Anatomical Landmark Tracking for the Analysis of Animal Locomotion in X-ray Videos Using Active Appearance Models

Daniel Haase and Joachim Denzler

Friedrich Schiller University of Jena, Chair for Computer Vision
Ernst-Abbe-Platz 2, 07743 Jena, Germany
{daniel.haase,joachim.denzler}@uni-jena.de
http://www.inf-cv.uni-jena.de/

Abstract. X-ray videography is one of the most important techniques for the locomotion analysis of animals in biology, motion science and robotics. Unfortunately, the evaluation of vast amounts of acquired data is a tedious and time-consuming task. Until today, the anatomical landmarks of interest have to be located manually in hundreds of images for each image sequence. Therefore, an automatization of this task is highly desirable. The main difficulties for the automated tracking of these landmarks are the numerous occlusions due to the movement of the animal and the low contrast in the x-ray images. For this reason, standard tracking approaches fail in this setting. To overcome this limitation, we analyze the application of *Active Appearance Models* for this task. Based on real data, we show that these models are capable of effectively dealing with occurring occlusions and low contrast and can provide sound tracking results.

Keywords: Active Appearance Models, X-ray Videography, Landmark Tracking, Locomotion Analysis

1 Introduction

An important field of ongoing research in biology, motion science and robotics is concerned with the analysis of how the morphology of animals constrains their locomotion. Discovering the underlying relations means not only obtaining a better understanding of common principles of locomotion, but also learning about the adaptivity of the locomotor system to certain circumstances or gaining a more precise knowledge of evolution [8]. It also provides deep insight into the mechanical properties and self-stabilization techniques of animals, which is, for instance, of great interest for the construction of walking robots.

To enable reliable conclusions regarding these open questions, extensive studies have to be carried out on many specimens across different species. These studies are focused on analyzing the movement of the locomotor system. For the case of bipedal terrestrial locomotion, the parts of interest are mainly the *pelvis*, the *femur* and joints like the hip and knee joints [8]. X-ray videography has gained
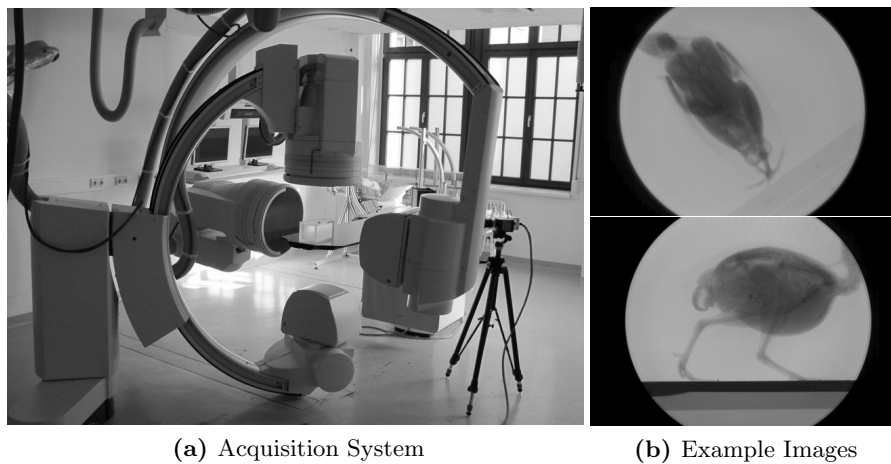
(a) Acquisition System   **(b)** Example Images

Fig. 1. (a) Biplanar high-speed x-ray acquisition system (Neurostar®, Siemens AG). (b) Two example images of a quail (*Coturnix coturnix*) for the dorsoventral (*top row*) and lateral (*bottom row*) camera view acquired with this system.

large popularity for this sort of locomotion analysis over the past decades, as it allows for a relatively unobstructed observation compared to external marker based videography [4, 8]. A typical state-of-the-art x-ray acquisition system is shown in Fig. 1a. Two C-arms allow biplanar recordings at 1000 Hz with a resolution of 1536×1024 pixels. For locomotion analysis, a treadmill is placed on the table to enable recordings of walking animals. Two example images of the dorsoventral (top row) and lateral (bottom row) view obtained with this system are given in Fig. 1b. The images show the locomotion of a quail (*Coturnix coturnix*).

The evaluation of the recorded data is based on anatomical landmarks which have to be located in each image of the sequence. The amount of landmarks differs from sequence to sequence, but common values range from ten to thirty per image. To this day, the labeling task mainly has to be carried out by the human expert, because common tracking algorithms fail due to the overlaps and the low contrast present in the x-ray projections (see Fig. 1b). To speed up the tedious task of manual labeling and to enable the evaluation of large amounts of data, an automatic tracking approach for anatomical landmarks is necessary. The goal of this work is to develop a method which can deal with the problem of overlapping body parts and low contrast x-ray images, and which allows to substantially reduce the human effort spent on manual landmark labeling. In the following we propose the application of Active Appearance Models [5, 7, 6]. The primary reason for the choice of Active Appearance Models is that relationships between landmarks and gray values are modeled in the context of the entire image (*i.e.* globally) and not just locally, which is a promising way of dealing with the problems stated above.

The remainder of this paper is organized as follows. After a short literature review and a motivation for the use of Active Appearance Models in Sect. 2, we will give a brief introduction to these models in Sect. 3. In Sect. 4 we will discuss general aspects and specific properties of Active Appearance Models applied to the scenario of anatomical landmark tracking. The results of our experiments are presented in Sect. 5. At the end we will summarize our findings and discuss future work.

2 Related Work and Motivation

Tracking is an important field of computer vision and a subject of research for many years. It can be distinguished between data-driven and model-based tracking approaches. For the former, prominent representatives are optical-flow-based tracking [10], the "KLT tracker" [1], region-based tracking [9, 11] or trackers based on SIFT descriptors [14]. All these approaches use local image features and allow for a tracking solely based on the given data. The local treatment is the main weakness for the present case, as occlusions in the x-ray images can often only be resolved by using global context information.

Model-based approaches, on the contrary, try to explain the given data by using an underlying model. In the field of medical x-ray analysis, for instance, target regions are tracked by registering a 2D image sequence to a previously recorded 3D computer tomography dataset [15]. In the biological context, this approach is also known as *X-ray Reconstruction of Moving Morphology* (XROMM) [2, 3]. However, this approach is very demanding and complex in our scenario, as not only a full-body computer tomography scan, but also a skeletal model for each specimen need to be provided for each tracking task.

For our application, Active Appearance Models [5, 7, 6] combine the advantages of both tracking principles. On the one hand, training is based on the image sequence and given landmarks, and no explicit model information is necessary in advance. Instead, a combined model of shape and texture is learnt automatically based on the training data. This model describes landmarks and gray values within a combined global framework. Active Appearance Models have been applied to numerous tasks, the most prominent being face modelling and tracking [7, 19] and medical applications (*e.g.* [16]). A non-exhaustive overview of example applications is given by Stegmann [17]. Important extensions of Active Appearance Models for our application are for instance presented by Walker *et al.* [19], who make use of the sequential nature of their data or by Lelieveldt *et al.* [12], who extend Active Appearance Models to multiple camera views.

3 Active Appearance Models

Active Appearance Models [5, 7, 6] are generative statistical models which jointly describe the shape (represented by landmarks) and the appearance (represented by gray values) of non-rigid objects pictured in digital images. The application of such models generally involves two steps, namely the training and the fitting

step. For training, annotated images showing an instance of the object to be modeled are needed. The annotations only consist of landmarks, *i.e.* 2D points which define the shape of the object instance in the according image. In our case, these landmarks consist of parts of the locomotor system (*e.g.* joints) and the torso. Once trained, an Active Appearance Model can be fit to new images in an easy and quick manner. The following two subsections give a brief overview of the basics of Active Appearance Models.

3.1 Training Step

Given the N training images $\boldsymbol{I}_n \in \mathbb{R}^{Y \times X}$, $1 \le n \le N$ and their corresponding M landmarks $\boldsymbol{l}_n = (x_{n,1}, \ldots, x_{n,M}, y_{n,1}, \ldots, y_{n,M})^{\mathrm{T}} \in \mathbb{R}^{2M}$, Active Appearance Models are trained in three sub-steps: the creation of a statistical shape model, a texture model and a combined model. The next passage will give a short overview of these three steps.

Shape model. At first, the combined variation of the landmarks over the training set is analyzed. The goal is to reveal how the position of each landmark correlates with the positions of the other landmarks in order to obtain a specific description of the object's shape. After removing the effects of rotation, scaling and shifting, principal component analysis (PCA) is applied on the centered and aligned landmarks. The result of the PCA is the matrix $\boldsymbol{P}_{\mathrm{L}}$ of shape eigenvectors, which can be used to represent each shape $\boldsymbol{l}'$ via

$$\boldsymbol{l}' = \boldsymbol{l}_0 + \boldsymbol{P}_{\mathrm{L}}\boldsymbol{b}_{\mathrm{L}}, \tag{1}$$

where $\boldsymbol{l}_0$ is referred to as the *mean shape*. The elements of the vector $\boldsymbol{b}_{\mathrm{L}} = \boldsymbol{P}_{\mathrm{L}}^{\mathrm{T}}(\boldsymbol{l}' - \boldsymbol{l}_0)$ are the *shape parameters* of $\boldsymbol{l}'$.

Texture model. The second step is to build a statistical model of the image gray values given in the training data. The approach is very similar to the previous step. The gray values of every training image $\boldsymbol{I}_n$ are warped into a common reference shape. The remaining actions for the texture model follow those from the shape model. Again, PCA is applied and each texture vector $\boldsymbol{g}'$ in the given reference shape can be represented via

$$\boldsymbol{g}' = \boldsymbol{g}_0 + \boldsymbol{P}_{\mathrm{G}}\boldsymbol{b}_{\mathrm{G}}, \tag{2}$$

where $\boldsymbol{g}_0$ is the *mean texture*, $\boldsymbol{P}_{\mathrm{G}}$ are the texture eigenvectors and $\boldsymbol{b}_{\mathrm{G}} = \boldsymbol{P}_{\mathrm{G}}^{\mathrm{T}}(\boldsymbol{g}' - \boldsymbol{g}_0)$ are the *texture parameters* of $\boldsymbol{g}'$.

Combined model. To model the dependencies between shape and texture, an additional PCA is applied on the vectors $\boldsymbol{c}_n = (w\boldsymbol{b}_{\mathrm{L,n}}^{\mathrm{T}}, \boldsymbol{b}_{\mathrm{G,n}}^{\mathrm{T}})^{\mathrm{T}}$, where $\boldsymbol{b}_{\mathrm{L,n}}$ and $\boldsymbol{b}_{\mathrm{G,n}}$ are the shape and texture parameters for the nth training example and $w \in \mathbb{R}$ is a scaling factor (to account for the different units of shape and

texture). In the end, each object instance with the concatenated parameters $\boldsymbol{c}'$ can be represented by

$$\boldsymbol{c}' = \boldsymbol{P}_\mathrm{C}\boldsymbol{b}_\mathrm{C}. \tag{3}$$

Here, $\boldsymbol{b}_\mathrm{C}$ are the *combined parameters* or *appearance parameters* and the matrix $\boldsymbol{P}_\mathrm{C}$ are the combined eigenvectors. If this matrix is restricted on the first eigenvectors which explain a certain amount of model variance, a vast dimension reduction can be achieved. This typically leads to statistical combined models which are capable of explaining the appearance of an object with a very compact set of appearance parameters.

3.2 Model Fitting

Model fitting describes the process of finding suitable appearance parameters for a given model such that the model instance fits a previously unseen image. As every Active Appearance Model describes one specific object, it can be assumed that all fitting tasks are similar. Therefore, we do not need to carry out a separate time-consuming optimization each time we see a new image, but instead can learn the solution for these similar tasks in an offline step. This is achieved by using multivariate regression where parameter changes are predicted based on the texture difference between the model and the real image. The necessary training data is obtained by systematically displacing known model instances from the training set. Once learnt, this relationship is used to iteratively fit a model instance to a given image in a quick and easy way.

4 Application to X-ray Locomotion Landmark Tracking

The general application of Active Appearance Models for tracking tasks in video sequences is straightforward and has been widely discussed in the literature, however mainly under the aspect of face tracking [7, 19]. For the application to high-speed x-ray locomotion sequences, there are two important differences compared to the case of usual tracking. First of all does the training data not consist of miscellaneous instances of the object to be modeled (*e.g.* a face database for face tracking), but rather of images taken from the sequence to be tracked itself. The reason for this approach is that often only one sequence per species is available or that available sequences differ considerably, either in their visual appearance or in the labeled landmarks.

The second specific characteristic compared to usual Active Appearance Model tracking is the property which is characterized by the shape model. Instead of the variation of landmarks between static instances of an object, the shape model describes the *dynamic* variation of landmarks during the locomotion of one specific specimen. Therefore, the shape model becomes actually a very basic locomotion model. An example for this effect can be seen in Fig. 2. It shows the first two eigenmodes of the statistical shape model for the trunk and *femora* (thighs) landmarks of a quail trained on the lateral view of the dataset shown in Fig. 1b. The first eigenmode explains 85% of the total shape variation,
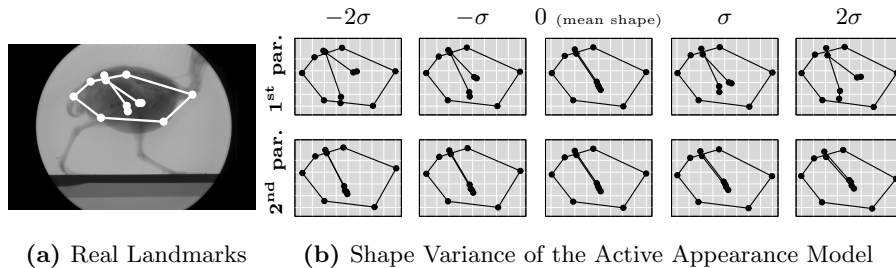
(a) Real Landmarks   (b) Shape Variance of the Active Appearance Model

Fig. 2. Shape variance of the trunk and *femora* landmarks for the lateral view of the dataset shown in Fig 1b. The first and second shape parameters explain 85% and 11% of the total variance of the landmarks, respectively. Due to the specific application, the statistical shape model actually characterizes a very basic locomotion model.

and it can be seen that it mainly expresses the forward and backward *femora* movement during the locomotion. Therefore, the first eigenmode of the shape model roughly corresponds to the angle which is spanned between the two projections of the *femora*. The second eigenmode shows a large movement of the rightmost landmark relative to the rest of the trunk. As this particular landmark corresponds to the quail's 5th *vertebra*, the second eigenmode models the typical cervical movement of a quail while walking.

5   Experiments and Results

The experiments presented in the following were performed on the quail dataset as shown in Fig. 1b. The sequence has a total length of 2.245 s (2245 images) and covers $5^{1}/_{2}$ walking periods (about 11 strides) at a resolution of 1536×1024 pixels. Because the labeling for all sequences was done by human experts so far, plenty of groundtruth data is available. For this data set, the groundtruth data consists of 10 and 12 anatomical landmarks for the dorsoventral and the lateral view in 68 and 81 images of the sequence, respectively. Approximately every 20th image was labeled. The landmarks of interest cover the 5th *vertebra* (neck), the *pelvis*, the *acetabula* (hip joints), the *pygostyle* (pearson's nose), the *caudal carina* (rear breastbone), the *furcula* (wishbone) and the knee joints. The most part of the tracking relevant occlusions occur in the region of the knee joints.

In our experiments, we wish to investigate the following issues:

(1) Are Active Appearance Models suitable for this kind of tracking task?
(2) Does the global modeling lead to better results compared to local methods?
(3) How do image size and preprocessing influence the tracking quality?
(4) Which and how many images of a sequence are best suited for training?
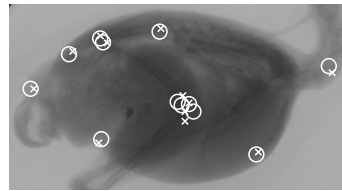
Based on the *point to point error* [17], which is the Euclidian distance between tracking result and groundtruth landmark position, we examine general suitability, the *generalization ability* and the *model accurateness* for various scenarios.
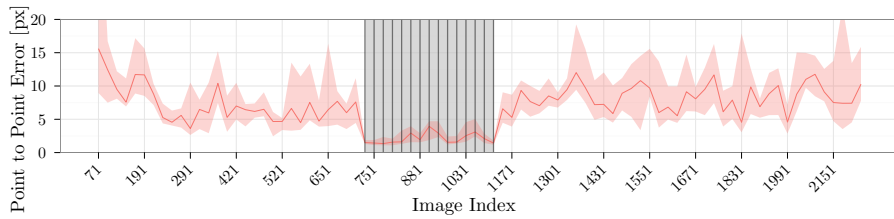
(a) Point to Point Errors
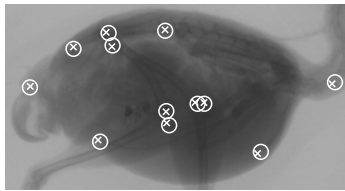
(b) Result for Image 261

(c) Result for Image 1361

Fig. 3. (a) Point to point errors of the tracked landmarks of the lateral view using a basic Active Appearance Model compared to the groundtruth landmark positions. For each image index, the median error as well as the first and third quartiles of the landmark errors are shown. The training images are selected from equally spaced frames (indicated by vertical lines) of the walking period marked with a shaded background. Subfig. (b) and (c) show the tracking results for the images 261 and 1361 in detail, where crosses and circles denote tracked and groundtruth landmarks.

5.1 General Suitability

Proof-of-Concept. As a general proof-of-concept for the application of Active Appearance Models to this kind of tracking task, we trained a basic model on the given dataset. As training set we chose 15 images evenly spread over one walking period in the middle of the sequence. After a coarse initialization of the landmark positions for the first image of the sequence, the landmarks were tracked solely based on the trained model and without any further user interaction. To ensure temporal consistency, we used the result of frame t as initial solution for frame $t + 1$. Fig. 3a shows the point to point errors of the tracked landmarks for each image having groundtruth data available. For each image index, the median error and the first and third quartiles of the landmark errors are shown as a measure of accuracy and precision, respectively. The images used for training are indicated by vertical lines and the according walking period is marked with a gray background.

First of all, the difference between the results on the training and non-training images is clearly visible. Both accuracy and precision are about two to four times larger in the previously unseen images compared to the training images. However, median errors of ten pixels for the non-training images are a promising result, taking the image resolution of 1536×1024 pixels into account. To sup-

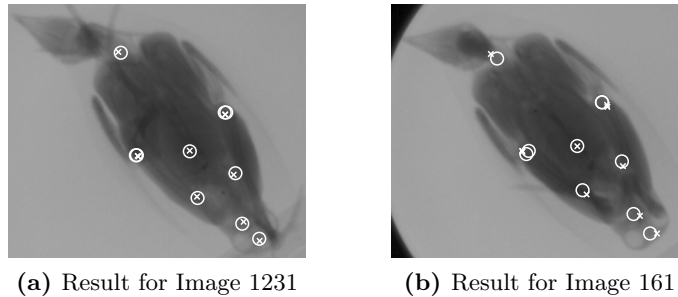**(a)** Result for Image 1231   **(b)** Result for Image 161

Fig. 4. Tracking results for the images (a) 1231 and (b) 161 of the dorsoventral camera view. Image 1231 has one of the best and image 161 has one of the worst tracking results on the non-training images of the sequence.

port this claim, Fig. 3b and 3c show the detailed tracking result for the images 261 and 1361, where crosses and circles denote tracked and groundtruth landmarks, respectively. From Fig. 3a it can be seen that image 261 has one of the best tracking results amongst the non-training images. This observation can be verified by Fig. 3b. Image 1361, on the contrary, has one of the worst performances of the tracked sequence according to the point to point error. This result is also visible in Fig. 3c, where especially the knee joint landmarks are imprecise. One reason for the different tracking qualities of these two images is probably the amount of relevant occlusions (*i.e.* occlusions of the *femora*, hip joints and knee joints), which differs substantially in the two images.

Nevertheless, above results show that Active Appearance Models are capable of dealing with the difficulties of the given data. Despite the considerable occlusions, no landmark is completely lost. In contrast to the human expert, no temporal model or anatomical knowledge was exploited. However, the tracking accuracy is promising for real applications, and it has to be considered that even the hand-labeled groundtruth landmarks may deviate from their true anatomical positions by several pixels.

For the dorsoventral view, the results are closely related to those from the lateral view. For a similar training set, the point to point error curve is akin to the one of the lateral view as shown in Fig 3a. In Fig. 4a and 4b, examples for one of the best and worst tracking results of the dorsoventral sequence are shown in detail. Again, both results demonstrate that Active Appearance Models can handle the existent difficulties of the data very well.

Comparison to Local Approaches. In contrast to global approaches like the Active Appearance Models, local tracking methods are likely to fail in this setting. To verify this claim, we tested the Horn-Schunck optical flow tracking method [10] on the same dataset. As expected, the results show that indeed this method is adequate to track landmarks which are not subject to occlusions, like the 5th *vertebra*, the *pelvis* or the *pygostyle*. The other landmarks, however, were irretrievably lost as soon as occlusions occurred in the x-ray projections

Table 1. Computational considerations for identical models trained on various image scales of the quail dataset. Training was performed on 15 images taken from one walking period. The tracking was performed on the entire sequence of 2245 images. The landmark errors were calculated on the non-training images of the sequence only. All error values refer to the original image size of 1536×1024 pixels.

Image Size	Texture Size	Computing Time		Error Quartiles		
		Training	*Tracking*	1st	2nd	3rd
100.0%	146,055 px	95.4 min	362.4 min	3.43 px	6.45 px	10.73 px
50.0%	36,432 px	29.2 min	82.7 min	3.39 px	6.45 px	10.93 px
25.0%	9,068 px	6.2 min	19.8 min	3.52 px	6.81 px	10.96 px
12.5%	2,257 px	1.7 min	4.6 min	4.09 px	7.33 px	12.16 px

due to the locomotion of the quail. This result underlines another advantage of Active Appearance Models for this tracking task, which can usually recover after suboptimal model fits in the image sequence.

5.2 Impact of Image Resolution and Preprocessing

Image Resolution. For 15 training images, the learning and tracking step of an Active Appearance Model took about 7.63 h on a modern desktop PC (Intel® Core™ i5 CPU 760 @ 2.80 GHz). We therefore examined the performances of Active Appearance Models for several resolutions of the input data in order to find out whether full resolution images are necessary. The experiments were made for image scales of 100% (1536×1024), 50% (768×512), 25% (384×256) and 12.5% (192×128). The results of these experiments are listed in Tab. 1. It can be seen, that despite the enormous differences in the training and tracking time, the results for image scales of 100%, 50% and 25% do not deviate substantially. For a scale of 12.5%, however, the loss of quality due to the resolution reduction becomes apparent.

Preprocessing. Another important aspect we analyzed was how sensitive Active Appearance Models react on different methods of preprocessing of the input data. We compared the performance of Active Appearance Models applied to (1) the original data with background-subtracted and contrast scaled images, (2) images sharpened based on the Laplacian operator and (3) gradient images. All three methods have their justification, as the first two improve the contrast in the images and reveal faint structures, whereas the latter emphasizes the anatomical structures of interest, such as the *femur*. The results, as listed in Tab. 2, however, show that only the first two methods benefit the tracking performance, while the gradient approach even worsens the result. This result suggests that homogeneous areas in the image (like certain organs) are important for the fitting process. Based on these findings, we used background-subtracted and contrast scaled images for the majority of the conducted experiments.

Table 2. Influence of different preprocessing methods on the tracking performance.

Preprocessing	Error Quartiles		
	1st	2nd	3rd
Original (no preprocessing)	3.77 px	7.40 px	11.85 px
Background subtraction & contrast stretching	3.52 px	6.81 px	10.96 px
Laplace-based enhancement	3.45 px	6.44 px	10.99 px
Gradient image (Sobel)	4.03 px	8.59 px	15.77 px

5.3   Selection of Training Images

For real applications, the amount of human effort spent in landmark labeling to create training data is the main limiting factor for data evaluation. Our goal is to achieve the desired tracking quality with as much as necessary, but as little as possible human interaction. Therefore, two very important questions about the given training data arise. The first one is how much training images are actually necessary to achieve the desired tracking quality, and the second question is which images of the given sequence are most suitable as training images. To answer these questions, we trained several Active Appearance Models with varying sizes of the training set and selection schemes. In the first case we started with images entirely taken from one walking period. Then, we successively added new images, one walking period at a time. In the end, the training size ranged from 3 images ($^1/_5$ walking period) to 58 images (4 walking periods). In the other case, the same image amounts were used, but the images were selected from equally spaced images of the entire sequence.

We evaluated the experiments in two different ways. In one case we only used the unseen images of the series for testing, which gives the generalization ability of the according models. In the second case, only the errors made on the training set were evaluated, giving an estimation for the model accurateness. The evaluations for both cases are shown in Fig. 5a and 5b. It can be noticed that the results for both selection methods differ substantially. Considering both the generalization ability and the model accurateness, the period-based method seems to be the better choice for few training examples, whereas the equally spaced selection gives better results for many training images. In both cases, the turning point is located around 15 images, which is the maximum amount of images taken from one period. That is, as long as images from one period are to be used for training, the period-based method is to be preferred. As soon as images from more than one walking period should be used, the equally spaced selection is more advantageous. The reason for this result is quite clear: few equally spaced images will generally not cover all parts of a walking period, which is a disadvantage compared to period-based selection. For many images, equally spaced methods will perform better because they cover all walking periods of the sequence in contrast to the period-based selection.
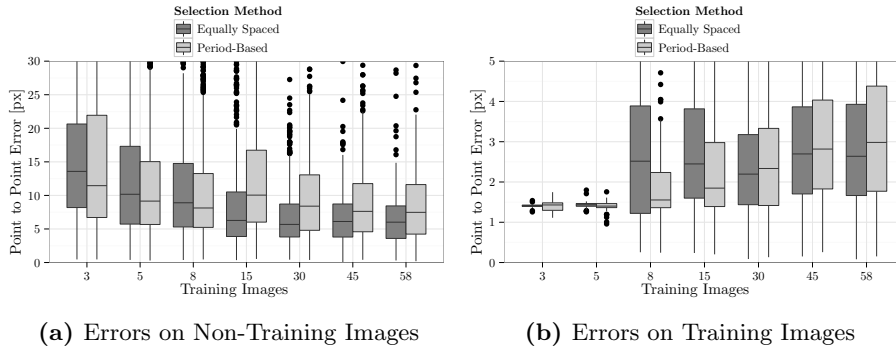
(a) Errors on Non-Training Images   **(b)** Errors on Training Images

Fig. 5. Tracking errors on (a) non-training images (generalization ability) and (b) on training images (model accurateness) for varying sizes and selection schemes of the training set. For few training examples, the period-based selection appears to be more advantageous.

6   Conclusions and Further Work

We analyzed the application of Active Appearance Models for anatomical landmark tracking in x-ray videos of animal locomotion. As landmarks and gray values are modeled in a global manner, these models are well suited to deal with occlusions and low contrast in images. We showed that the Active Appearance Model based approach performed substantially better than local approaches on real data. We also studied the effect of different preprocessing and image selection methods on the performance and generalization ability of the models.

Further work should focus on the combined modelling of both camera views (for instance based on [12]) to improve the performance for frames with high occlusion and thus uncertainty. Additionally, the knowledge that the training data is actually a sequence and not just a set of images should be exploited. For the reduction of user interaction, bootstrapping methods for Active Appearance Models based on [18, 13] could be utilized. Another important issue will be to derive a confidence value from the texture error of the Active Appearance Models to automatically detect ill-fitted frames.

References

1. Baker, S., Matthews, I.: Lucas-kanade 20 years on: A unifying framework. Int. J. Comput. Vision 56(3), 221–255 (2004)
2. Bey, M.J., Zauel, R., Brock, S.K., Tashman, S.: Validation of a new model-based tracking technique for measuring three-dimensional, in vivo glenohumeral joint kinematics. J. Biomech. Eng. 128, 604–609 (2006)
3. Brainerd, E.L., Gatesy, S.M., Baier, D.B., Hedrick, T.L.: A method for accurate 3D reconstruction of skeletal morphology and movement with CTX imaging. Comp. Biochem. Physiol. 146, 119 (2007)
4. Brainerd, E.L., Baier, D.B., Gatesy, S.M., Hedrick, T.L., Metzger, K.A., Gilbert, S.L., Crisco, J.J.: X-ray reconstruction of moving morphology (XROMM): Precision, accuracy and applications in comparative biomechanics research. J. Exp. Zool. A 313A(5), 262–279 (2010)
5. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. In: Burkhardt, H., Neumann, B. (eds.) Proceedings of the 5th European Conference on Computer Vision. LNCS, vol. 1407, pp. 484–498. Springer (1998)
6. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE T. Pattern Anal. 23(6), 681–685 (2001)
7. Edwards, G.J., Cootes, T.F., Taylor, C.J.: Face recognition using active appearance models. In: Burkhardt, H., Neumann, B. (eds.) Proceedings of the 5th European Conference on Computer Vision. LNCS, vol. 1407, pp. 581–595. Springer (1998)
8. Gatesy, S.M.: Guineafowl hind limb function. I: Cineradiographic analysis and speed effects. J. Morphol. 240(2), 1097–4687 (1999)
9. Hager, G.D., Belhumeur, P.N.: Efficient region tracking with parametric models of geometry and illumination. IEEE T. Pattern Anal. 20(10), 1025–1039 (1998)
10. Horn, B.K.P., Schunck, B.G.: Determining optical flow. Artif. Intell. 17(1-3), 185–203 (1981)
11. Jurie, F., Dhome, M.: Hyperplane approximation for template matching. IEEE T. Pattern Anal. 24(7), 996–1000 (2002)
12. Lelieveldt, B., Üzümcü, M., van der Geest, R., Reiber, J., Sonka, M.: Multi-view active appearance models for consistent segmentation of multiple standard views. International Congress Series 1256, 1141–1146 (2003)
13. Liu, X.: Video-based face model fitting using adaptive active appearance model. Image Vision Comput. 28(7), 1162–1172 (2010)
14. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60(2), 91–110 (2004)
15. Rohlfing, T., Denzler, J., Gräßl, C., Russakoff, D.B., Jr., C.R.M.: Markerless real-time 3-d target region tracking by motion backprojection from projection images. IEEE T. Med. Imaging 24(11), 1455–1468 (2005)
16. Roussos, A., Katsamanis, A., Maragos, P.: Tongue tracking in ultrasound images with active appearance models. In: Proceedings of the IEEE International Conference on Image Processing. pp. 1733–1736 (2009)
17. Stegmann, M.B.: Active Appearance Models: Theory, Extensions and Cases. Master's thesis, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby (2000)
18. Sung, J., Kim, D.: Adaptive active appearance model with incremental learning. Pattern Recogn. Lett. 30(4), 359–367 (2009)
19. Walker, K.N., Cootes, T.F., Taylor, C.J.: Automatically building appearance models from images sequences using salient features. In: Proceedings of the British Machine Vision Conference. British Machine Vision Association (1999)