Individual Animal Identification using Visual Biometrics on Deformable Coat-Patterns

Tilo Burghardt and Neill Campbell

Dept. of Computer Science
University of Bristol, UK

Abstract. In this paper we propose and evaluate a recognition approach to *individual* animal identification in patterned species based on video filmed in widely unconstrained, natural habitats. The key issue addressed is a distortion robust detection and comparison of unique, *deforming* camouflage markings as found in a wide range of species. We propose a coarse-to-fine methodology specifically extending and combining vision techniques in a three-stage approach, that is 1) a rapid, coarse key-view detection based on patch appearance, 2) pose estimation and 3D model fitting using a (pre-computed) dynamic *Feature Prediction Tree* (FPT) followed by bundle adjustment and 3) texture back-projection, extraction of unique phase singularities and final encoding using an extended variant of *Shape Contexts* [2]. Distortion-robust animal identification is then achieved by solving associated bipartite graph matching tasks for pairs of templates. Experimental results indicate a performance of GAR 92% at FAR 10^{-1}% for a prototype trained on non-occluded African penguins operating in a real-world animal colony of thousands. Independently producing time-stamped identification data, the system marks a first step towards a partial automation of biological field observations that may permit for a truly non-intrusive behavioural as well as conservational analysis of population dynamics.

1 Introduction

Visual identification exploiting biometric individuality in humans attracted significant research interest during the last decade. However, current biometric recognition systems employing unique entities such as the human iris [7, 19], fingerprints [18, 17] or facial features [14, 22], are still highly dependent on artificially controlled acquisition conditions and some form of normalised, cooperative user behaviour [11]. Hence, visual biometrics almost exclusively focus on the human subject despite a high potential for an application to wildlife.

As formalised by Murray [15], a large variety of animal species carry permanent camouflage markings on their coats. Such highly unique (but lifelong stable) patterns, mainly compositions of stripes and spots, originate from Reaction-Diffusion systems originally described by Turing [21]. Well-known examples include 'eye-spots' on butterflies and stripes on zebras. Due to quasi-chaotic formation processes, coat markings often differ significantly from individual to individual while following a wider theme typical for a species (see Fig. 1).

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

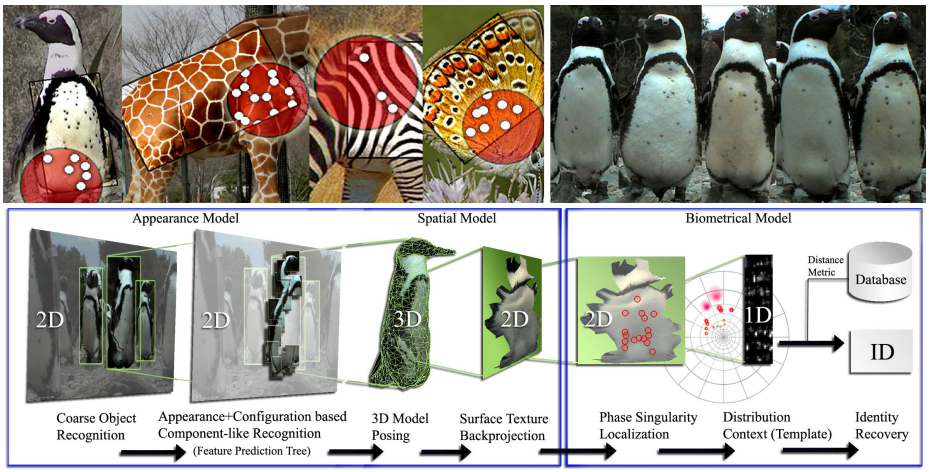

Fig. 1. *(Top Left)* Superimposition of phase singularities on coat patterns of several species. *(Top Right)* Individually unique spot patterns of African penguins. *(Bottom)* Conceptual outline of our proposed recognition pipeline and its components.

We propose, inspired by minutae representations of fingerprints, to compactly capture this individuality by spatial configurations of inherent phase singularities, e.g. landmark sets of spot centres or line bifurcations/endings. To utilise these sets as identifiers, a robust detection of animals and their pose is crucial since coat patterns (and any landmark set on them) undergo significant, non-linear deformation during motion altering the spatial configuration of features.

Various methodologies have been suggested to limit the impact of deformation: *distortion avoidance*, as suggested by Dorai et al. [8], aims at selecting a minimally deformed measurement. The strategy relies on both the availability of multiple measurements and means of rating the severeness of distortion[1].

Building on a different strategy, *Shape Contexts*, introduced by Belongie et al. [2], provide a rich, deformation robust representation of landmark sets, e.g. shapes, through spatial histograms that emphasise local coherence. However, the discrete nature of the histograms used requires larger landmark sets.

If a nominal pattern form can be recovered, distortion may be removed using image *warping techniques*[2] before matching. Here, the difficulty lies in an appropriate definition of a 'nominal representation'. Such techniques require dense, regular feature maps and are, in contrast to Shape Contexts, not easily applicable to sets of landmark data only (such as spot patterns of penguins). Moreover, all published systems dealing with fingerprint or iris consider 2D patterns only. In contrast, face recognition systems consider highly view-point dependent intra-

[1] Assuming the availability of dense time-series data, it has been proposed to utilise flow models [5] to determine the extent of deformation.

[2] For instance, Senior et al. [18] use equal spacing between ridge curves in fingerprints to normalise patterns. Ross et al. [17] imply a smooth nature of the distortion field and model deformation employing thin plate splines anchored at ridge curves.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

class variation as well as natural variance in face geometry and appearance; the detection of face location and pose becomes a task in its own right. Viola and Jones [23] suggested to utilise AdaBoost to train Haar-like classifiers for finding *image patches* containing faces. A recent extension [9] organises similar edge based classifiers in decision trees to pose a rigid model aligned with the face. Huang, Heisele and Blanz [10] propose a *component-based detection* system trained on 3D models rendered in variable lighting and created as mixtures of both geometry and texture generated from 3D face scans. In contrast, several attempts in the graphics community aim at describing deformation by *physically accurate models* that simulate the interaction of underlying structure [1, 6]. Such models are rarely applicable for detection since they rely on knowledge about the forces involved or on inverse kinematic calculations.

However, the key requirement for the use of surface-model approaches is the establishment of correspondences between model and image on the basis of local (or global) features. To achieve this, current systems mainly use *wide baseline matching* employing, for instance, SIFT [13] or patch classification via Randomised Trees [16]. As we found, these methods (based on local appearance alone) perform poorly detecting features on deforming species with individually varying patterning. We therefore turned towards modelling the flexible spatial relationship of features. Ozuysal et al. [16] model *rigid objects* by using a single underlying 3D model or, generalising the method, an ellipsoid model to relate features. Recently Carneiro and Lowe [4] proposed the use of *flexible models* operating on pre-matched SIFT [13] features for the discovery of correspondences.

For our system, as conceptually outlined in Fig. 1, we model the spatial relationship of features on the basis of representative 2D projections observed from 3D animation of a deformable model. The system operates in a coarse-to-fine manner combining and extending previous work: we initially train Haar-like classifiers [23] on the entire object at key views of interest centred at a key feature (see Fig. 2). This detector (coupled with a tracking technique as outlined in [3]) yields a coarse estimate of the initial feature location and object size. On this basis we progressively extract other correspondences where the search is guided by predictions that spatially constrain the search. The method resembles ideas in Carneiro and Lowe [4] but differs conceptually by 1) actually *guiding* the feature extraction and by 2) incorporating 3D-visibility information. We furthermore employ adaptive classifiers as local descriptors that capture deformation invariant features as inspired by [16]. This establishes a middle ground between component detection [10] and low-level, local feature extraction such as SIFT [13]. We utilise Haar-like features for modelling the appearance as described in Sec. 3.1. We model the spatial coherence of patches by a Feature Prediction Tree as outlined in Sec. 3.2. In both cases we utilise 3D animation to train the model.

2 Learning from Animated 3D Model

In order to create training information, a deformable 3D model M with K surface features $F = \{\mathbf{f_1}, ..., \mathbf{f_K}\}$ is animated through T animation steps, that is $M = \{\mathbf{m}_1, ..., \mathbf{m}_T\}$ where $\mathbf{m}_t$ represents the model pose at step t. We aim to learn both 1) the (changing) appearance of feature patches and 2) the structure of their

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

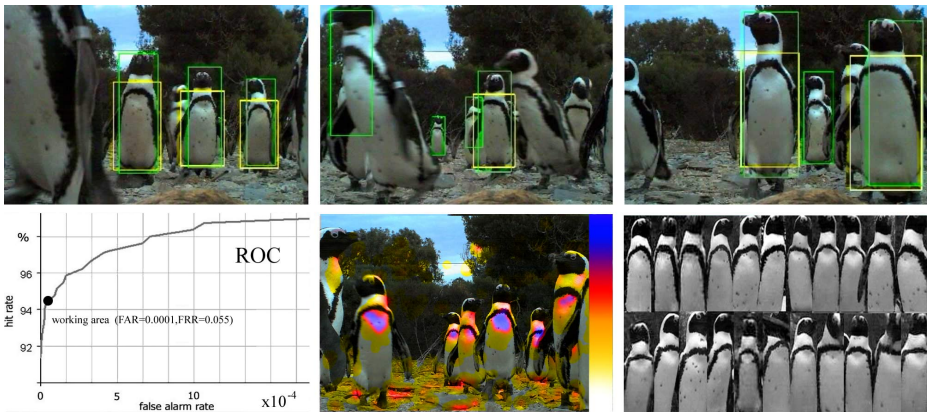

Fig. 2. Detection for Key-Features. *(Top)* frontal example detections; *(Bottom Left)* Receiver Operating Characteristic (ROC) of key-feature detector; *(Bottom Centre)* superimposition of matching probability (chest centre) using Eq. 2; *(Bottom Right)* positive samples from the training set.

spatial configuration. To limit the data set, we sample a sparse set of surface features that are chosen by detecting local interest points[3] on the appearance model under orthogonal viewing conditions.

The model is observed through H different virtual cameras placed at equally distributed angles around the model (see Fig. 3) to register different view aspects. During animation the cameras generate a set of $L = HT$ model projections. Hence, this yields a $K \times L$ matrix $\mathbf{M} = [\ m_1\ ...\ m_L\ ]$ of 2D projections m_l, each represented by a column vector $[x_{t,h,\mathbf{f}_1}\ ...\ x_{t,h,\mathbf{f}_K}]^T = [x_{1,\mathbf{f}_1}\ ...\ x_{L,\mathbf{f}_K}]^T$ for pairs $(t \in T, h \in H)$ containing object-centered 2D vertex coordinates. We will refer to each m_l as a *formation* describing a specific feature configuration. In addition, the projected image patch around each feature $\mathbf{f}_k$ (using multiple textures to capture species-variability) is stored creating an image set $D_{\mathbf{f}_k}$ per feature (see Fig. 3) for the training of local appearance classifiers as described in Sec. 3.1. To preserve visibility information for each m_l a set $\mathbf{v}_l = \{v_{l,\mathbf{f}_1}, ..., v_{l,\mathbf{f}_K}\}$ is constructed holding values $v_{l,\mathbf{f}_k} \in \{0, 1\}$ that indicate visibility of feature $\mathbf{f}_k$.

3 Probabilistic Model of Feature Correspondence

From this given training data we aim to establish a set of consistent feature correspondences, that is an injective mapping E from model features F to features $\hat{F}$ found in a novel image $\mathbf{I}$. The mapping can be expressed as a correspondence set $E = \{(\mathbf{f}_k, \hat{\mathbf{f}}_{c(k)}) | 1 \leq k \leq K\} = \{e_1, ..., e_K\}$ where $c(k)$ is a correspondence function that maps indices. Let $P(\omega_{\mathbf{f}_k}|\hat{\mathbf{f}}_j)$ represent the probability of $k = c(j)$, that is the estimated degree of association between a measured feature $\hat{\mathbf{f}}_j$ centred at

[3] The choice of the detector is application dependent. For the penguin application, we simply utilise a subset of $G = \{x \mid thresh > |\bigtriangledown [\mathbf{I}_{tex}(x) * \mathbf{G}(x)]|\ \}$ where $\mathbf{G}(x)$ is a Gaussian kernel.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

an interest point location $x_{\hat{\mathbf{f}}_j} \in \mathbf{I}$ and the class $\omega_{\mathbf{f}_k}$ representing model feature $\mathbf{f}_k$. Assuming statistical independence, we aim to approximate this probability by an evaluation of both the similarity of appearance and the positional validity yielding:

$$P(\omega_{\mathbf{f}_k}|\hat{\mathbf{f}}_j) = P_A(\omega_{\mathbf{f}_k}|\hat{\mathbf{f}}_j)P_{\bar{F}}(\omega_{\mathbf{f}_k}|\hat{\mathbf{f}}_j) \quad (1)$$

where P_A reflects a view-point independent 'similarity of feature appearance' and $\mathrm{P}_{\bar{F}}$ describes the probability of 'the feature being in a valid place' relative to some subset of already detected features $\bar{F} \subset \hat{F} \backslash \{\hat{\mathbf{f}}_j\}$. We now turn to an explicit modelling of these two essential measures.

3.1 Appearance Model using Haar-like Descriptors

Aiming for robust local appearance description, we (again) employ Haar-like feature classifiers [23]. We utilise AdaBoost to train weak Haar-like classifiers for each model feature $\mathbf{f}_k$ describing a feature class $\omega_{\mathbf{f}_k}$ based on positive data sets $D_{\mathbf{f}_k}$ and the negative set $D_{neg} \cup D_{\mathbf{f}_1} \cup ...D_{\mathbf{f}_{j\neq k}}... \cup D_{\mathbf{f}_K}$ respectively[4]. Following ideas described in [20], we modify the AdaBoost classifier to yield estimates for matching probabilities instead of binary classifications:

$$P_A(\omega_{\mathbf{f_k}}|\hat{\mathbf{f}}_j) = \max_{s\in S} \frac{e^{\bar{\mathrm{H}}_\mathrm{s}}}{e^{\bar{\mathrm{H}}_\mathrm{s}} + e^{-\bar{\mathrm{H}}_\mathrm{s}}} \qquad \bar{\mathrm{H}}_\mathrm{s} = \sum_{i=1}^{N}[\alpha_i \cdot h_i(s, x_{\hat{\mathbf{f}}_j})] \quad (2)$$

where $h_i(s, x_{\hat{\mathbf{f}}_j})$ represents the i^{th} weak Haar-like classifier with associated weight α_i calculated relative to image position $x_{\hat{\mathbf{f}}_j}$ using all scales $s \in S$.

3.2 Probabilistic Spatial Prediction Model

We model the degree of validity of feature positions by multivariate Gaussian distributions around rigid predictions. The parameters of interest are the multivariate variance $\Sigma_{j,\bar{F}}$ and the mean $\mu_{j,\bar{F}}$. We aim to estimate the latter by a known feature triple $\{\bar{\mathbf{f}}_m, \bar{\mathbf{f}}_n, \bar{\mathbf{f}}_o\} \subseteq \bar{F}$, that serves as reference system, and an affine parameter vector $\bar{\mathbf{a}}_j$ providing the relative location within this system. The triple indices and the associated vector form an affine predictor. During off-line training, for pairs $(\bar{F}, \mathbf{f}_k)$ the indices of the most stably performing feature triple as well as the associated affine displacement $\bar{\mathbf{a}}$ are learned on the bases of the related visible formation set, that is $\bar{\mathbf{M}} = \{m_l \in M | \forall_{\bar{\mathbf{f}}_k \in \bar{F}} : v_{l,\mathbf{f}_{c^{-1}(k)}} = 1\}$. To find this best affine predictor we solve the linear system:

$$\mathbf{a}_{l,j}(o,m,n) = (x_{l,\mathbf{f}_{c^{-1}(j)}} - x_{l,\mathbf{f}_{c^{-1}(o)}}) \cdot [\,(x_{l,\mathbf{f}_{c^{-1}(m)}} - x_{l,\mathbf{f}_{c^{-1}(o)}})\,(x_{l,\mathbf{f}_{c^{-1}(n)}} - x_{l,\mathbf{f}_{c^{-1}(o)}})]^{-1} \quad (3)$$

for $\mathbf{a}_{l,j}$ in all formations $m_l \in \bar{M}$ for chosen three-element subsets. We confine the tests to subsets of the N neighbouring features closest to $\{\hat{\mathbf{f}}_j\}$ since there would be up to $\binom{3}{|\bar{F}|} HT$ linear systems to solve for a complete affine stability search. The predictor $(\bar{\mathbf{a}}_j, c^{-1}(m), c^{-1}(n), c^{-1}(o))$ performing with smallest normalised variance, that is $\min(||\mu_{\bar{\mathbf{a}}_j}||^{-1} \cdot |\Sigma_j|)$, is chosen and the parameters for Gaussian approximation are estimated as:

$$\Sigma_j = \frac{1}{|\bar{\mathbf{M}}|} \cdot \sum_{l \mid m_l \in \bar{\mathbf{M}}} \left(\mathbf{a}_{l,j} - \mu_{\bar{\mathbf{a}}_j}\right)^2; \qquad \mu_{\bar{\mathbf{a}}_j} = \frac{1}{|\bar{\mathbf{M}}|} \cdot \sum_{l \mid m_l \in \bar{\mathbf{M}}} \mathbf{a}_{l,j} \quad (4)$$

[4] D_{neg} represents a set of randomly sampled patches of input images not containing instances of the feature.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

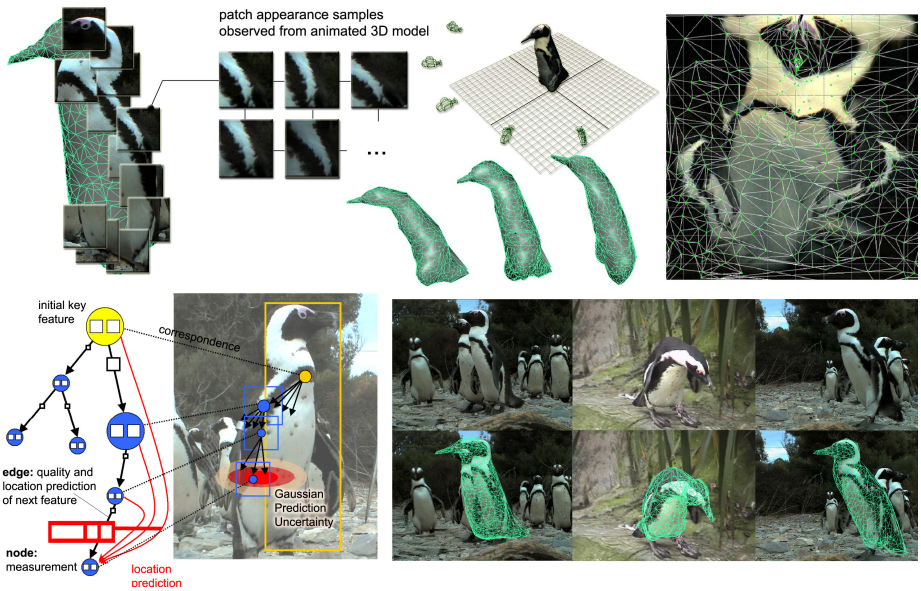

Fig. 3. *(Top Left)* Our appearance model recruits small, overlapping and localised patch-components as features. Also shown are 5 out of $L = 5,000$ sample patches (per feature) observed during animation used to train a Haar-like appearance classifier for the feature location. *(Top Centre)* Camera arrangement used (where $H = 5$, $K = 48$, $T = 1,000$) and samples of poses occurring during training animation. *(Top Right)* One of the texture maps used for animation. *(Bottom Left)* Visualisation of conceptual aspects of FPT's. *(Bottom Right)* Examples of correct detection results and a false detection (leftmost image) after performing bundle adjustment.

$$\mu_{j,\bar{F}} = x_{\bar{\mathbf{f}}_o} + \bar{\mathbf{a}}_j \cdot \left[\left(x_{\bar{\mathbf{f}}_m} - x_{\bar{\mathbf{f}}_o}\right) \left(x_{\bar{\mathbf{f}}_m} - x_{\bar{\mathbf{f}}_o}\right) \right] \tag{5}$$

Note that the linear systems are solved in the training stage while only the positional estimation is dynamically calculated based on actual locations of the features in the chosen set $\{\bar{\mathbf{f}}_m, \bar{\mathbf{f}}_n, \bar{\mathbf{f}}_o\}$. This division decouples the object structure (that is feature coherence expressed in the feature choice, average feature position and variance) from the measurements of feature locations. At the beginning of detection, when no three features are available as a reference system, we utilise the image reference system scaled by the initial detection size.

3.3 Tree Construction

The posterior matching probability of a certain formation m in the image $\mathbf{I}$ given a certain feature set $\hat{F}$ can be approximated by a product of probabilities of feature match $P(m|\hat{F}) \propto \prod_k P(\omega_{\mathbf{f_k}}|\hat{\mathbf{f}}_j)$. We aim to detect all possible such formations covered during training. The core idea of a *Feature Prediction Tree* is to interweave the detection process for different formations to dynamically order the evaluation sequence trying to calculate probabilities for highly predictable features first based on knowledge about 1) the learned object structure and

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

2) the previously identified features $\bar{F}$. The structural component is organised in a (large) tree where each node represents a stage of the recognition process and holds associations to features. Figure 3 provides a sketch of an FPT with annotations of its components.

The root, the initial detection stage, is represented by the initial key feature. Edges branching from a node n describes the minimal set of the most robust predictions of unmeasured features $F_N = \{\mathbf{f}_k\}$ that together increase the information about the presence of *all* hypothesised formations given the features on the path from n to the root. Robustness of a prediction depends on the co-visibility probability of the feature given $\bar{F}$, that is $P_{Vis}(\mathbf{f}_k) = |\bar{\mathbf{M}} \cap \{m_l \in M | v_{l,\mathbf{f}_k} = 1\}|/|\bar{\mathbf{M}}|$, and the success rates of the classifiers for appearance and position. The performance of the Haar-like classifiers as given in Eq.(2) is approximated by the true positive rate $P_{App}(\mathbf{f}_k)$ of the associated binary Haar-like classifiers evaluated at the constant false positive rate in the working area. The performance of a spatial predictor is approximated by its variance $P_{Pos}(\mathbf{f}_k) = \alpha/|\Sigma_k|$ where α is a weighting factor. For each node the set F_N is determined in a greedy procedure progressively adding $\mathbf{f}_{next} = \arg\max_{\mathbf{f}_k \notin (\bar{F} \cup F_N)} [P_{Vis}(\mathbf{f}_k) \cdot P_{App}(\mathbf{f}_k) \cdot P_{Pos}(\mathbf{f}_k)]$ until $(\forall_{m_l \in \bar{\mathbf{M}}} \exists_{\mathbf{f}_k \in F_N} : v_{l,\mathbf{f}_k} = 1)$ or a leaf is reached when $\bar{\mathbf{M}}$ contains only formations with the same visible features given $\bar{F}$. For each $\mathbf{f}_k$ a branch and a node is established and $\mathbf{f}_k$ as well as all features $\mathbf{f}_j \notin (\bar{F} \cup F_N)$ that share the same visibility given $\bar{F}$, that is $(\forall_{m_l \in \bar{\mathbf{M}}} : v_{l,\mathbf{f}_k} = v_{l,\mathbf{f}_j})$, are associated with the node. For all associated features, in sequence of their robustness, the prediction parameters detailed in Section 3.2 are calculated and stored. The procedure is repeated for all nodes without branches until all nodes have branches or are leaves.

During detection, we apply a depth-first search with backtracking until the partially evaluated posterior probability reaches either an acceptance threshold or drops below a rejection threshold. Dynamic Programming is used to aid the search since subproblems overlap, e.g. different paths share common features. Successful matching naturally yields a correspondence set E between all visible model and image features as well as a set of possible formations. This estimate is then refined using bundle adjustment initialised on the dense polygonal model and its view parameters associated with the formations found in the leaf (see Figure 3, right). The approach is not robust to severe occlusion. However, this fits the task since biometric identification requires the entity (spots) to be visible.

4 Biometric Identification

We apply our system to find and identify individual African penguins *(Spheniscus demersus)* in video sequences. The identification is based on the chest pattern of typically 5 to 25 stable black spots. We utilise the technique described to locate penguins in the video sequences and derive the correspondences between model and image features to pose a polygonal model into the scene. Back-projection then generates a texture map of the visible skin parts that is widely normalised for global deformation. Segmentation using adaptive thresholding is employed to extract the unique configuration of spot locations on the penguin chest. However, the spot pattern still suffers from non-linear local distortion of the skin.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

4.1 Robust Matching using Distribution Contexts

To address this problem, we employ a deformation robust matching technique: Shape Contexts [2] provide the means to incorporate local coherence into matching schemes. The method partitions the measurement space with respect to a reference landmark by building polar histograms of feature positions with increasing bin size further away from the reference. Since the technique is of statistical nature it performs well for large landmark sets even under distortion. For sparse sets it suffers from 'jumping' features close to bin borders. We propose an extension, that we refer to as *Distribution Context*, replacing feature landmarks by feature distributions describing their positional uncertainty relative to the reference spot. This addresses the problem of bin jumps by creating dense feature distributions. Fig. 4 shows a visualisation of two polar divisions from one Distribution Context (note that as many histograms as landmarks are constructed). We derive the form of these distributions from statistical measurements on penguin patterns under motion. Based on 20 sample clips of 5s length we find that the relative variance of feature pairs is correlated with their distance (see Fig. 4). We approximate the relationship fitting an exponential function $\sigma_{x_i}(x) = e^{(\alpha \cdot ||x - x_i||)} - \beta$ where the parameters were found to be $\alpha = 1.07 \cdot 10^{-2}$ and $\beta = 0.97$. This model describes the data with a correlation coefficient of $R = 0.976$. Using this finding we model a single landmark distribution by a Gaussian $f_{\mu,x_i}(x) = N \cdot e^{-\frac{1}{2e^{(\alpha \cdot ||x - x_i||)} - \beta}(x-\mu)^T(x-\mu)}$ that is *radially* distorted relative to the reference where N is a numerically pre-calculated normalisation factor. The matching of two distribution contexts employs the Hungarian method [12] to solve the assignment problem (identical to Shape Contexts). The sum of the associated matching costs is used as distance metric in pattern space. The definition of a threshold value for maximal distance between two patterns of the same individual allows the comparison of two patterns.

4.2 A Real-World Application

A static camera with a resolution of 1280x856 pixels was used for image acquisition at a penguin colony on Robben Island, South Africa. In order to estimate system performance 1,000 chest patterns of 114 individuals were used. The patterns were manually identified to build a ground truth. Each pattern was then compared to all other patterns. Various threshold values d in pattern space were applied to generate the Receiver Operating Characteristic (ROC) curve for the identification process as presented in Fig. 4. For comparison, we evaluated two other techniques, that is Shape Contexts [2] and landmark matching using the mean squared error of the closest matching feature pairs where the two patterns were aligned using Procrustes algorithm. The evaluation shows that distribution contexts slightly outperform Shape Contexts over the entire spectrum of the characteristic. At the working area at a false positive rate of 10^{-1} % distribution contexts perform with about 92 % genuine acceptance rate, that is 2 % better than Shape Contexts. The technique allows a machinised construction of identification time-lines (for example see Fig. 4). This can be a first step towards a partial automation of biological field observations that may permit for a non-intrusive behavioural and conservational analysis of population dynamics.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

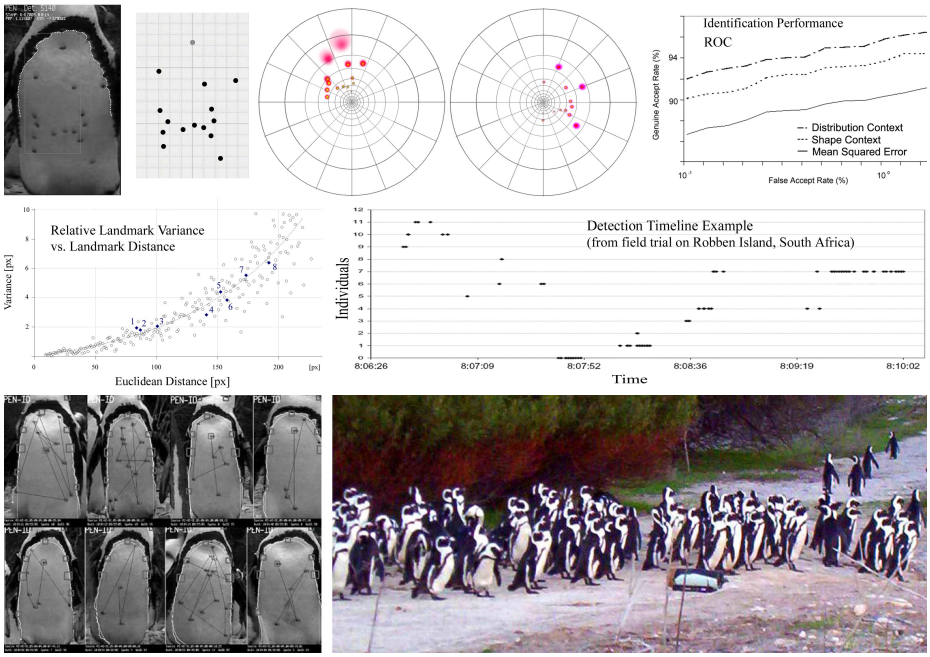

Fig. 4. *(Top Left)* Original detection image, normalised penguin spot pattern and two histograms (out of 15, one for each centre spot) from its Distribution Context. *(Top Right)* Receiver Operating Characteristic for individual penguin identification, comparison of techniques. *(Middle Left)* Variance of spot distance plotted against spot distance. *(Middle Right)* Time-line exsample (4min) of identification events with static camera in penguin colony. *(Bottom Left)* Various examples of identified African penguins with detected spot pattern superimposed. *(Bottom Right)* Application scenario on Robben Island: note the camera (middle) observing penguins returning to colony.

Acknowledgements: This work was made possible by funding of 'The Leverhulme Trust' and support of the 'Earthwatch Institute'. We would like to thank Peter J. Barham and Innes C. Cuthill for their project related guidance, inspirational discussions and advice. We would also like to thank Les G. Underhill, Rob Crawford and Mario Leshoro for their practical help and support during the realisation of field tests.

References

1. A. Aubel and D. Thalmann. Realistic deformation of human body shapes. In *Computer Animation and Simulation*, pages 125–135, 2000.
2. 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.
3. T. Burghardt and J. Calic. Analysing animal behaviour in wildlife videos using face detection and tracking. *IEE Proceedings Vision, Image and Signal Processing*, 153(3):305–312, 2006.

4. G. Carneiro and D. Lowe. Sparse flexible models of local features. In *European Conference on Computer Vision*, number 3, pages 29–43, 2006.
5. 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.
6. 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.
7. J. Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):21–30, 2004.
8. 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.
9. M. Everingham and A. Zisserman. Identifying individuals in video by combining generative and discriminative head models. In *Proc. International Conference on Computer Vision*, pages 1103–1110, October 2005.
10. 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.
11. A. K. Jain, S. Pankanti, S. Prabhakar, L. Hong, A. Ross, and J. L. Wayman. Biometrics: A grand challenge. volume 2, pages 935–942, 2004.
12. H. W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistic Quarterly 2*, pages pp. 83–97, 1955.
13. D. G. Lowe. Object recognition from local scale-invariant features. In *Proc. International Conference on Computer Vision*, pages 1150–1157, 1999.
14. 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.
15. J. D. Murray. *Mathematical Biology 2 - Spatial Models and Biomedical Applications*. Springer-Verlag Berlin Heidelberg, 3rd edition, 2003. ISBN 0-387-95228-4.
16. 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.
17. A. Ross, S.C. Dass, and A.K. Jain. Fingerprint warping using ridge curve correspondences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):19–30, 2006.
18. A. Senior and R. Bolle. Improved fingerprint matching by distortion removal. *IEIC Transactions on Information and Systems*, 8(7):825–831, 2001.
19. C. Tisse, L. Martin, L. Torres, and M. Robert. Person identification technique using human iris recognition. In *Vision Interface*, pages 294–299, 2002.
20. 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.
21. 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.
22. M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.
23. P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de