
Recognizing Animals Using Motion Parts
Changming Kong, Andrew Calway and Majid Mirmehdi

Presented at BMVC 2005

DISCLAIMER NOTICE

Readers should be aware that following the publication of this paper we discovered that there were errors in the experimental methodology. We are in the process of clarifying this with the first author. As such, although we stand by the ideas presented in the paper, we do not currently have sufficient confidence in the results as published.

Please accept our apologies for any inconvenience.

Andrew Calway and Majid Mirmehdi
University of Bristol

January 2006

Recognizing Animals Using Motion Parts

Changming Kong, Andrew Calway and Majid Mirmehdi

Department of Computer Science
University of Bristol, UK

{kong, andrew, majid}@cs.bris.ac.uk

Abstract

We describe a method for automatically recognizing animals in image sequences based on their distinctive locomotive movement patterns. The 2-D motion field associated with the animal is represented using a ‘configuration of motion parts’ model, the characteristics of which are learned from training data. We adopt an unsupervised approach to learning model parameters, based on minimal a priori knowledge of the physical or locomotive characteristics of the animals concerned. Results are presented demonstrating excellent classification performance, with accuracy exceeding 98% on a test set consisting of over 100 sequences of 7 different species.

1 Introduction

It has long been recognized that the visual motion patterns exhibited by a human or an animal in a given mode of locomotion can be powerful discriminatory features for recognition. A large body of work on human gait analysis has made convincing use of such patterns [7, 15]. Less work has been done on animal locomotion, see e.g. [9, 6], although the current interest in developing video archiving systems means that it would find considerable application for wildlife film making. This is the intended application for the work reported here. Motion information is important when dealing with wildlife sequences since alternative features, such as appearance or shape, are often difficult to extract due to natural camouflage, large variations in viewing conditions and the presence of significant background clutter in many of the sequences.

Many approaches have been investigated to capture the distinctive characteristics of human and animal movement. They fall broadly into two categories. First, those that use high-level models, such as skeletons or silhouettes, to represent physical attributes, with the variation amongst model parameters being used to characterize movement [2, 15, 7]. Secondly, those that use the spatio-temporal patterns of pixels or low level features, such as optical flow or point trajectories, with the patterns characterized using techniques such as spectral or principal component analysis [8, 1, 10]. The difference here is primarily one of specificity versus generality. Based on physical make up, the model based approaches are designed to capture the specific characteristics of the locomotion for a given subject. This can give high discrimination performance, but at the expense of generality and a reliance on good feature extraction, e.g. of limbs and joints. On the other hand, the spatio-temporal techniques generalize well, but rely on global motion properties, such as dominant periodicities, which are necessarily less discriminatory.

In this paper we describe a approach to representing and recognizing locomotive motion which seeks to address these limitations. It is based on a ‘configuration of motion parts’ model, in which the parts represent component motion patterns exhibited during locomotion. Importantly, the model characterizes the positional and cyclic correlations amongst these parts, as well as their individual motion properties. This structural component gives the potential for greater discrimination than previous spatio-temporal techniques, whilst being sufficiently flexible to allow generalization across species. The key point is that we hypothesize only the existence of such parts, not their precise configurations and inter-relationships as done in previous model based schemes. Instead, we learn the latter from training sequences within an unsupervised framework. The approach is motivated by the recent work by Song *et al* on representing human motion using parts [13] and by the work on parts based recognition of object categories in images, see e.g. [3, 14, 11, 4], all of which have demonstrated the utility of modeling with parts. Here, we extend the ideas to recognising animals through their characteristic locomotions.

The remainder of the paper is organised as follows. In the next section we outline the basis of modelling with motion parts and then in Section 3 describe a classification algorithm for recognizing species in a specific mode of locomotion using a configuration of parts framework. Results of experiments on a test set of over 100 sequences containing 7 different species are then presented in Section 4, demonstrating that the method gives excellent classification performance, achieving accuracy of over 98%.

2 Modeling with Motion Parts

Animals can be recognized from the patterns resulting from the component motions exhibited during a mode of locomotion, such as walking or running. These include cyclic movements of the legs, head and other body parts, as well as the more subtle rhythmic movements of muscles. The patterns are distinctive and uniquely characterize the animal. Because they result from the act of locomotion, they also have two important properties: they are often periodic in nature, as in the cyclic action of the legs, and they are often correlated - the bobbing up and down of a head being directly related to the cyclic leg movement. It is this periodicity and correlation amongst the components, in addition to their relative position, that we wish to harness in our recognition algorithm.

To do so, we model the 2-D motion field associated with an animal using ‘configurations of motion parts’. These correspond to constellations of regions which contain cyclic and consistent motion over an extended sequence of frames. As an example, Fig. 1a shows a configuration that might be used to represent a running Hyena. As well as the leg regions, the motion parts are centered on regions around the head and lower back, both of which exhibit cyclic motion patterns which characterize the gait of the Hyena. This can be seen from the vertical components of the average trajectory of points within the regions also shown in Fig. 1a. Note in particular the degree of positive and negative correlation between the trajectories.

Note that the motion parts need not necessarily correspond to individual physical parts of the animal, such as a leg or the head. Unlike approaches based on skeleton models, for example, we have deliberately adopted a weaker model based on distinctive regions within the motion field. This gives greater flexibility in terms of representing motion patterns (allowing the inclusion of muscle movements as well as body part motions, for

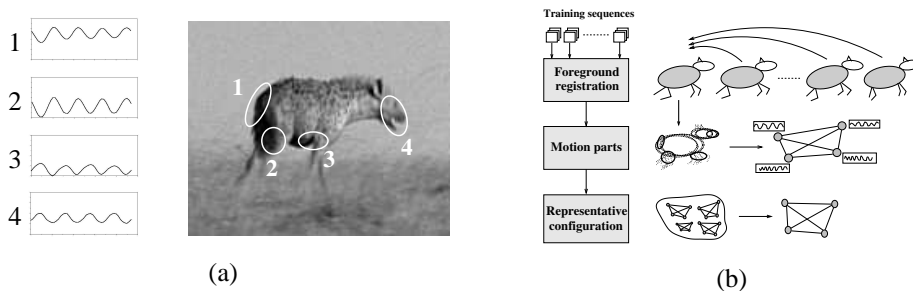


Figure 1: (a) Example motion part configuration representing the motion patterns of a running Hyena; (b) Main steps in deriving motion part configurations for recognition.

example) and hence greater generality across species. It also means that we are less demanding on the robustness of the feature extraction. Thus in Fig. 1a, parts represent the general motion in the vicinity of the legs, rather than the specific motion of individual legs or leg joints, which are difficult to distinguish accurately due to motion blur and the lack of contrast between the animal and the background.

Using a set of training sequences for a given species in a specific mode of locomotion, e.g. running, walking, etc, our approach is to identify a configuration of motion parts which characterises the data set. This is then used to recognize the same species in a similar mode of locomotion in unseen data. Importantly, this is done without pre-defining where the parts occur in the animal’s motion field, nor how they are inter-related; this is determined from the training data automatically. Each part has an associated frequency vector representing the fundamental periods of the horizontal and vertical cyclic motion within the underlying region. Pairs of parts are then related by horizontal and vertical correlation values. Note that ‘configuration’ refers here not only to the spatial arrangement of parts, but also to their frequency values and the associated correlations.

3 Recognition by Motion Parts

To determine a representative configuration, we first extract a (possibly large) number of potential motion parts and then seek a configuration based on a fixed number of parts which occurs consistently across the training data (this mirrors the strategy used in [14, 4] for parts-based object recognition). In practice, this amounts to determining the configuration which is closest to this ideal based on a distance metric between configurations. The same metric is then used to classify unseen data based on its minimum distance from each of the representative configurations. The approach is illustrated in Fig. 1b and the main components are summarized below.

We first register frames according to the dominant foreground motion, which we assume corresponds to the global motion of the animal. This normalizes the locomotive motion patterns with respect to a common reference. Potential motion parts are then identified by analysing the covariance map of registered pixels; motion part regions will exhibit motions distinct from the dominant motion and hence give rise to high pixel covariances. Regions surrounding local maxima in the covariance map are thus denoted as potential motion parts. The periodicities and correlations amongst the parts are then estimated by analysing the average trajectories of tracked feature points within the under-

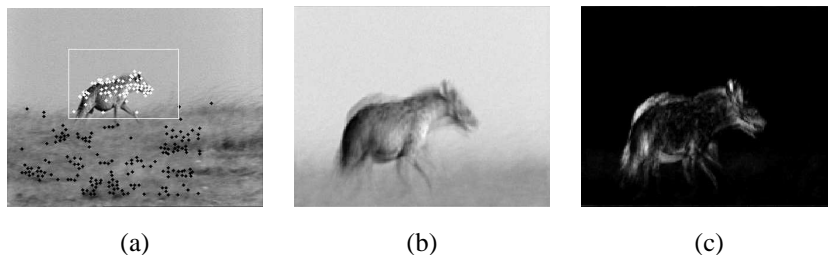


Figure 2: Foreground registration: (a) Segmented feature points and bounding box; (b) mean reference image ; (c) covariance image.

lying regions. Using the motion parts extracted from training sequences, we then derive a representative configuration based on a fixed, small number of parts (typically 4 or 5). This is achieved using an iterative comparison of potential configurations within the sets of motion parts based on a distance metric. Unseen sequences are then classified based on their distance from the representative configurations extracted for each category in the training data. Details of each of these components are given in the following sections.

3.1 Foreground Registration

Our training sequences contain a single animal in a specific mode of locomotion, roughly parallel to the image plane of a stationary or moving camera. For each sequence we register the frames with respect to a reference frame according to the global motion of the animal. Observing ‘frame differences’ following registration then enables us to identify significant motion parts. We use the transformed Gaussian method [5] to perform the registration. This models the frames as a series of transformations of a reference image within a probabilistic framework, with the reference image and the transformations treated as hidden variables. With the pixels in frame t stacked into a vector \mathbf{u}_t , the method is based on a generative model of the form

$$p(\mathbf{u}_t, T, \mathbf{z}) = P(T) \mathcal{N}(\mathbf{u}_t; T\mathbf{z}, \Psi) \mathcal{N}(\mathbf{z}; \bar{\mathbf{z}}, \Phi) \quad (1)$$

where $\mathcal{N}(\mathbf{z}; \bar{\mathbf{z}}, \Phi)$ denotes the Gaussian density with mean $\bar{\mathbf{z}}$ and covariance Φ , \mathbf{z} is the reference image (or *latent image* [5]), and T is a matrix representing a 2-D transformation. A frame is therefore ‘generated’ by drawing a reference image \mathbf{z} from $\mathcal{N}(\mathbf{z}; \bar{\mathbf{z}}, \Phi)$, applying a transformation T drawn with probability $P(T)$ from a set of allowable transformations and then adding noise according to the covariance Ψ . The model parameters are the mean reference image $\bar{\mathbf{z}}$, the covariances Ψ and Φ , and the transformation probabilities $P(T)$. These are estimated using the EM algorithm over all of the frames, with the reference images and transformations treated as hidden variables. For our application, the key parameter is the covariance matrix Φ , the diagonal elements of which indicate agreement or otherwise with the ‘transformation of a reference image’ model. Thus, pixels with high covariance are likely to be within regions exhibiting motion patterns distinct from the global motion and hence are either potential motion parts or background clutter.

We found it beneficial to first isolate the animal within a bounding box in each frame prior to registration. As well as minimizing the number of pixels used in the EM algorithm, this also minimizes the likely occurrence of background clutter which may be

confused as motion parts. To do this we adopt a feature tracking approach which identifies significant points on the animal and in the background, and then uses RANSAC based segmentation to distinguish between them assuming a stationary background [6]. The bounding box isolating the animal is then set according to the distribution of foreground points. As an example, Fig. 2a shows tracked feature points [12] and the bounding box computed for a frame from a Hyena sequence. It turns out that we also need the tracked points to estimate the trajectories associated with motion parts and thus this pre-processing operation fits well within the algorithm framework.

We then apply the registration algorithm to the pixels within the bounding box. Computationally, it is unrealistic to register every frame, and so we typically use 1 frame in every 5. As in [5], we limit the transformations to horizontal and vertical integer translations. The mean reference image and associated covariance obtained for the Hyena sequence are shown in Fig. 2b-c, where the values in the covariance image are the diagonal elements of Φ . In this and the other experiments, the EM algorithm converged within around 15 iterations. Note the blurred areas in the mean reference image around the leg, head and lower back resulting from the locomotive motion in those regions, i.e. corresponding to potential motion parts. There are also corresponding local maxima in the covariance image and it is these that we use to identify the location of the parts.

3.2 Motion Part Extraction

To extract potential motion parts, we therefore identify local maxima in the covariance image. This gives the relative spatial positions of the parts. We then estimate the fundamental frequency of the underlying motion region based on the trajectories of the tracked feature points and compute correlations of the trajectories between pairs of parts. The local maxima are extracted using a window size set according to the size of the bounding box. The spatial covariance about each maxima then defines the motion part region, i.e. we model the region using an ellipse. Thus, if $\phi(\mathbf{x})$ is the covariance at pixel \mathbf{x} , then \mathbf{p} is the location of a potential motion part if $\phi(\mathbf{p}) > \phi(\mathbf{x}), \forall \mathbf{x} \in \Gamma(\mathbf{p})$, where $\Gamma(\mathbf{p})$ denotes pixels within the window surrounding \mathbf{p} . The spatial covariance is then given by

$$R(\mathbf{p}) = \sum_{\mathbf{x} \in \Gamma(\mathbf{p})} \phi(\mathbf{x})(\mathbf{x} - \mathbf{p})(\mathbf{x} - \mathbf{p})^T \quad (2)$$

which defines the region associated with the part. As an example, the top 10 ranked parts extracted for the Hyena sequence are shown in Fig. 3a, where the ellipses indicate one standard deviation along each of the principle axes. Note that the regions are centred on the high value clusters in the covariance image and hence around areas of high motion activity with respect to the global motion of the animal.

For each motion part region we then estimate horizontal and vertical trajectories across the sequence based on the feature point tracks within the vicinity of the local maximum. We use a weighted average of the tracks, where the weights are given by a Gaussian about the local maximum as defined by the spatial covariance $R(\mathbf{p})$. The vertical trajectories computed for parts 2, 3 and 5 and the horizontal trajectory for part 8 are shown in Fig. 3b. The periodicity of the underlying regions is clearly evident, as is the positive and negative correlation between the parts. From these trajectories we can estimate the periodicity of the patterns by extracting the fundamental frequencies using spectral analysis. Finally, we determine the relationship amongst the motion parts by computing pairwise correlation

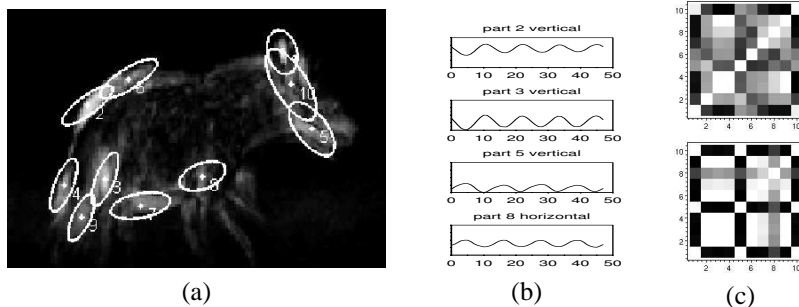


Figure 3: Top 10 ranked motion parts for Hyena sequence: (a) part regions on covariance image; (b) average trajectories for parts 2, 3, 5 and 8; (c) grids of horizontal (top) and vertical correlation coefficients for all pairs of parts (white \equiv positive, black \equiv negative).

coefficients for the horizontal and vertical trajectories. For example, if $h_k(t)$ denotes the horizontal trajectory for part k , then we compute the normalized horizontal correlation coefficient between parts k and l as

$$\rho_{kl}^h = \frac{1}{E_k E_l} \sum_t (h_k(t) - \bar{h}_k)(h_l(t) - \bar{h}_l) \quad (3)$$

where $E_k = \sum_t (h_k(t) - \bar{h}_k)^2$ and \bar{h}_k is the mean value of $h_k(t)$. The horizontal and vertical correlation coefficients amongst the 10 parts extracted from the Hyena sequence are shown as grids in Fig. 3c. These show clear positive correlations between parts 1, 5 and 10 on the head and their negative correlations with those parts on the legs and lower back.

3.3 Representative Configurations

Having determined potential motion parts for each of the training sequences, the next step is to derive a representative configuration for each species. Ideally, this is a configuration which occurs consistently in each set of motion parts. In practice, we seek a configuration which is closest to this with respect to a distance metric. However, computing this turns out to be not straightforward. The key point is that we do not have correspondence between the parts detected in different sequences and thus we are forced to test a potentially large number of permutations in order to find a consistent configuration. To make this tractable, we adopt a similar strategy to that used in [14, 4] and fix the number of parts allowed in the configuration. Limiting the number of parts and using some obvious exclusion rules, then allows a naive exhaustive search amongst the possible permutations. As the results show, this does not affect performance significantly since the number of distinctive motion patterns exhibited by an animal is typically small.

The approach we take is best described by first considering a single pair of training sequences. Using the top N ranked parts from each sequence and configurations consisting of K parts, we find the configuration pair (\hat{C}, \hat{C}') such that

$$(\hat{C}, \hat{C}') = \underset{(C_r, C'_s) \in \Lambda}{\operatorname{argmin}} \mathcal{D}(C_r, C'_s) \quad (4)$$

where $\mathcal{D}()$ denotes the distance metric, C_r and C'_s are the configurations in the two sequences, and Λ is the set of all possible correspondences between all the K component

configurations amongst the N motion parts from each sequence. In other words, (\hat{C}, \hat{C}') are the closest pair of configurations from the two sets of motion parts. A representative configuration is then obtained by taking the average of these - giving the configuration which is closest to at least one configuration from each set of parts. Note that taking the average is possible since the correspondences between the parts in \hat{C} and \hat{C}' are known. Assuming corresponding parts are numbered the same, we use the following for $\mathcal{D}()$:

$$\mathcal{D}(C, C') = w_p \sum_{k=1}^K \|\mathbf{p}_k - \mathbf{p}'_k\|^2 + w_f \sum_{k=1}^K \|\mathbf{f}_k - \mathbf{f}'_k\|^2 + w_\rho \sum_{k=1}^{K-1} \sum_{l=k+1}^K \|\rho_{kl} - \rho'_{kl}\|^2 \quad (5)$$

where \mathbf{p}_k and \mathbf{f}_k denote the relative position and frequency vectors for part k and $\rho_{kl} = (\rho_{kl}^h, \rho_{kl}^v)$ is a vector containing the horizontal and vertical correlations between parts k and l as defined in (3). The weights w_p , w_f and w_ρ normalise the relative contributions of the three terms based on the range of values observed in the training data. We also normalize the position vectors according to the size of the bounding box to give a degree of scale invariance and use the global motion to correct for asymmetries in position caused by animals moving in opposite directions.

Extending the above to more than two sequences is problematic due to the number of permutations. To address this, we adopt an iterative approach which works as follows. We assume that we have a set of motion parts for each training sequence. Denoting the current estimate of the representative configuration by \bar{C}_i , we find the configuration in each set which is closest to \bar{C}_i according to the distance metric in (5). The average of these then forms the updated estimate \bar{C}_{i+1} . Note that we can form this average since the distances are all computed with respect to the same configuration and hence we know the correspondences between all the parts. The process continues until convergence or after a given number of steps. To initialise the process we used a configuration derived from the covariance images for each of the training sequences, based on a clustering of the top ranked motion parts and a weak requirement for sufficient spatial separation of the parts. In the experiments the process typically converged after 4 or 5 iterations.

Having determined the representative configurations, one for each species, we can use these to classify unseen data. For a new sequence, we detect potential motion parts and then determine the configurations within those parts which are closest to each of the representative configurations based on the distance metric in (5). Classification of the sequence is then determined by the smallest distance.

4 Experiments

We used a data set of 140 sequences of approximately 8-10 seconds each, comprising in equal parts individual examples of seven different species in a specific mode of locomotion: Camels (c), Elephants (e), Hyenas (h), Lions (l), Cheetahs (ch), Giraffes (g), and Kangaroos (k). The Hyenas and Kangaroos are all running or hopping, whilst the other animals are all walking. Some sequences were taken using a stationary camera and others with the camera tracking the animal. The majority of animals are well camouflaged and there is a large variation in viewing conditions (see example frames in Fig. 4). There is also considerable clutter, particularly around the animals' legs. Recognition based purely on appearance or shape is therefore likely to be problematic.



Figure 4: Frames from 10 of 140 sequences in the data set.

We trained on 5 sequences per species to learn the representative configurations, and used the remaining 15 in each case for classification. Fig. 5 shows the representative configurations obtained for Hyenas, Camels, Cheetahs and Lions using 10 potential motion parts and 5-part configurations. For example, in Fig. 5a, the leftmost images show the top ten ranked motion parts extracted from two of the Hyena training sequences, and the rightmost images show the representative configuration in terms of its horizontal (top) and vertical components. The same are shown for the other species in Fig. 5b-d. The images illustrating the representative configurations can be interpreted as follows: the ellipses show the covariance on mean part position, the thickness of the connecting lines show the strength of the pairwise correlation between motion parts (positive correlations white, negative correlations black), and the shade of the circle at the centre of each ellipse indicates the periodicity of the trajectories (the darker the circle, the higher the frequency), while the size of the circle indicates the frequency covariance. Note in particular the strong negative and positive correlations between the head and leg motions, and the uncertainty in spatial position of the parts. There are also significant differences between the animals: the part frequencies detected for the Hyenas are very similar, whilst for the Camels and Lions there is noticeable variation (the low part of the neck on the Camel, for instance, has a high relative frequency in both directions); and the pairwise correlations within the configurations are clearly different.

Classification experiments based on the representative configurations were performed using a total of 105 test sequences. Of these, 103 were correctly classified in terms of species, giving an accuracy of over 98%. Two sequences were misclassified: a lion was classified as a cheetah and vice versa, although the distance measures were close. This is perhaps not so surprising given the obvious similarities between the species. In this case we may need to employ additional features, based on appearance or shape, for example, in order to sort out the ambiguity. The confusion matrix illustrating the classification performance is shown on the right in Table 1. Also shown are the confusion matrices resulting when only one of the terms in the distance metric is used, i.e. position, frequency or correlation. The main point to note is that performance is well below that achieved using the combined metric; there is considerable mis-classification across the majority of species. Note in particular that if only relative spatial position of parts is used, then there is considerable error in recognition. Only by incorporating frequency and correlation of motion can the species be successfully distinguished. The one exception is the Giraffe, which can be completely distinguished by position alone, and quite accurately by frequency alone. This is, however, to be expected given its very distinctive shape and gait.

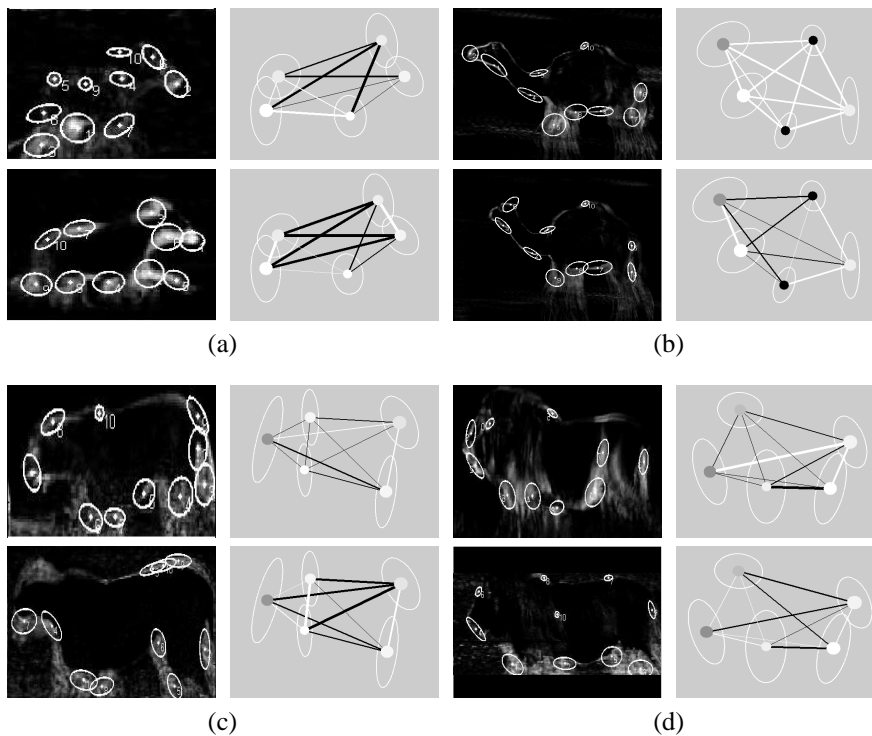


Figure 5: Top 10 ranked motion parts for two training sequences and the corresponding representative configuration (horizontal components top and vertical components bottom) for (a) Hyenas; (b) Camels; (c) Cheetahs and (d) Lions.

5 Conclusions

A method for representing and recognizing animal locomotion has been presented. The results demonstrate that the method is able to successfully distinguish sequences containing the same animal in a similar mode of locomotion with very high accuracy. There are several key contributions in the work. The ‘configuration of motion parts’ model gives flexibility to account for variation within species, whilst at the same time allowing generalization across different species. This extends the ideas used by Song *et al* for representing human motion [13] and further demonstrates the effectiveness of a parts based approach for recognition. In particular, the use of the correlation measure amongst the locomotive parts proved to be a significant distinguishing feature when combined with positional and frequency information. There are, however, many areas for further work. This includes dealing with missing or occluded parts, and the associated uncertainties in part detection. We are investigating whether this can be achieved using a probabilistic framework similar to that used in [14, 4, 13] for object recognition. We are also looking at the recognition of different forms of locomotion within the same species and to extend the technique to deal with examples containing multiple animals and animals viewed from different viewpoints.

Position						Frequency						Correlation						Combined													
	c	e	h	l	ch	g	k		c	e	h	l	ch	g	k		c	e	h	l	ch	g	k		c	e	h	l	ch	g	k
c	7	0	0	2	6	0	0	c	7	2	0	0	0	6	0	c	8	2	2	0	2	1	0	c	15	0	0	0	0	0	0
e	1	3	0	5	6	0	0	e	3	10	0	0	0	2	0	e	1	9	0	1	1	2	1	e	0	15	0	0	0	0	0
h	0	0	11	0	0	3	1	h	0	0	9	0	0	0	6	h	0	2	10	2	0	1	0	h	0	0	15	0	0	0	0
l	1	0	0	6	8	0	0	l	0	0	0	7	5	1	2	l	1	3	0	8	3	0	0	l	0	0	0	14	1	0	0
ch	4	0	0	3	8	0	0	ch	1	0	0	4	9	0	1	ch	1	1	1	1	8	1	2	ch	0	0	0	1	14	0	0
g	0	0	0	0	0	15	0	g	1	1	0	0	0	13	0	g	2	1	0	1	0	10	1	g	0	0	0	0	0	15	0
k	0	0	6	0	0	1	8	k	0	0	9	0	0	0	6	k	0	4	1	0	0	0	10	k	0	0	0	0	0	0	15

Table 1: Confusion matrices illustrating classification performance using single and all terms in the distance metric.

Acknowledgement This research forms part of the ICBR project within the 3c Research programme of convergent technology research for digital media processing and communications whose funding and support is gratefully acknowledged.

References

- [1] A.F. Bobick and J.W. Davis. The recognition of human movement using temporal templates. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 23(3):257–267, 2001.
- [2] C. Bregler. Learning and recognizing human dynamics in video sequences. In *Proc CVPR*, 1997.
- [3] M.C. Burl, M. Weber, and P. Perona. A probabilistic approach to object recognition using local photometry and global geometry. In *Proc ECCV*, 1998.
- [4] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proc. CVPR*, 2003.
- [5] B.J. Frey and N. Jovic. Transformation-invariant clustering using the EM algorithm. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 25(1):1–17, 2003.
- [6] D. Gibson, N. Campbell, and B. Thomas. Quadruped gait analysis using sparse motion information. In *Proc. ICIP*, 2003.
- [7] A. Kale, A. Sundaresan, A.N Rajagopalan, N.P Cuntoor, A.K Roy-Chowdhury, V. Kruger, and R. Chellappa. Identification of humans using gait. *IEEE Trans on Image Processing*, 13(9):1163–1173, 2004.
- [8] S.A. Niyogi and E.H. Adelson. Analyzing and recognizing walking figures in xyt. In *Proc. CVPR*, 1994.
- [9] D. Ramanan and D. A. Forsyth. Using temporal coherence to build models of animals. In *Proc. ICCV*, 2003.
- [10] J. Rittscher, A. Blake, and S.J. Roberts. Towards the automatic analysis of complex human body motions. *Image and Vision Computing*, 20(12):905–916, 2002.
- [11] C. Schmid. Constructing models for content-based image retrieval. In *Proc CVPR*, 2001.
- [12] J. Shi and C. Tomasi. Good features to track. In *Proc CVPR*, 1994.
- [13] Y. Song, L. Goncalves, and P. Perona. Unsupervised learning of human motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):814–827, 2003.
- [14] M. Weber, M. Welling, and P. Perona. Unsupervised learning of models for recognition. In *Proc. ECCV*, 2000.
- [15] C.Y. Yam, M.S. Nixon, and J.N.Carter. Automated person recognition by walking and running via model-based approaches. *Pattern Recognition*, 37(5), 2004.