

RESEARCH

Open Access

# 2D and 3D analysis of animal locomotion from biplanar X-ray videos using augmented active appearance models

Daniel Haase\* and Joachim Denzler

## Abstract

For many fundamental problems and applications in biomechanics, biology, and robotics, an in-depth understanding of animal locomotion is essential. To analyze the locomotion of animals, high-speed X-ray videos are recorded, in which anatomical landmarks of the locomotor system are of main interest and must be located. To date, several thousand sequences have been recorded, which makes a manual annotation of all landmarks practically impossible. Therefore, an automatization of X-ray landmark tracking in locomotion scenarios is worthwhile. However, tracking all landmarks of interest is a very challenging task, as severe self-occlusions of the animal and low contrast are present in the images due to the X-ray modality. For this reason, existing approaches are currently only applicable for very specific subsets of anatomical landmarks. In contrast, our goal is to present a holistic approach which models all anatomical landmarks in one consistent, probabilistic framework. While active appearance models (AAMs) provide a reasonable global modeling framework, they yield poor fitting results when applied on the full set of landmarks. In this paper, we propose to augment the AAM fitting process by imposing constraints from various sources. We derive a general probabilistic fitting approach and show how results of subset AAMs, local tracking, anatomical knowledge, and epipolar constraints can be included. The evaluation of our approach is based on 32 real-world datasets of five bird species which contain 175,942 ground-truth landmark positions provided by human experts. We show that our method clearly outperforms standard AAM fitting and provides reasonable tracking results for all landmark types. In addition, we show that the tracking accuracy of our approach is even sufficient to provide reliable three-dimensional landmark estimates for calibrated datasets.

**Keywords:** Active appearance models; X-ray videography; Landmark tracking; Animal locomotion analysis

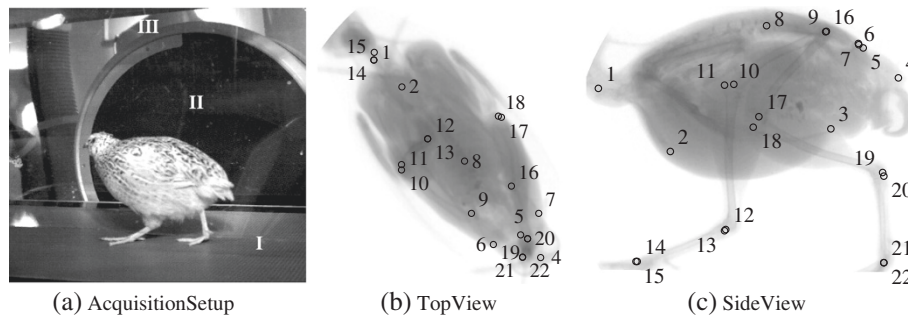
## 1 Introduction

For many fundamental problems of ongoing research in biomechanics, zoology, evolutionary biology, and robotics, the key element is a thorough knowledge on animal locomotion [1-8]. Ideally, this knowledge is obtained by analyzing skeletal movements of locomoting animals. While many methods have been developed over time, the state-of-the-art approach for obtaining noninvasive *in vivo* measurements of the locomotor system is biplanar X-ray videography. In contrast to reflective marker-based methods, it allows for unobstructed observations at an unrivaled accuracy [5]. In general, the animal to

be analyzed is placed on a treadmill and filmed from a side camera view (*lateral* camera) and a top camera view (*dorsoventral* camera) at a very high frequency, usually 1,000 frames per second. A typical experimental setup is shown in Figure 1.

For an evaluation of acquired data, anatomical landmarks - usually skeletal joints of the locomotor system such as hip joints, knee joints, intertarsal joints, and phalangeal joints [6,7] - have to be located in the images. Most evaluations to date solely rely on human experts (e.g., [5,6]), which is an extremely time-consuming process and complicates the realization of large-scale studies. An automation of this process would therefore greatly benefit research in the aforementioned areas [9]. However, as almost all parts of an animal's skeletal system undergo severe self-occlusions during locomotion

\*Correspondence: daniel.haase@uni-jena.de  
Computer Vision Group, Friedrich Schiller University of Jena, Ernst-Abbe-Platz  
2-4, Jena 07743, Germany



**Figure 1 X-ray acquisition system.** Biplanar X-ray acquisition system (I, treadmill; II, lateral X-ray detector; III, dorsoventral X-ray detector) used for locomotion analysis (a). Example data of a quail with typical landmarks (1 to 8, torso; 9 and 16, hip joints; 10, 11, 17, and 18, knee joints; 12, 13, 19, and 20, intertarsal joints; 14, 15, 21, and 22, feet) are shown in (b, c).

(cf. Figure 1), developing fully automatic tracking methods for this application is a challenging task.

In this paper, we address the issue of landmark tracking in X-ray sequences of grounded locomotion of birds. We present a novel method which, unlike previous approaches, is able to track all landmarks used in locomotion analysis and can overcome many other practically relevant drawbacks of existing methods (see Subsection 1.2) using a unified, consistent, and probabilistic framework that combines the complementing paradigms of model-driven and data-driven tracking.

### 1.1 Related work

For very simple scenarios of locomotion analysis, straightforward tracking approaches such as template matching can be applied [9]. Due to severe occlusions, however, template matching and a variety of other standard methods such as optical flow/KLT and its extensions [10-12], region tracking [13,14], and SIFT-based tracking [15] were proven to be unsuited for X-ray analyses in the challenging scenario at hand [16,17]. A more advanced approach for skeletal tracking is based on image registration between recorded X-ray images and backprojected CT scans [3,18,19]. However, in most cases this method is only feasible for medical applications, as a full CT scan is necessary for each subject to be analyzed.

An alternative, completely data-driven approach for robust template tracking in X-ray sequences was recently proposed in [16]. As standard template tracking fails due to the severe occlusions, the idea is to divide the template to be tracked into certain sub-templates. For each frame, all sub-templates are matched to the target image individually, and the results of these sub-templates are then merged to obtain one consistent parameter transformation for the whole template. The important difference between [16] and existing sub-template-based approaches such as [20-22] lies in the fusion of sub-template results. While previous approaches employ a

hard decision between occluded and non-occluded sub-templates, the authors in [16] use a soft decision which exploits special properties of X-ray images. It has proven to be well suited for X-ray bone tracking under moderate occlusions (e.g., for the lower leg landmarks in the side view) [16]. However, due to its data-driven nature, landmarks undergoing severe occlusions (landmarks occluded by the torso, e.g., knee landmarks of the side view or feet landmarks of the top view, cf. Figure 1) cannot be handled.

To overcome such problems of data-driven approaches, model-driven methods generally are able to estimate landmark positions - even for total occlusions - by using global context. One prominent example of global models are *active appearance models* (AAMs) [23-25]. Besides many applications for human face modeling (e.g., [24,26,27]) and medical image analysis (e.g., [28,29]), AAMs have also been successfully applied to landmark tracking in X-ray locomotion scenarios [17,30]. One major problem in our scenario, however, is that the movement of the animals often is very complex. As a result, especially for the lower legs, landmark configurations during locomotion substantially differ from the mean landmark configuration, i.e., the motions are non-stationary [31,32]. As discussed in [31] and [33], this situation drastically complicates the fitting of AAM-like models. Besides the non-stationary motion, another major problem is the *non-discriminative texture information* of the lower leg landmarks (cf. landmarks 12 to 15 and 19 to 22 in Figure 1b), which additionally complicates the fitting process of AAMs. Thus, the aforementioned standard AAM-based approaches only work when neglecting the set of non-stationary landmarks, as in [17,30,34].

To combine the benefits of data-driven and model-driven methods, several hybrid models were developed over time. One straightforward example are *combined local models* [35], where the shape is modeled globally, as for AAMs, but the texture is modeled locally around each landmark. A recently proposed probabilistic example

of this approach are *discriminative Bayesian active shape models* [36], where many local detectors are used to estimate a global landmark configuration. Both approaches, however, model landmark motions similarly to AAMs and are thus very likely to suffer from the same problems as well.

## 1.2 Motivation

As mentioned in the last subsection, data-driven [16] as well as model-driven approaches [17,30] exist for landmark tracking in X-ray locomotion analysis. However, all previously published works in this field suffer from at least one of the following shortcomings, which is a major drawback for the usage of these methods for actual zoological and biomechanical studies:

- Only very specific anatomical landmark subsets can be tracked, e.g., the torso landmarks [17,30] or the lower leg landmarks [16]. In addition, certain landmarks exist which are covered by neither of the current approaches, e.g., the lower leg landmarks of the top camera view (*cf.* landmarks 12 to 15 and 19 to 22 in Figure 1b).
- Although model-driven and data-driven approaches can generally complement each other (e.g., [36]), only one paradigm is used at a time [16,17,30,34].
- For data-driven approaches, merely landmarks of the side camera view are considered due to severe self-occlusions in the top camera view [16].
- Modeling, tracking, and evaluation are performed solely using two-dimensional (2D) approaches, although an accurate camera calibration can be obtained for newly recorded datasets [16,17,30,34].
- The validation of the tracking methods is performed only on very few real-world datasets, as obtaining ground-truth landmarks is a tedious and time-consuming task [17,30,34].

As a consequence, while model-driven as well as data-driven approaches exist for very specific landmark subsets, neither of them alone is applicable for the full tracking problem. The trivial option of simply merging their results is not an option, because on the one hand the landmark subsets would be tracked independently of another and hence would not be consistent. On the other hand, not all landmarks would be covered by these methods, as for instance the lower leg landmarks in the top view (*cf.* Figure 1b). Our goal in this work is to overcome all drawbacks mentioned above and to present an approach which is holistic in the sense that all landmarks of the animal are modeled in one consistent framework. We base the approach on the fact that existing methods [16,30] are complementary, i.e., the first method works well on a landmark subset the second method is unsuited for and vice

versa. Our main idea is to unify these ‘subset approaches’ within a probabilistic framework to obtain consistent estimates for all landmarks. While AAMs applied on the full set of locomotion landmarks yield poor fitting results, they are still well suited for modeling interrelationships between landmarks. Therefore, we use AAMs as base model for our approach. However, in contrast to standard AAMs, we augment the fitting process by imposing constraints obtained from sources such as subset methods [16,30]. We first derive a probabilistic framework that allows AAM fitting under arbitrary types of constraints. While similar approaches such as [36] and [37] only utilize positional priors, we aim to include additional constraints, e.g., the anatomical context or the epipolar geometry of the camera setup. As opposed to existing works in this field, this framework allows to consistently incorporate all landmarks of both camera views while combining the advantages of data-driven and model-driven approaches. In addition, we evaluate our approach based on 32 real-world datasets from three zoological studies [6,7,34], including 175,942 manually labeled ground-truth landmarks and birds of different morphology and locomotion characteristics, which by far exceeds the amount of data used in recent studies. An outline of our approach is shown in Figure 2.

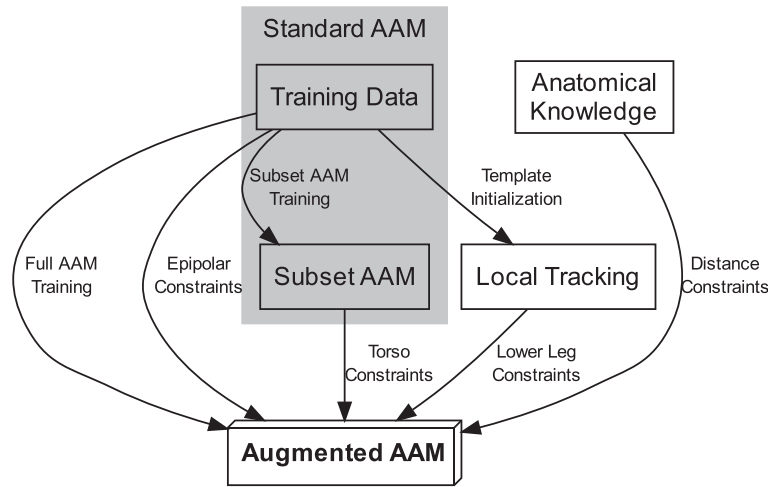
The remainder of this paper is structured as follows. First, an overview of standard AAMs is given in Section 2, as AAMs form the baseline of our method. In Section 3, we present augmented AAMs as our approach for landmark tracking in X-ray locomotion sequences. After deriving a general fitting framework, we describe the constraints used in our specific case. The validation of our approach is presented and discussed in Section 4.

## 2 Active appearance models

This section gives an overview of standard AAMs [23-25], which form the baseline of our augmented approach presented in Section 3. AAMs are parametric statistical models which describe the visual appearance of arbitrary object classes. The variation in object appearance is modeled by a shape component (represented by image landmarks) and a shape-free texture component. AAMs are trained from sample images with annotated landmark positions. Once learned, a trained model can be fit to unseen images automatically. In the following subsections, the basic training and fitting procedure of standard AAMs will be described.

### 2.1 AAM training

AAM training is based on annotated sample images, i.e.,  $N$  images  $I_1, \dots, I_N$  with  $M$  corresponding landmarks  $\mathbf{l}_n = (x_{n,1}, y_{n,1}, \dots, x_{n,M}, y_{n,M})$ ,  $1 \leq n \leq N$ . As first step, the shape model is built by aligning the given



**Figure 2 Overview of augmented AAMs.** Our holistic approach for landmark tracking in X-ray locomotion sequences. The fitting process of standard AAMs is augmented by imposing constraints from methods that only perform well on a subset of landmarks ('subset AAM' and 'local tracking'). Additionally, further knowledge such as anatomical constraints and epipolar priors are included.

shape samples  $I_1, \dots, I_N$  with respect to translation, rotation, and scale via *Procrustes analysis* [38,39], resulting in shapes  $s_1, \dots, s_N$ . The shape variations are then parameterized by applying *principal component analysis* (PCA) to the matrix  $S = (s_1 - \bar{s}, \dots, s_N - \bar{s})$ , where  $\bar{s} = \frac{1}{N} \sum_{n=1}^N s_n$  is the *mean shape*. The result is a linear model which describes an arbitrary shape  $s$  based on its *shape parameters*  $b_s$ , the *shape eigenvectors*  $P_s$ , and the *mean shape*  $\bar{s}$  of all samples via

$$s = \bar{s} + P_s b_s. \quad (1)$$

An example of an AAM shape model is shown in Figure 3 for an animal locomotion dataset used in this paper. It demonstrates that the movements of the lower legs are very complex in both camera views and thus cannot be handled well in the fitting process of standard AAMs.

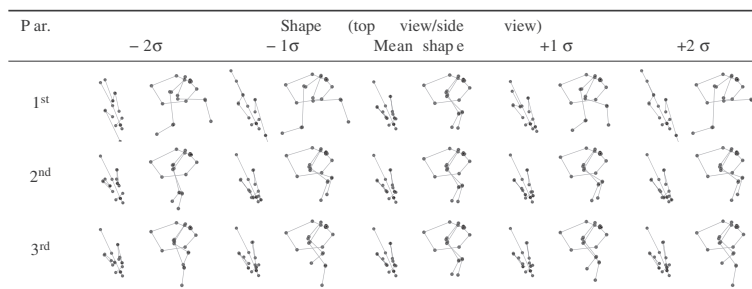
The second step of AAM training consists of building a texture model. Firstly, each image  $I_1, \dots, I_N$  is warped into a common reference frame - usually the mean shape  $\bar{s}$ .

The shape-normalized images are then vectorized, resulting in the texture vectors  $g_1, \dots, g_N$ . Afterwards, the very same PCA-based procedure as for the shape model is employed, which results in the linear texture model

$$g = \bar{g} + P_g b_g, \quad (2)$$

where  $g$  is an arbitrary shape-normalized texture with *texture parameters*  $b_g$ ,  $P_g$  are the *texture eigenvectors*, and  $\bar{g} = \frac{1}{N} \sum_{n=1}^N g_n$  is the *mean texture* of the samples.

To obtain a combined representation of both shape and texture, the third - albeit optional - step of AAM training is to merge shape and texture parameters into one parameter set. This is achieved by concatenating the variance-normalized shape and texture parameter vectors for each training sample and again applying PCA. Therefore, each object instance can then be represented by its *combined parameters*  $b_c$ . The final parameter count, i.e., the dimension of  $b_c$ , is then reduced by discarding parameters which explain only a small fraction of the total variance.



**Figure 3 Shape components of an AAM trained on a bird sequence.** In this example, variations of the first three shape parameters for the top view (left) and the side view (right) are shown. The movements of the lower legs are very complex in both camera views and cannot be handled well in the fitting process of standard AAMs.

## 2.2 AAM fitting

The goal of AAM fitting is to find the model parameter vector  $\hat{\mathbf{b}}_c$  that best fits an object instance shown in a given input image. Technically, the optimization criterion is to minimize the squared difference  $\delta\mathbf{g} = (\mathbf{g}_{\text{image}} - \mathbf{g}_{\text{model}})$  between the given image and the synthesized appearance of the AAM instance, i.e.,

$$\hat{\mathbf{b}}_c = \underset{\mathbf{b}_c}{\operatorname{argmin}} \delta\mathbf{g}^\top \delta\mathbf{g}. \quad (3)$$

In its original formulation [23-25], this problem was solved in an iterative manner by assuming a linear relationship  $\delta\mathbf{b}_c = \mathbf{A}\delta\mathbf{g}$  between the necessary model parameter changes  $\delta\mathbf{b}_c$  and the current image difference  $\delta\mathbf{g}$ , where  $\mathbf{A}$  can be learned in advance. In general, however, such a simple constant relationship between  $\delta\mathbf{b}_c$  and  $\delta\mathbf{g}$  does not exist, which can lead to suboptimal fitting results [40]. An alternative optimization approach for Equation 3 is the *inverse compositional/project-out* algorithm [40]. By decoupling shape and texture parameters, it allows for a very efficient alignment that eliminates many drawbacks of the original AAM fitting method.

Note, however, that our augmented AAM approach presented in Section 3 is independent of the actual optimization scheme - it is possible to base it on both the additive as well as the inverse compositional methods (cf. Subsection 3.1).

## 2.3 Multi-view extension

While standard AAMs can only be used for a single camera view, possible extensions are available for scenarios which contain more than one camera, e.g., [41] or [42]. In our case, a biplanar image acquisition is usual, albeit also monocular sequences exist. In addition, for many previously recorded datasets from biological studies such as [7], a calibration of the camera setup is not available. Therefore, in our locomotion scenario, it is generally not possible to apply any of the methods mentioned above, as they rely on certain assumptions about the scene. However, it is still possible to exploit relationships between multiple camera views using *multi-view AAMs* [43,44], as shown in [30].

The construction of multi-view AAMs is closely related to standard AAMs. Let  $K$  denote the number of camera views. As first step, the aligned landmark vectors  $\mathbf{s}_n^{(1)}, \dots, \mathbf{s}_n^{(K)}$  of all camera views are concatenated into one vector  $\mathbf{s}'_n$ . Afterwards, PCA is applied to obtain the multi-view shape model in the same manner as for standard AAMs. As for the multi-view landmarks, for each training sample the texture vectors of all views are concatenated to form the vector  $\mathbf{g}'_n$  and PCA is applied. Note that this multi-view extension is used in exactly the same manner for augmented AAMs, which are presented in the following section.

## 3 Augmented AAM approach

In the following augmented AAMs, our extension of standard AAMs are presented. As stated in the motivation (cf. Subsection 1.2), the goal is to overcome poor fitting results in cases of non-stationary shape activities [31,32] and non-discriminative texture information, which is particularly true for the locomotion analysis scenario presented in this paper. We achieve this goal by augmenting the fitting process of standard AAMs by including various types of constraints. A general overview of augmented AAMs is shown in Figure 2. It depicts the different components which contribute to the final system, whereas most parts are directly based on the given training data. An AAM trained on all landmarks of the training data forms the baseline of our augmented AAM ('full AAM training' in Figure 2). The fitting step of this AAM is then augmented using constraints derived from (1) a standard AAM trained only on the subset of stationary (i.e., torso and upper leg) landmarks, (2) local tracking methods for lower leg landmarks, (3) anatomical knowledge, and (4) the epipolar geometry of the scene.

In Subsection 3.1, we first derive a general framework for the inclusion of AAM fitting constraints. The remainder of this section gives a detailed description of the particular constraints used for the application on locomotion sequences. In Subsection 3.6, the necessary conditions of our approach and the generalization ability to other scenarios is discussed.

### 3.1 AAM fitting with constraints

For standard AAMs, it is not possible to include further knowledge - i.e., constraints - into the fitting process. We therefore reformulate AAM fitting within a *maximum a posteriori* (MAP) framework, which includes the approach of [37] as a special case. By definition, the MAP estimate  $\hat{\mathbf{b}}_{c,\text{MAP}}$  of the combined AAM parameter vector maximizes the posterior probability given the observations, in our case the input image  $\mathbf{I}$  and the fitting constraints  $\boldsymbol{\pi}$ , i.e.,

$$\hat{\mathbf{b}}_{c,\text{MAP}} = \underset{\mathbf{b}_c}{\operatorname{argmax}} p(\mathbf{b}_c|\mathbf{I}, \boldsymbol{\pi}). \quad (4)$$

By assuming *conditional independence* of the image data  $\mathbf{I}$  and the provided constraints  $\boldsymbol{\pi}$  given the parameter vector  $\mathbf{b}_c$ , we can rewrite Equation 4 as

$$\hat{\mathbf{b}}_{c,\text{MAP}} = \underset{\mathbf{b}_c}{\operatorname{argmax}} p(\mathbf{I}|\mathbf{b}_c) \cdot p(\boldsymbol{\pi}|\mathbf{b}_c) \cdot p(\mathbf{b}_c). \quad (5)$$

For the first likelihood term  $p(\mathbf{I}|\mathbf{b}_c)$ , not the whole input image  $\mathbf{I}$  is relevant, but only its sampled version  $\mathbf{g}_{\text{image}}$  which is based on the AAM shape configuration specified by  $\mathbf{b}_c$ , i.e.,  $p(\mathbf{I}|\mathbf{b}_c) = p(\mathbf{g}_{\text{image}}|\mathbf{b}_c)$ . As for standard AAMs, we assume the fitting process to be initialized at a parameter combination close to the optimal value.

The likelihood can then be modeled as a Gaussian distribution  $\mathbf{g}_{\text{image}}|\mathbf{b}_c \sim \mathcal{N}(\mathbf{g}_{\text{model}}, \Sigma_{\mathbf{g}_{\text{image}}-\mathbf{g}_{\text{model}}})$  or equivalently  $p(\mathbf{g}_{\text{image}}|\mathbf{b}_c) = p(\delta\mathbf{g})$  with  $\delta\mathbf{g} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\delta\mathbf{g}})$ . The covariance matrix  $\Sigma_{\delta\mathbf{g}}$  of the texture errors can be estimated in the training step of the AAM and is usually assumed to be diagonal due to its large dimensionality (cf. Subsection 2.2).

The likelihood term  $p(\boldsymbol{\pi}|\mathbf{b}_c)$  of Equation 5 integrates constraints into the fitting process. Here,  $\boldsymbol{\pi}$  is a vector which contains the differences between given target values (constraints) and the actual values based on the current AAM parameters  $\mathbf{b}_c$ . We again assume a Gaussian distribution, i.e.,  $\boldsymbol{\pi}|\mathbf{b}_c \sim \mathcal{N}(\mathbf{0}, \Sigma_{\boldsymbol{\pi}})$ . Concrete configurations of  $\boldsymbol{\pi}$  for different types of priors will be presented in the following subsections. Note that if multiple prior types are used, as is the case in our scenario, Equation 5 contains one likelihood term for each prior type.

The prior term  $p(\mathbf{b}_c)$  of Equation 5 can be modeled in various ways, e.g., using a uniform distribution (resulting in a *maximum likelihood* estimation) or a zero-mean Gaussian distribution [37]. To favor model configurations with a low complexity, in this work we prefer the latter method.

As a result of the above considerations, maximizing Equation 5 is equivalent to minimizing its *negative log likelihood*, thus

$$\hat{\mathbf{b}}_{c,\text{MAP}} = \underset{\mathbf{b}_c}{\operatorname{argmin}} \delta\mathbf{g}^\top \Sigma_{\delta\mathbf{g}}^{-1} \delta\mathbf{g} + \boldsymbol{\pi}^\top \Sigma_{\boldsymbol{\pi}}^{-1} \boldsymbol{\pi} + \mathbf{b}_c^\top \Sigma_{\mathbf{b}_c}^{-1} \mathbf{b}_c. \quad (6)$$

As mentioned above, Equation 6 can be optimized using arbitrary methods. One possible approach is based on the standard *additive* AAM parameter update scheme [25], which is derived in [37] and is used in this work. However, it is also possible to reformulate Equation 6 - i.e., AAM fitting with constraints - for the inverse composition/project-out approach [40], which in detail is described in [45].

### 3.2 Anchor AAM

The first type of constraints we use for fitting the full-body AAM are the results of an ‘anchor AAM’ or ‘subset AAM,’ which is an AAM applied on the subset of stationary landmarks, i.e., the torso and upper leg landmarks (cf. Figure 1). We include the results using the tracked landmark locations as positional constraints. Therefore,  $\boldsymbol{\pi}_{\text{anchor}}$  is the difference vector between target and current landmark positions. To estimate the reliability of the constrained positions and thus  $\Sigma_{\boldsymbol{\pi}_{\text{anchor}}}$ , robust confidence measures derived from the AAM fitting process (e.g., [46] or [47]) can be applied.

### 3.3 Robust local tracking constraints

While standard AAMs have problems with landmarks located at distal limb segments such as the lower legs,

the data-driven approach in [16] was specifically designed for tracking in X-ray sequences containing occlusions. In former studies, the method was proven to be well-suited for tracking the subset of lower leg landmarks of the side camera view, but it is inapplicable for landmarks with more severe occlusions such as the knee landmarks of the side view or feet landmarks of the top view. We include the tracking results for the lower leg landmarks as additional constraints  $\boldsymbol{\pi}_{\text{local}}$  into the augmented AAM. As for  $\boldsymbol{\pi}_{\text{anchor}}$ , the vector  $\boldsymbol{\pi}_{\text{local}}$  is the difference between target and current landmark positions. For the estimation of the corresponding covariance matrix  $\Sigma_{\boldsymbol{\pi}_{\text{local}}}$ , the same options as for the local detector used in [36] apply. In our case, due to the high accuracy of the local method [16], it is sufficient to use an isotropic covariance.

### 3.4 Anatomical constraints

For the challenging tracking scenario at hand, the inclusion of anatomical context knowledge is an important point to consider. As demonstrated in [16,48], one possibility is to perform a segmentation of the images into relevant anatomical parts - in our case, the *torso*, *left leg*, and *right leg*. For the side view of the bird locomotion scenario at hand, this segmentation can be obtained in three simple steps:

1. Global thresholding and contour finding  $\rightarrow$  *whole-body segment*
2. Iterative ellipse fitting on the whole body  $\rightarrow$  *torso segment*
3. Removing the torso segment from the whole-body segment  $\rightarrow$  *leg segments*

Here, the main problem is to find the correct correspondence between the two leg segments in the images and their anatomical counterparts. We propose to use the anchor AAM’s training data to train a regression model which can predict the correct correspondence for the entire sequence based on the AAM’s model parameters.

To include the results of the anatomical image segmentation into the fitting process, we define  $\boldsymbol{\pi}_{\text{anatomical}}$  to be the vector which for each landmark  $\mathbf{p}_m = (x_m, y_m)^\top$  contains the minimum Euclidean distance to its corresponding segment  $S(m)$ , i.e.,

$$\boldsymbol{\pi}_{\text{anatomical},m} = \min_{q \in S(m)} d(\mathbf{p}_m, \mathbf{q}). \quad (7)$$

To quickly obtain values for Equation 7 during the fitting process, we precompute distance transformed images for each segment using the algorithm presented in [49,50]. However, also, faster approximations for the distance transform such as [51] can be used, as small errors in the computed distances do not affect the overall result.

Because anatomical region constraints can only provide a coarse estimate for individual landmark positions,

for the covariance matrix  $\Sigma_{\pi_{\text{anatomical}}}$ , we assume a scaled identity matrix  $\sigma^2 \mathbf{I}$ , where  $\sigma^2$  is chosen to be substantially smaller than the covariances of other priors. This has the effect that the fitting process at first is completely driven by the anatomical constraints. When, as a result, each landmark  $\mathbf{l}_n$  is aligned to its corresponding anatomical segment  $S(m)$ , i.e.,  $\mathbf{p}_m \in S(m)$ , the vector  $\pi_{\text{anatomical}}$  becomes zero and the fitting procedure is governed by other constraints.

### 3.5 Epipolar priors

Although a camera calibration is not available for all datasets, it is still possible to include knowledge about the camera geometry into the fitting process. We can estimate the fundamental matrix  $F$  by exploiting the fact that point correspondences for the two camera views are available from the anchor AAM's training data. For each pair  $(\mathbf{v}_n, \mathbf{u}_n)$  of homogenous landmarks from the top and side view, we then add the additional constraint  $\pi_{\text{epipolar}}$  with

$$\pi_{\text{epipolar},n} = \mathbf{v}_n^\top F \mathbf{u}_n. \quad (8)$$

Equation 8 becomes zero if  $\mathbf{v}_n$  is located on the epipolar line  $F \mathbf{u}_n$  and vice versa. The covariance matrix  $\Sigma_{\pi_{\text{epipolar}}}$  can be estimated by applying the points used for the estimation of  $F$  on Equation 8.

### 3.6 Generalization to other scenarios

The presented method was specifically designed for the skeletal locomotion tracking scenario at hand. For this particular case, X-ray acquisition is a necessity, as all skeletal landmarks of interest must be observable. In addition, all parts of interest of an animal must remain in the field of view during the whole sequence, which generally implies the use of a treadmill. As the appearance of the animal is modeled using multi-view AAMs [43,44] (cf. Subsection 2.3), the camera setup must remain static during a recording. Similarly, if a trained model is to be reused for another sequence, the recordings must share an identical camera setup. However, as for standard multi-view AAMs, the number of cameras used for a sequence is flexible - in fact, the validation of our approach presented in Section 4 includes datasets with one camera view as well as datasets with two camera views.

More generally, the main characteristics of the data which led to our approach are non-stationary landmark movements and non-discriminative local texture information of certain landmarks (cf. Subsection 1.2). Therefore, the idea of *augmented active appearance models* should be applicable for all scenarios (1) in which landmarks and texture can be modeled by active appearance models, (2) which suffer from the data characteristics mentioned above, and (3) for which sufficient fitting constraints can be obtained. One possible example might be a medical

scenario, in which certain anatomical structures are to be tracked in an image sequence.

## 4 Experiments and results






The evaluation of our holistic approach for anatomical landmark tracking is performed on 32 real-world X-ray bird locomotion sequences. The datasets were recorded in the course of three large-scale zoological studies - namely [7], [6], and [34] - and comprise five species (quails, jackdaws, tinamous, bantams, and lapwings) which differ in morphology and locomotion characteristics. The acquisition of all sequences was carried out using a state-of-the-art biplanar high-speed X-ray system, based on the Neurostar<sup>®</sup> X-ray device (Siemens AG, Munich, Germany). All images have a resolution of  $1,536 \times 1,024$  pixels and were recorded at 1,000 frames per second. A total of 42,909 frames (approximately 125 GB of raw image data) was used in the course of this evaluation. Except for lapwings, all datasets have a biplanar camera setup and use the multi-view version of AAMs and augmented AAMs. Camera calibration allowing three-dimensional (3D) triangulation and evaluation of the tracking results is available for exactly one dataset. For each dataset, landmark positions manually located by human experts (biologists) are available, usually for every tenth frame of a sequence. Typical landmarks used for these datasets are depicted in Figure 1. A total of 175,942 ground-truth landmark positions were used for the comparisons presented in this paper. The actual number of ground-truth landmarks defined for each image varies per dataset and ranges from 14 to 24, with typical values being 20 landmarks per image. An overview of the employed datasets is shown in Figure 4.

We evaluate our approach based on the *point-to-point error* [52], i.e., the Euclidean error (in pixels for the 2D case and in millimeters for the 3D case) between manually located and automatically tracked landmark positions. For each sequence, an AAM was trained based on exactly one stride, using the provided landmark data. In any case, at most ten frames of a sequence were used for AAM training. Afterwards, all frames of the sequence were tracked using our presented augmented AAM approach.

### 4.1 Comparison to standard AAMs

As a proof of concept, we first compare our augmented AAMs to the results obtained by standard AAMs. For both methods, identical experimental setups were used - they only differ in the fitting method. The quantitative and qualitative comparisons for the real-world bird locomotion datasets are shown in Figure 5, grouped by camera view and bird species. For a better overview, landmarks are grouped into anatomical subsets: the *torso*



	SPECIES					TOTAL
	QUAIL ( <i>Coturnix coturnix</i> )	JACKDAW ( <i>Corvus monedula</i> )	TINAMOU ( <i>Eudromia elegans</i> )	BANTAM ( <i>Gallus gallus</i> )	LAPWING ( <i>Vanellus vanellus</i> )	
EXAMPLE IMAGES						5
INDIVIDUALS	6	3	2	1	1	13
SEQUENCES	15	8	6	1	2	32
FRAMES	20,952	11,563	6,070	796	3,528	42,909
GROUND-TRUTH LANDMARKS	71,597	24,683	70,002	2,160	7,500	175,942

**Figure 4 Dataset overview.** Overview of the 32 real-world bird locomotion datasets used for the evaluation of our approach. The datasets originate from three zoological studies [6,7,34], including birds of different morphology and locomotion characteristics.

(e.g., pelvis, furcula, and neck), *upper legs* (hip joints and knee joints), and *lower legs* (intertarsal joints and feet).

From the quantitative results presented in Figure 5a, it can be seen that augmented AAMs substantially outperform standard AAMs in terms of fitting accuracy in any case. This is particularly apparent for lower leg landmarks, where *median* errors of up to 150 pixels are constantly reduced to below 25 pixels for image sizes of  $1,536 \times 1,024$  pixels. As a typical example, for all 15 quail sequences, the median point-to-point error of lower leg landmarks of the side camera view is about 110 pixels for standard AAMs and only about 20 pixels for augmented AAMs. The reason for this result is that especially lower leg landmarks are prone to non-stationary shape movements and non-discriminative texture information, which drastically complicates standard AAM fitting but can be handled well by augmented AAMs. For other landmark groups, augmented AAMs are also clearly superior to their standard AAM counterparts: for the example of the 15 quail datasets, the median point-to-point error of torso landmarks of the top camera view is about 25 pixels for standard AAMs and about 15 pixels for augmented AAMs. The general performance disparity between the five bird species can be explained by different locomotion characteristics. For birds such as tinamous, the movement of the lower leg landmarks is less dominant compared to species such as jackdaws (*cf.* images in Figure 4).

In Figure 5b, qualitative tracking results for standard AAMs and augmented AAMs are presented for the lower leg landmarks of a jackdaw. It can be stated that the landmarks located by standard AAMs are clearly inaccurate in most cases, while augmented AAMs give reliable results. An example video showing tracking results of standard AAMs and augmented AAMs is provided in Additional file 1.

The above comparison clearly shows that our augmented AAM approach, as opposed to standard AAMs, is well suited for tracking the entire set of anatomical

landmarks in this challenging scenario. Based on a large-scale study which analyzes the accuracy of manually located landmarks in X-ray locomotion scenarios [34], it can be stated that the accuracy of our approach is comparable to the performance of human experts.

#### 4.2 Comparison to non-holistic approaches

While our augmented AAM approach is holistic in the sense that all landmarks are modeled in one consistent framework, it uses constraints obtained from methods which only perform well on very specific landmark subsets (*cf.* Section 3).

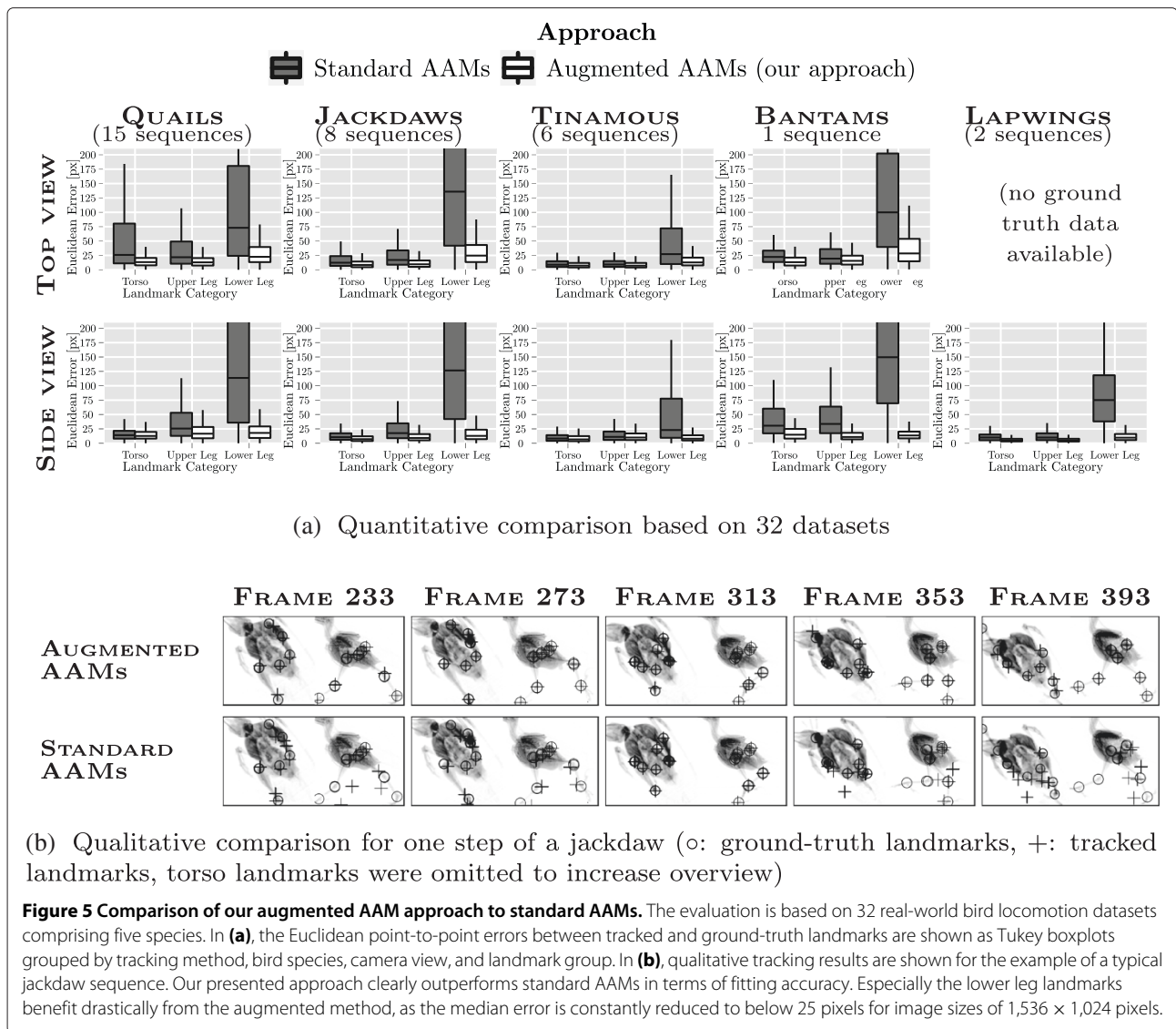
The question that we therefore would like to address is how an augmented AAM performs in direct comparison to each of the non-holistic approaches which provide its constraints. Quantitative results of this comparison are shown in Figure 6 for the two non-holistic tracking methods:

1. Subset AAM: standard multi-view AAM for the subset of torso and upper leg landmarks only [30]
2. Local tracking: robust local template tracking for lower leg landmarks of the side view only [16]

It is important to note that for both cases, the evaluation is performed only on the specific landmark subset of the respective non-holistic method.

As can be seen in the top row of Figure 6, the median error of the subset AAM is between 2 pixels (tinamou, top view) and 5 pixels (quails, side view) smaller than for corresponding landmarks of the augmented AAM. For the example of quails, the median error of the side view landmarks is about 10 pixels for subset AAMs, and about 15 pixels for augmented AAMs. This effect can be explained by the fact that the subset AAM is optimized for these specific landmarks, while the augmented AAM mediates between various fitting constraints for all landmarks - even those not covered in





**FRAME 233**   **FRAME 273**   **FRAME 313**   **FRAME 353**   **FRAME 393**

**AUGMENTED AAMs**

**STANDARD AAMs**

(b) Qualitative comparison for one step of a jackdaw (o: ground-truth landmarks, +: tracked landmarks, torso landmarks were omitted to increase overview)

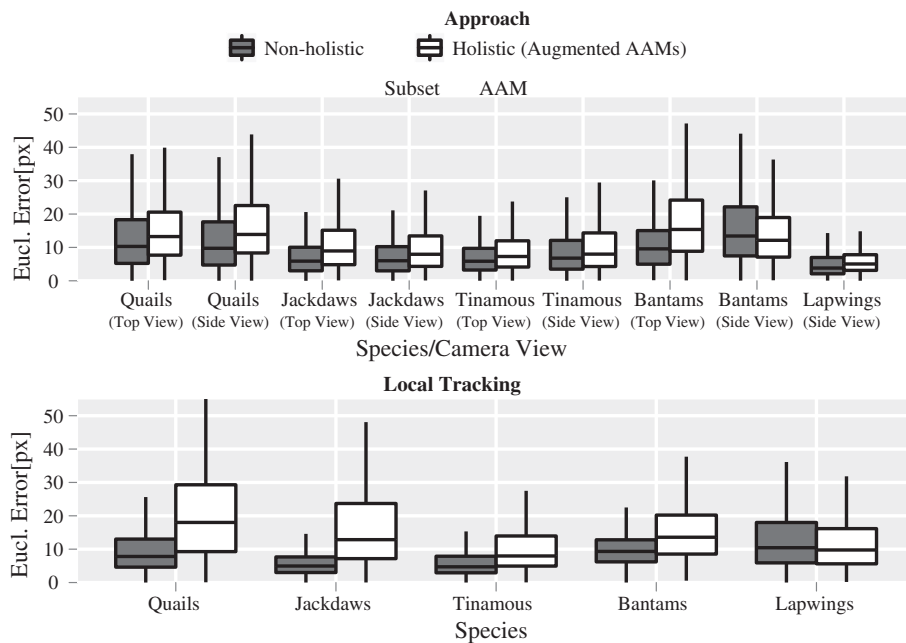
this comparison. In addition, the shape and texture models of the augmented AAM are more complex due to the increased scope and thus are harder to optimize. The results of the second non-holistic method, robust template tracking, are presented in the bottom row of Figure 6 and show the same tendency. While local tracking is even more accurate than the subset AAM, the performance of the augmented AAM is similar for both comparisons. Here, the very same explanations as before apply.

As a result, we can state that both non-holistic methods are more accurate on their specific landmark subsets than our holistic approach. However, the holistic approach has the essential advantage that it also can reliably and consistently track landmarks which are covered by neither of the non-holistic approaches, as for instance the lower leg landmarks of the top camera view (*cf.* Figure 5).

### 4.3 Influence of constraints

As our approach combines several fitting constraints, an important aspect is the practical relevance of individual constraint types. It is to be expected that positional constraints such as local tracking priors will have a larger benefit on fitting accuracy than, e.g., anatomical constraints. However, the question is whether a combination of several constraints can improve the fitting results. We therefore compare the performance of augmented AAMs using different combinations of constraints described in Section 3.

In Figure 7, quantitative results of this analysis are depicted. Due to the large amount of comparisons, results are exemplarily shown for jackdaws and tinamous, which according to Figure 5 have the worst and best tracking performance, respectively. It can be seen that torso and upper leg landmarks behave similarly for either case. Whenever

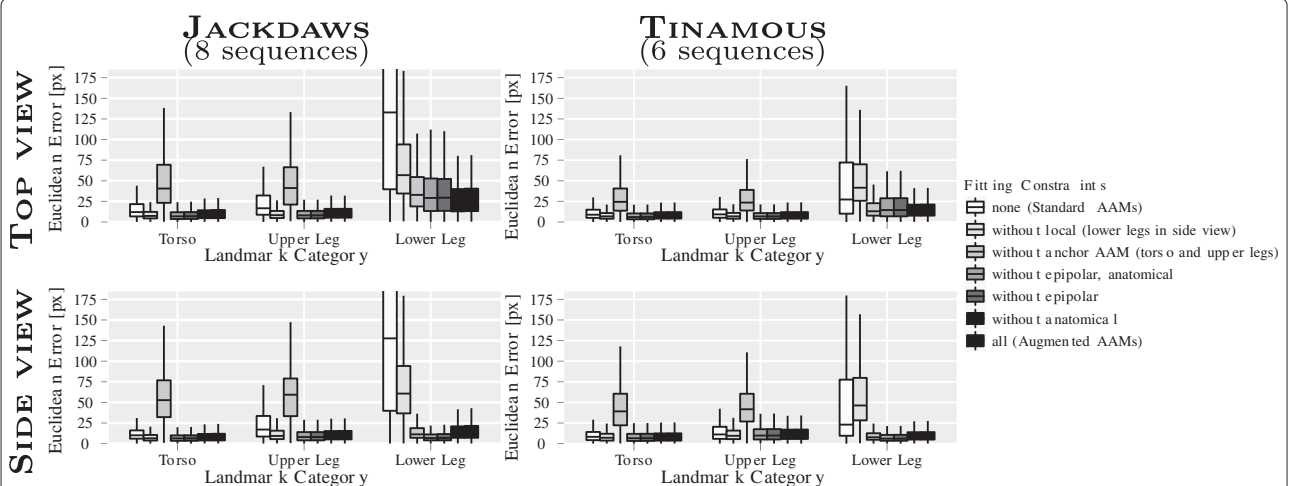


**Figure 6 Comparison of augmented AAMs to non-holistic methods.** Subset AAM corresponds to non-holistic method of standard AAMs applied on its subset of well-trackable landmarks (torso and upper leg, cf. Subsection 3.2), as presented in [17,30,34]. Local tracking denotes the non-holistic data-driven tracking approach of [16] applied on its subset of well-trackable landmarks (lower leg, cf. Subsection 3.3). On their specific landmark subsets, non-holistic methods are superior to augmented AAMs. However, augmented AAMs incorporate all landmarks in a consistent framework and also provide reliable positions for landmarks which are not covered by any of the non-holistic approaches.

constraints of the anchor AAM are provided for these landmarks, the holistic model seems to reach its maximum accuracy and no other constraints are beneficial.

Similarly, for lower leg landmarks, it is sufficient to use local template tracking constraints in easy scenarios (tinamous). However, in more challenging

scenarios (jackdaws), all constraints contribute to the final fitting performance. In both scenarios, epipolar constraints primarily improve the results of the top view. This is mainly due to the fact that the lower leg landmarks have no positional constraints for the top view and thus have a larger inaccuracy. While



**Figure 7 Influence of constraints for augmented AAM fitting.** The Euclidean point-to-point errors between tracked and ground-truth landmarks are shown as Tukey boxplots grouped by the type of used constraints for AAM fitting, bird species, camera view, and landmark group. The types of used constraints range from 'none' (standard AAMs) to 'all' (augmented AAMs). Results are exemplarily shown for a difficult bird species (jackdaws) and an easy scenario (tinamous). In challenging scenarios, all constraints contribute to the final fitting performance, as can be seen for the lower leg landmarks of jackdaws.

anatomical constraints do not increase accuracy when used together with local constraints, they improve results of standard AAMs in complicated scenarios (jackdaws). This fits their intended purpose of providing a rough initial landmark estimate for the other constraints (*cf.* Section 3).

An example in Figure 7 which demonstrates all aspects of the above argumentation is the case of lower leg landmarks in the top view for jackdaws. In case that no constraints are used ('none', standard AAMs), the median point-to-point error is larger than 125 pixels. If all but local tracking priors are employed ('without local'), a median error of about 55 pixels is obtained. When using all priors ('all', augmented AAMs), the median error is smallest with about 25 pixels.

#### 4.4 3D evaluation of tracking results

To allow an analysis of uncalibrated animal locomotion datasets recorded in previous biological studies such as [7], augmented AAMs do not rely on a calibrated camera setup, albeit both X-ray camera views are modeled in a consistent manner. However, for datasets having calibration information available, using 3D landmark positions instead of projected 2D positions is desired for biological evaluations (e.g., [8]). In the case of a known camera calibration, this can be achieved by triangulating the 2D tracking results of both X-ray camera views [53,54]. Similarly, we obtain 3D ground-truth landmarks by triangulating the given 2D ground-truth landmark locations. In the following, we evaluate the 3D accuracy of the landmarks tracked with our approach in order to

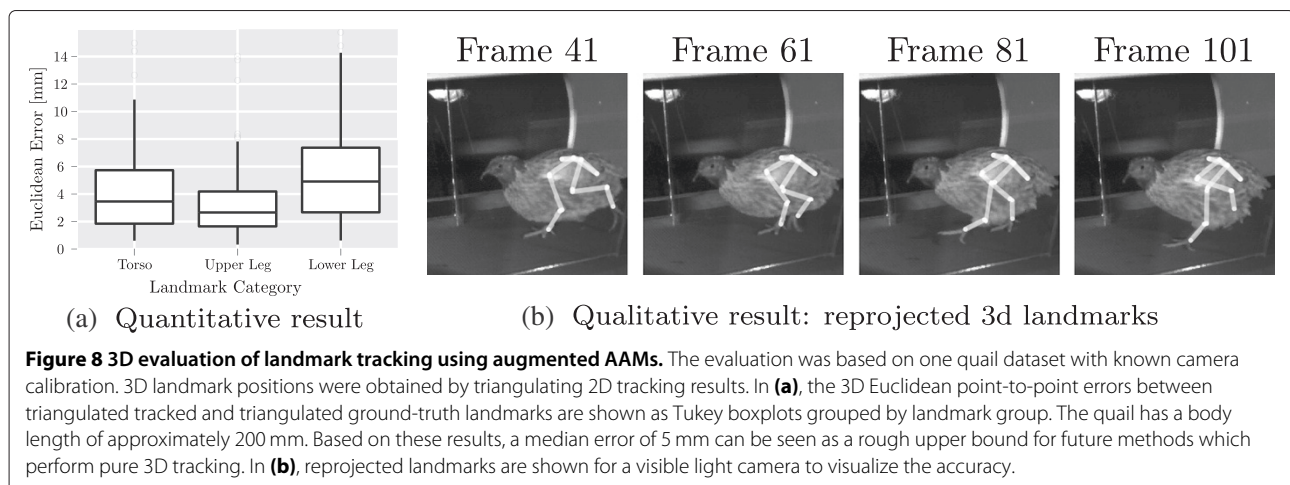
1. State whether our approach is accurate enough to produce reliable 3D results
2. Obtain an upper error bound for pure 3D tracking methods

Currently, camera calibration is only available for exactly one of the 32 datasets presented above - namely a quail dataset having 1,841 frames which cover 24 steps. For the calibration of this dataset, a custom-built metal plate with a size of 140 mm × 60 mm × 0.5 mm was employed. It contains 18 uniquely identifiable holes which are easily detectable in both X-ray as well as visible light cameras. For the actual calibration, we use the method of Zhang [55]. The mean backprojection error of the intrinsic camera calibration is 0.27 pixels at an image size of 1,536 × 1,024 pixels.

In Figure 8, both qualitative as well as quantitative results of the 3D evaluation are presented. From the quantitative results in Figure 8a, it can be seen that the median point-to-point error of all landmark types is below 5 mm. Compared to the animal's body length of 200 mm, this error is negligible for many practical biological evaluations. Additionally, this error serves as a rough upper bound for methods which perform pure 3D tracking. The largest median error (5 mm) is obtained for lower leg landmarks, which is in accordance with the results of the 2D evaluation (*cf.* Figure 5). The rather surprising result that upper leg landmarks have a slightly lower median error (2.7 mm) than torso landmarks (3.5 mm) is caused by 2D tracking inaccuracies in the top view of this particular dataset. To allow a visual assessment of the 3D accuracy, Figure 8b shows the reprojected landmark positions for one step of the animal which was additionally filmed with a visible light camera. A video showing these reprojected 3D landmarks is provided in Additional file 2.

#### 4.5 Implementation details

Both the augmented AAM approach presented in this work as well as the standard AAMs were entirely implemented in the programming language R (<http://www.r-project.org/>). The robust template tracking approach of [16] which is used to provide local AAM fitting



**Figure 8 3D evaluation of landmark tracking using augmented AAMs.** The evaluation was based on one quail dataset with known camera calibration. 3D landmark positions were obtained by triangulating 2D tracking results. In (a), the 3D Euclidean point-to-point errors between triangulated tracked and triangulated ground-truth landmarks are shown as Tukey boxplots grouped by landmark group. The quail has a body length of approximately 200 mm. Based on these results, a median error of 5 mm can be seen as a rough upper bound for future methods which perform pure 3D tracking. In (b), reprojected landmarks are shown for a visible light camera to visualize the accuracy.

constraints was implemented in C++ using the OpenCV library [56]. All experiments were performed on a typical desktop computer with an Intel® Core™ i5-760 CPU at 2.80 GHz. On average, the creation of all fitting constraints for the augmented AAM was performed at 13.2 frames per second (fps) for the anchor AAM, 11.4 fps for local tracking constraints, and 2.8 fps for torso and leg distance constraints (*cf.* Section 3). Given these constraints, our implementation of augmented AAMs runs at 0.5 fps. Note that this value could be drastically increased using a pure C/C++ implementation, employing the *inverse compositional/project-out* [40,45] optimization instead of the *additive method* [37] and by exploiting the vast parallelization capability of the approach. In the animal locomotion scenario at hand, however, a real-time processing of datasets is not of primary importance.

## 5 Conclusions and further work

In this paper, we presented augmented active appearance models, a general approach for AAM fitting in cases of non-stationary shape motions and non-discriminative local texture information. Our method is based on a holistic, probabilistic framework which allows the inclusion of arbitrary fitting priors. We applied our approach to the challenging scenario of landmark tracking in X-ray animal locomotion sequences, for which until now only methods for specific landmark subsets existed. For this particular scenario, we presented various types of suitable fitting constraints that were included into our probabilistic framework. Extensive experiments based on 32 real-world datasets including 175,942 ground-truth landmark positions showed that our approach clearly outperforms standard AAM fitting and allows to reliably track all landmarks of interest. In addition, we could show that the accuracy of our approach is sufficient to provide reliable 3D landmark estimates for calibrated datasets.

For further work, an interesting and relevant point to consider is the scenario of non-cyclic locomotion, for instance birds running over obstacles. Another important problem we want to solve is how to transfer already trained models to different tracking scenarios, such as adapting a quail model to be able to track tinamous. Both points mentioned require an adaptation of a given model to novel cases, and we plan to utilize methods from incremental learning [47] and domain adaptation for this task. Inspired by the promising results of 3D landmark estimation for calibrated datasets, another idea for further work is the inclusion of additional imaging modalities such as visible light cameras into the tracking process.

## Additional files

**Additional file 1: Video 1.** Qualitative tracking examples. This video shows some qualitative landmark tracking results for the scenario of grounded bird locomotion.

**Additional file 2: Video 2.** 3D landmark reconstruction examples. This video shows an example of 3D landmark estimation using augmented AAMs for the scenario of grounded bird locomotion.

## Competing interests

Both authors declare that they have no competing interests.

## Acknowledgements

The authors would like to thank Alexander Stöbel from the Department of Human Evolution at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany for providing the quail, jackdaw, and tinamou datasets. Furthermore, we would like to thank John Nyakatura from the Institute of Systematic Zoology and Evolutionary Biology with Phyletic Museum at the Friedrich Schiller University of Jena, Germany for providing the bantam and lapwing datasets as well as one additional quail dataset. This research was supported by grant DE 735/8-1 of the German Research Foundation (DFG).

Received: 31 January 2013 Accepted: 31 July 2013

Published: 12 August 2013

## References

1. SM Gatesy, Guinea fowl hind limb function, I: cineradiographic analysis and speed effects. *J. Morphol.* **240**(2), 1097–14687 (1999)
2. MS Fischer, N Schilling, M Schmidt, D Haarhaus, H Witte, Basic limb kinematics of small therian mammals. *J. Exp. Biol.* **205**(Pt 9), 1315–1338 (2002)
3. S Tashman, W Anderst, In vivo measurement of dynamic joint motion using high speed biplane radiography and CT: application to canine ACL deficiency. *J. Biomech. Eng.* **125**, 238–245 (2003)
4. EL Brainerd, DB Baier, SM Gatesy, TL Hedrick, KA Metzger, SL Gilbert, JJ Crisco, 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. SM Gatesy, DB Baier, FA Jenkins, KP Dial: Scientific rotoscoping: a morphology-based method of 3-D motion analysis and visualization. *J. Exp. Zool. A* **313A**(5), 244–261 (2010)
6. JA Nyakatura, E Andrada, N Grimm, H Weise, MS Fischer, Kinematics and center of mass mechanics during terrestrial locomotion of northern lapwings (*Vanellus vanellus*, Charadriiformes). *J. Exp. Zool. A* **317**(9), 580–594 (2012)
7. A Stöbel, MS Fischer, Comparative intralimb coordination in avian bipedal locomotion. *J. Exp. Biol.* **215**, 4055–4069 (2012)
8. MS Fischer, K Lilje, *Dogs in Motion* (VDH, Dortmund, 2011)
9. TL Hedrick, Software techniques for two- and three-dimensional kinematic measurements of biological and biomimetic systems. *Bioinspir. Biomim.* **3**(3), 034001 (2008)
10. BD Lucas, T Kanade, in *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81)*. An iterative image registration technique with an application to stereo vision (William Kaufmann, Vancouver, August 1981), pp. 674–679
11. J Shi, C Tomasi, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Good features to track (Seattle, 21–23 June 1994), pp. 593–600
12. S Baker, I Matthews, Lucas-Kanade 20 years on: a unifying framework. *Int. J. Comput. Vis.* **56**(3), 221–255 (2004)
13. GD Hager, PN Belhumeur, Efficient region tracking with parametric models of geometry and illumination. *IEEE T. Pattern Anal.* **20**(10), 1025–1039 (1998)
14. F Jurie, M Dhome, Hyperplane approximation for template matching. *IEEE T. Pattern Anal.* **24**(7), 996–1000 (2002)
15. DG Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vis.* **60**(2), 91–110 (2004)
16. M Amthor, D Haase, J Denzler, in *Proceedings of the Vision, Modeling and Visualization (VMV) Workshop*. Fast and robust landmark tracking in X-ray locomotion sequences containing severe occlusions (Magdeburg, 12–14 November 2012), pp. 15–22
17. D Haase, J Denzler, in *Image Analysis* ed. by A Heyden, F Kahl. Proceedings of the 17th Scandinavian Conference on Image Analysis (SCIA 2011), no.



- 6688 in LNCS, Ystad, May 2011. Anatomical landmark tracking for the analysis of animal locomotion in X-ray videos using active appearance models (Springer, Heidelberg, 2011), pp. 604–615
18. T Rohlfing, J Denzler, C Gräßl, DB Russakoff, CR Maurer Jr, Markerless real-time 3-D target region tracking by motion backprojection from projection images. *IEEE T. Med. Imaging* **24**(11), 1455–1468 (2005)
19. DL Miranda, JB Schwartz, AC Loomis, EL Brainerd, BC Fleming, JJ Crisco, Static and dynamic error of a biplanar videoradiography system using marker-based and markerless tracking techniques. *J. Biomech. Eng.* **133**(12), 121002 (2011)
20. T Ishikawa, I Matthews, S Baker. Efficient image alignment with outlier rejection. Technical Report CMU-RI-TR-02-27. Carnegie Mellon University Robotics Institute, 2002
21. F Jurie, M Dhome, in *Proceedings of the British Machine Vision Conference 2002, (BMVC)*. Real time robust template matching. (British Machine Vision Association, Cardiff, 2–5 September 2002)
22. J Pan, B Hu, JQ Zhang, Robust and accurate object tracking under various types of occlusions. *IEEE Trans. Circuits Syst. Video Techn.* **18**(2), 223–236 (2008)
23. TF Cootes, GJ Edwards, CJ Taylor, in *Computer Vision—ECCV'98*, ed. by H Burkhardt, B Neumann. Active appearance models. Proceedings of the 5th European Conference on Computer Vision, Freiburg, 2–6 June 1998. Lecture Notes in Computer Science, vol. 1407 (Springer, Berlin, 1998), pp. 484–498
24. Edwards G J, TF Cootes, CJ Taylor, in *Computer Vision—ECCV'98*, ed. by H Burkhardt, B Neumann. Face recognition using active appearance models. Proceedings of the 5th European Conference on Computer Vision, Freiburg, 2–6 June 1998. Lecture Notes in Computer Science, vol. 1407 (Springer, Berlin, 1998), pp. 581–595
25. TF Cootes, GJ Edwards, CJ Taylor, Active appearance models. *IEEE T. Pattern Anal.* **23**(6), 681–685 (2001)
26. AB Ashraf, S Lucey, JF Cohn, T Chen, Z Ambadar, KM Prkachin, PE Solomon, The painful face – pain expression recognition using active appearance models. *Im. Vis. Comp.* **27**(12), 1788–1796 (2009)
27. L van der Maaten, E Hendriks, Action unit classification using active appearance models and conditional random fields. *Cogn. Process* **13**(Suppl 2), S507–S518 (2012)
28. TF Cootes, CJ Taylor, in *Medical Imaging: Image Processing*, vol. 4322. Statistical models of appearance for medical image analysis and computer vision. ed. by M Sonka, KM Hanson. Proceedings of SPIE, Bellingham August 2001, (SPIE, Bellingham, 2001), pp. 236–248
29. SC Mitchell, BPF Lelieveldt, RJ van der Geest, JG Bosch, JHC Reiber, Sonka M, Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac MR images. *IEEE Trans. Med. Imaging* **20**(5), 415–423 (2001)
30. D Haase, JA Nyakatura, J Denzler, in *Pattern Recognition*. Multi-view active appearance models for the X-ray based analysis of avian bipedal locomotion. ed. by R Mester, M Felsberg. Proceedings of the 33rd DAGM Symposium (DAGM), no. 6835 in LNCS, Frankfurt, 31 August to 2 September 2011, (Springer, Berlin, 2011), pp. 11–20
31. S Das, N Vaswani, Nonstationary shape activities: dynamic models for landmark shape change and applications. *IEEE T. Pattern Anal.* **32**(4), 579–592 (2010)
32. N Vaswani, AKR Chowdhury, R Chellappa, "Shape activity": a continuous-state HMM for moving/deforming shapes with application to abnormal activity detection. *IEEE Trans. Image Process* **14**(10), 1603–1616 (2005)
33. TF Cootes, CJ Taylor, DH Cooper, J Graham, Active shape models—their training and application. *Comput Vis. Image Underst.* **61**, 38–59 (1995)
34. D Haase, J Denzler, in *Proceedings of the 8th Open German-Russian Workshop Pattern Recognition and Image Understanding (OGRW-8-2011)*. Comparative evaluation of human and active appearance model based tracking performance of anatomical landmarks in locomotion analysis, (Nizhny Novgorod, November 2011), pp. 96–99
35. D Cristinacce, TF Cootes, Automatic feature localisation with constrained local models. *Pattern Recognit.* **41**(10), 3054–3067 (2008)
36. P Martins, R Caseiro, JF Henriques, in *Proceedings of the 12th European Conference on Computer Vision*. Discriminative Bayesian active shape models (Florence, 7–13 October 2012), pp. 57–70
37. TF Cootes, CJ Taylor, in *IEEE International Conference on Computer Vision (ICCV)*. Constrained active appearance models (Vancouver BC, 7–14 July 2001), pp. 748–754
38. FL Bookstein, Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Med. Image Anal.* **1**(3), 225–243 (1997)
39. IL Dryden, KV Mardia, *Statistical Shape Analysis*. (Wiley, Chichester, 1998)
40. I Matthews, Baker S, Active appearance models revisited. *Int. J. Comput. Vis.* **60**(2), 135–164 (2004)
41. J Xiao, S Baker, I Matthews, T Kanade, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Real-time combined 2D+3D active appearance models (Washington D.C., June 2004), pp. 535–542
42. J Sung, D Kim, in *Third International Conference on Image Analysis and Recognition, Póvoa de Varzim, 18–20 September 2006, Lecture Notes in Computer Science vol. 4142*. Estimating 3D facial shape and motion from stereo image using active appearance models with stereo constraints (Springer Berlin, 2006), pp. 457–467
43. B Lelieveldt, M Üzümcü, R van der Geest, J Reiber, M Sonka, Multi-view active appearance models for consistent segmentation of multiple standard views. *Int. Congr. Ser.* **1256**, 1141–1146 (2003)
44. E Oost, G Koning, M Sonka, PV Oemrawsingh, JHC Reiber, BPF Lelieveldt, Automated contour detection in X-ray left ventricular angiograms using multiview active appearance models and dynamic programming. *IEEE T. Med. Imaging* **25**(9), 1158–1171 (2006)
45. G Papandreou, P Maragos, in *Conference on Computer Vision and Pattern Recognition (CVPR 2008)*. Adaptive and constrained algorithms for inverse compositional active appearance model fitting (Anchorage, 23–28 June 2008)
46. GJ Edwards, TF Cootes, CJ Taylor, in *IEEE International Conference on Computer Vision (ICCV) vol. 1*. Advances in active appearance models (Kerkyra, 20–27 September 1999), pp. 137–142
47. J Sung, D Kim, Adaptive active appearance model with incremental learning. *Pattern Recogn. Lett.* **30**(4), 359–367 (2009)
48. M Zhou, L Liang, J Sun, Y Wang, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. AAM based face tracking with temporal matching and face segmentation (San Francisco, 13–18 June 2010), pp. 701–708
49. PF Felzenszwalb, DP Huttenlocher, Distance transforms of sampled functions. *Theory Comput.* **8**(19), 415–428 (2012)
50. R van den Boomgaard, Mathematical morphology: extensions towards computer vision. PhD thesis. University of Amsterdam, 1992
51. G Borgefors, Distance transformations in digital images. *Comput. Vis., Graph., Image Process.* **34**(3), 344–371 (1986)
52. MB Stegmann, Active appearance models: theory, extensions and cases. Master's thesis. Technical University of Denmark, DTU, 2000
53. R Hartley, A Zisserman, *Multiple View Geometry in Computer Vision*, 2nd edn. (Cambridge University Press, Cambridge, 2003)
54. P Mittra, GN DeSouza, AC Kak, in *International Conference on Robotics and Automation (ICRA 2004)*. Calculating the 3D-pose of rigid-objects using active appearance models (New Orleans, 26 April to 1 May 2004), pp. 5147–5152
55. Z Zhang, A flexible new technique for camera calibration. *TPAMI* **22**(11), 1330–1334 (2000)
56. G Bradski, A Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library* (O'Reilly, Cambridge, 2008)

doi:10.1186/1687-5281-2013-45

**Cite this article as:** Haase and Denzler: 2D and 3D analysis of animal locomotion from biplanar X-ray videos using augmented active appearance models. *EURASIP Journal on Image and Video Processing* 2013 **2013**:45.