standards_30Nov05.pdf].

[208] A. M. Van Tienhoven, J. E. Den Hartog, R. A. Reijns, and V. M. Peddemors. A computer-aided program for pattern-matching of natural marks on the spotted raggedtooth shark carcharias taurus. *Applied Ecology*, 44:273–280, 2007.

[209] A. Vezhnevets and V. Vezhnevets. Modest adaboost - teaching adaboost to generalize better. In *International Conference on Computer Graphics and Applications*, 2005.
[weblink: http://graphics.cs.msu.ru/en/publications/text/gc2005vv.pdf].

[210] C. Vincent, L. Meynier, and V. Ridoux. Photo-identification in grey seals: Legibility and stability of natural markings. *Mammalia*, 65(3):363–372, 2001.

[211] P. Viola and M. J. Jones. Fast and robust classification using asymmetric adaboost and a detector cascade. *Advances in Neural Information Processing Systems*, 2:1311–1318, 2002.

[212] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[213] P. Viola, M. J. Jones, and D. Snow. Detecting predestrians using patterns of motion and appearance. In *International Conference on Computer Vision*, volume 2, pages 734–741, 2003.

[214] P. Waage and C. M. Guldberg. Forhandlinger: Videnskabs-selskabet i christiana (studies concerning affinity). 35, 1864.
[weblink to English translation: http://chimie.scola.ac-paris.fr/sitedechimie/hist_chi/text_origin/guldberg_waage/Concerning-Affinity.htm].

[215] M. Walter, A. Fournier, and D. Menevaux. Integrating shape and pattern in mammalian models. In *ACM Siggraph*, pages 317–326, 2001.

[216] B. K. Williams, J. D. Nichols, and M. J. Conroy. *Analysis and Management of Animal Populations*. Academic Press, 1st edition, 2002. ISBN 0127544062.

[217] C. Wilson, M. Garris, and C. Watson. Matching performance for the us-visit ident system using flat fingerprints. Technical report, US National Institute of Standards and Technology Interagency Report 7110, 2004.

[218] A. Witkin and M. Kass. Reaction diffusion textures. *Computer Graphics*, 25:299–308, 1991.

[219] I. H. Witten and F. Eibe. *Data Mining - Practical Machine Learning Tools and Techniques*. Elsevier, Morgan Kaufmann Publishers, 2005. ISBN 0-12-088407-0.

[220] H. J. Wolfson and I. Rigoutsos. Geometric hashing: An overview. *IEEE Computational Science and Engineering*, 4(4):10–21, 1997.

[221] B. Wu, H. Ai, C. Huang, and S. Lao. Fast rotation invariant multi-view face detection based on real adaboost. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 79–84, 2004.

[222] S. Yablo. Intrinsicness. *Philosophical Topics*, 26:479–505, 1999.

[223] Z. Zhang, M. Li, S. Li, and H. Zhang. Multi-view face detection with floatboost. In *IEEE Workshop on Applications of Computer Vision*, page 184, 2002.

SOURCES AND COPYRIGHTS OF PHOTOGRAPHS

[I00] Photographic collections of the author. © Tilo Burghardt

[I01] African Penguin photo and video material, © Peter Barham, Richard Sherley and Tilo Burghardt

[I02] Plains Zebra photographs, courtesy of Sophie Grange

[I03] Cheetah in the Serengeti, © Christof Abt

[I04] Masai Giraffes (*Giraffa camelopardalis tippelskirchi*), GNU public license, photo: Sandra Fenley

[I05] Whaleshark, © Eli Muh

[I06] Butterfly, free public domain license, photo available on http://www.pdphoto.org

[I07] Eastern Milk Snake (*Lampropeltis triangulum*), © John White

[I08] Ant, photograph and electron microscope image, courtesy of the University of Bristol

[I09] African Penguin underwater, public domain photograph

[I10] Wildlife video clips (CIF) of lions and zebras, unpublished, courtesy of Granada Media, Bristol

[I11] Slides and material from [103] presented at ICPR 2004 by Anil K. Jain

[I12] African penguin (albino), public domain photograph by Adrian Pingstone

ABSTRACT IN GERMAN LANGUAGE (ZUSAMMENFASSUNG DER DISSERTATION)

Die Grenzen der Anwendbarkeit maschinellen Sehens (*engl. computer vision*) befinden sich seit den ersten Anfängen dieses interdisziplinären Arbeitsfeldes in einem Prozess stetiger Neudefinition und sind bis heute nur unklar umrissen. Diese Dissertation diskutiert ein neues, praktisch relevantes Anwendungsfeld für Computer Vision: *Visuelle Tierbiometrie.*

Die Arbeit demonstriert, dass in natürlichen Lebensräumen gefilmte Tiere durch maschinelles Sehen vollautomatisch erkannt und identifiziert werden können. Die Dissertation beschreibt und validiert Algorithmen, welche eine visuelle Registrierung einer Tierart und – im Falle die beobachtete Spezies trägt Turingmuster – eine individuelle Identifikation von Einzeltieren in Bild- und Videomaterial ermöglichen. Das vorgeschlagene Detektionssystem ist robust gegenüber verschiedenen Ansichtskontexten (*engl. view contexts*), wechselnden Beleuchtungsverhältnissen und natürlichen, durch Störmuster geprägte Umgebungen. Die verschiedenen Leistungen und Anwendbarkeitsgrenzen des Systems werden an Löwen, Steppenzebras und Brillenpinguinen ausführlich beschrieben.

Zuerst wird ein algorithmisches System zur Artenerkennung diskutiert. Es wird gezeigt, dass der strukturelle Bildkontext in der Umgebung von Referenzpunkten auf tierischen Tarnmustern oft artspezifisch ist und daher für eine Erkennung von Tieren einer Art in Videomaterial geeignet ist. Das vorgeschlagene Model benutzt geboostete Punkt-Umgebungs-Klassifikatoren als lokale Musterdetektoren.

Im zweiten Teil der Arbeit wird illustriert, wie perspektivisch normalisierte Texturbilder Turing-gemusterter Tieroberflächen auf Basis der verfolgten Referenzpunkte extrahiert werden können. Es wird gezeigt, dass diese Bilder Merkmale enthalten, welche populationsweit einzigartig sind. Unter Benutzung von erweiterten Formkontexten (*engl. shape contexts*) können diese Merkmale in deformationsrobuste, biometrische Profile transformiert werden.

Die Anwendung von Distanzmaßen im multidimensionalen Raum dieser Profile ermöglicht letztlich eine automatische Zuordnung von Tieridentitäten zu Bildinhalten.

Zur Abrundung des Themas (sowie im Hinblick auf potentielle Anwendungen) beschreibt die Arbeit Tests an einem feldtauglichen Prototypen, welche die praktische Realisierbarkeit eines autonom operierenden Erkennungssystems demonstrieren. Experimente in einer afrikanischen Kolonie von Brillenpinguinen ergaben, dass mittels der vorgeschlagenen Methodik potentiell tausende Einzeltiere robust identifiziert werden können. Dieses Ergebnis markiert einen ersten Erfolg auf dem Weg hin zu automatisierten, nicht-intrusiven, rein visuellen Populationsüberwachungen, welche weitreichende Anwendungen in der Biologie, der Verhaltensforschung und in praktischen Studien zur Populationsentwicklung und zum Schutz gefährdeter Tierarten ermöglichen würden.

Appendix A

SYMBOLS AND FORMAL NOTATION

$[a_1, a_2, ..., a_n]$	a vector	$\mathbb{N}$	natural numbers	$=$	equality
$\{a_1, a_2, ..., a_n\}$	a set	$\mathbb{I}$	integers	$\neq$	disparity
$(x, y), [x, y]$	open and closed interval	$\mathbb{R}$	real numbers	$\approx$	approximately
$\{\ x \mid C\ \}$	set of all x that satisfy C	$\mathbb{C}$	complex numbers	$\cong$	congruence
$f : S_1 \to\ S_2$	mapping from S_1 to S_2	$\mathbb{N}_0$	$\mathbb{N} \bigcup \{0\}$	$<, >$	strict order signs
$S_1 \subset S_2$	S_1 is true subset of S_2	$S_1 \bigcup S_2$	union of sets	$\leq, \geq$	order symbols
$S_1 \subseteq S_2$	S_1 is subset of S_2	$S_1 \bigcap S_2$	common subset	$\equiv$	log. equality
$\exists\ x : C$	x exists and satisfies C	$S_1 \backslash S_2$	set difference	$\wedge$	conjunction
$\nexists\ x : C$	no x satisfies C	$S_1 \times S_2$	set product	$\vee$	disjunction
$\forall x : C$	all x satisfy C	$s \in S$	s is in S	$\neg$	negation
$\sup, \inf$	supremum, infimum	$s \notin S$	s is not in S	$\varnothing$	empty set
$\mathbf{M}^{-1}$	matrix inversion	S/R	factor set	∞	infinity
$\mathbf{M}^{T}$	matrix transposition	$[x]_R$	partition with x	f^{-1}	inverse function
$\mathbf{I}$	identity matrix	I	image function	$\mathbf{Re}(x)$	real-valued part of x
$\mathbf{tr}(\mathbf{M})$	matrix trace	$\|\vec{v}\|$	Euklid's distance	$\mathbf{pr}(\mathbf{x})$	product of vector entries
$\mathbf{M}_1 \times \mathbf{M}_2$	cross product	$\hat{R}$	hull of R	$\bigtriangledown$	nabla operator
$\mathbf{M}_1 \cdot\ \mathbf{M}_2$	dot product	$\mathfrak{P}(S)$	power set of S	$\parallel$	parallel
$\langle a_i \rangle \bullet \langle b_i \rangle$	element-wise operation	$\mathfrak{M}_{m \times n}$	$m \times n$ matrices	$\perp$	orthogonal
$\max(S)$	maximum element in S	$\|\boldsymbol{M}\|$	matrix determinant	$\measuredangle$	measured angle
$\min(S)$	minimum element in S	$\|S\|$	cardinality of set S	$\Rightarrow$	implication
arg	argument operator	(a_1, a_2)	open interval	$\Leftarrow$	re-implication
$x \bmod y$	modulo operator	$[a_1, a_2]$	closed interval	$\Leftrightarrow$	if and only if
$\ker f$	core of mapping f	$\circlearrowleft, \circlearrowright$	rotation direction	i	imaginary unit
$\sin, \cos, \tan$	trigonometric functions	∂	partial operator	$\lfloor x \rfloor$	floor operator
$\log_b x$	logarithm with basis b	$\int$	integral symbol	$\lceil x \rceil$	ceil operator
$\lg x$	natural logarithm	lim	limes	$x!$	factorial
$x^{[y]}$	y^{th} integration of x	x^y	exponentiation	$\sum$	sum operator
$\sqrt{x}\ , \sqrt[y]{x}$	square root, y^{th} root	e	Euler's number	$\prod$	product operator
$\otimes, \oplus, \oslash, \odot, \circ, \bullet$	abstract operators	π	Pi, Ludolf's number	$\triangle$	delta operator

Appendix B

EXTENDED MATERIALS

B.1 The Chemistry of Reaction Kinetics in Turing Systems

The pattern formation of Turing systems is driven by both diffusion and chemical reaction kinetics. Interestingly, an equation describing a chemical equilibrium reaction on a surface **I** can be *transformed* into reaction kinetics as used for the description of Turing systems.

According to Fick's *Law of Mass Action* [214], the transition can be generalised in the following transformation rule that reshapes a traditional chemical equilibrium equation into a system of partial differential kinetic equations:

$$\boxed{uA + vB \underset{k_1}{\overset{k_2}{\leftrightarrows}} wC} \iff \frac{\partial \mathbf{I}}{\partial t} = \begin{bmatrix} u \cdot (-k_1 a^u b^v + k_2 c^w) \\ v \cdot (-k_1 a^u b^v + k_2 c^w) \\ \dots \end{bmatrix} = \begin{bmatrix} f_a(a, b, \dots) \\ f_b(a, b, \dots) \\ \dots \end{bmatrix} \tag{B.1}$$

where A, B and C are placeholders for the chemical substances; a, b and c are their respective concentrations; u, v and w represent the reaction rates and k_1 and k_2 are the reaction coefficients which are constant for each reaction (expressing that rates are just proportional).

Using this scheme, chemical equilibria can be interpreted as a system of non-linear partial differential equations $f_i(\mathbf{I})$ that describe the reaction kinetics of the occurring reaction-diffusion. Together with the diffusion coefficients of the chemicals, the $f_i(\mathbf{I})$ fully determine the dynamics of a Turing system.

B.2 On the Structural Homogeneity of Polar Histograms

As discussed in the main body of the thesis, the *structural homogeneity* of a polar histogram is preserved if *all bins* have the *same proportions*.

It will now be shown that, for structurally homogenous polar histograms, the area of bins will grow linearly proportional to $(\log_n a)^2$ where n is the ring index and a is a constant. Given a sector of a concentric ring delimited by radii $(r,\, ar)$ cut out by an angle α (i.e. a single bin of a polar histogram as shown in Figure 5.14), the structural homogeneity is guaranteed if the bin shape, i.e. the ratio of bin width and height $w/h = \frac{ar-r}{\frac{2\pi ar}{\alpha}}$, can be maintained as equal for all bins.

This property is guaranteed if all neighbouring bins on adjacent rings with delimiting radii $(r,\, ar)$ and $(ar,\, br)$ show equal proportions, that is:

$$\frac{br-ar}{\frac{2\pi br}{\alpha}} = \frac{ar-r}{\frac{2\pi ar}{\alpha}} \quad \Leftrightarrow \quad \frac{br-ar}{b} = \frac{ar-r}{a} \quad \Leftrightarrow \quad \frac{b-a}{b} = \frac{a-1}{a} \tag{B.2}$$

Further simplification yields:

$$\frac{b-a}{b} = \frac{a-1}{a} \quad \Leftrightarrow \quad ab-a^2 = ab-b \quad \Leftrightarrow \quad a^2 = b \tag{B.3}$$

Thus, for homogeneity, the width of adjacent rings is linked so that the delimiting radii yield $(r,\, ar)$ and $(ar,\, aar)$=$a(r,\, ar)$. With respect to some ring index n, the radii grow exponentially by a constant exponent a. Consequently, the area of bins with respect to the ring index n must grow linearly in $\log^2$-space.

B.3 Information on the Study Site of Robben Island

The currently second largest African penguin colony – bearing more than 12,000 birds – is located on Robben Island[1], South Africa. The figure below illustrates the geography of the island at the southernmost tip of Africa.

(a) **(b)** **(c)**

Robben Island. **(a)** Robben Island (indicated by a red circle) is situated at the Western Cape of the African continent. **(b)** It lies 12 kilometers off the coast of Cape Town, South Africa in Table Bay. **(c)** An aerial photograph of the island: the study site is indicated by a red circle. [images: I02]

The penguin colony was selected for study since it provides close-to-optimal conditions for the observation of the species and the extensive data collection and testing of an artificial identification system:

1. The island is inhabited and accessible by ferry from Cape Town in about 40 minutes.
2. Mains power can be made available to supply technical devices at the major penguin highways, so there is no general need for expensive stand-alone power supplies.
3. The small size (appx. 6 km^2) of the island gives scope to monitor an entire colony in a future application scenario using only a small number of observation kits. It is estimated that 10 to 15 client devices will regularly monitor over 90% of the colony.
4. The University of Cape Town and the South African Marine and Coastal Management are involved in local research activities on the island. They agreed to collaborate on this project. ■

[1]With 507 ha it is the largest of the islands along the coastline of South Africa. The blue slate island is most famous for being the prison site that housed Nelson Mandela during apartheid.