

Fast and Robust Landmark Tracking in X-ray Locomotion Sequences Containing Severe Occlusions

M. Amthor, D. Haase and J. Denzler

Computer Vision Group
Department of Mathematics and Computer Science
Friedrich Schiller University of Jena, Germany

Abstract

Recent advances in the understanding of animal locomotion have proven it to be a key element of many fields in biology, motion science, and robotics. For the analysis of walking animals, high-speed x-ray videography is employed. For a biological evaluation of these x-ray sequences, anatomical landmarks have to be located in each frame. However, due to the motion of the animals, severe occlusions complicate this task and standard tracking methods can not be applied. We present a robust tracking approach which is based on the idea of dividing a template into sub-templates to overcome occlusions. The difference to other sub-template approaches is that we allow soft decisions for the fusion of the single hypotheses, which greatly benefits tracking stability. Also, we show how anatomical knowledge can be included into the tracking process to further improve the performance. Experiments on real datasets show that our method achieves results superior to those of existing robust approaches.

Categories and Subject Descriptors (according to ACM CCS): I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Tracking

1. Introduction

The in-depth understanding of animal locomotion is a key element and an ongoing field of research in biology, motion science and robotics [BBG*10]. Recently, this fact was impressively demonstrated when it was shown that all dogs—no matter what size or breed—have a nearly identical locomotion [FL11]. Not only has this study led to new insights, but it has also overthrown large parts of what was previously known about dog locomotion. Another interesting class of animals for which little is known about grounded locomotion are birds. Due to the countless variations in size, body mass, and walking speed, birds constitute an ideal field of investigation. To analyze the locomotion of birds, only x-ray recordings allow for accurate and unbiased *in vivo* insights into all parts of the locomotor system. In practice, this is realized by placing a bird on a treadmill and using a special x-ray system (C-arm) to record the bird during locomotion. In order to cover every important detail of the movement, these recordings are usually carried out at 1000 frames per second. For typical recording times of one to three seconds per sequence, this leads to huge amounts of data for a large scale

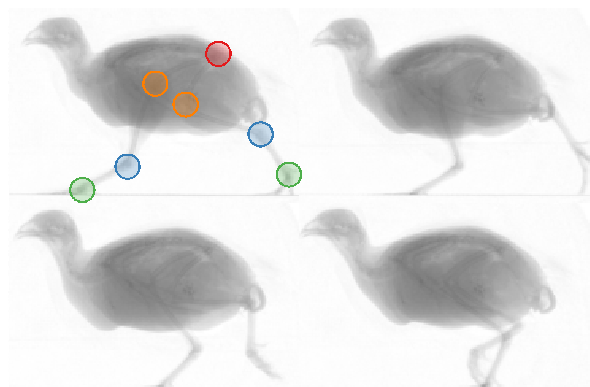


Figure 1: Example images of an x-ray locomotion sequence of a quail. The images show various states of locomotion and occlusions. Typical landmarks of the locomotor system are marked with colored circles (red: hip joints, orange: knee joints, blue: proximal tarsometatarsi, green: feet).

study. The biological evaluation of the recordings is mainly based on finding anatomical points of interest, *i.e.* landmarks

NAME	DEFINITION	RANGE
Sum of Squared Differences (SSD)	$R_{SSD}(\mathbf{I}_n, \mathbf{T})[x, y] = -\sum_{x', y'} (\mathbf{I}_n[x + x', y + y'] - \mathbf{T}[x', y'])^2$	$[-\infty, 0]$
Cross-Covariance (CCOV)	$R_{CCOV}(\mathbf{I}_n, \mathbf{T})[x, y] = \sum_{x', y'} \mathbf{I}_n^{(0)}[x + x', y + y'] \cdot \mathbf{T}^{(0)}[x', y']$	$[-\infty, +\infty]$
Cross-Correlation (CCORR)	$R_{CCORR}(\mathbf{I}_n, \mathbf{T})[x, y] = \frac{\sum_{x', y'} \mathbf{I}_n^{(0)}[x + x', y + y'] \cdot \mathbf{T}^{(0)}[x', y']}{\sqrt{\sum_{x', y'} (\mathbf{I}_n^{(0)}[x + x', y + y'])^2 \cdot \sum_{x', y'} (\mathbf{T}^{(0)}[x', y'])^2}}$	$[-1, 1]$

Table 1: Overview of prominent score functions used for template matching. $\mathbf{I}_n^{(0)}$ and $\mathbf{T}^{(0)}$ are the mean centered versions of the search image \mathbf{I}_n and the template image \mathbf{T} as defined in Subsect. 3.1.

of the locomotor system in every image of a sequence. An example image of a bird sequence including typical landmarks is shown in Fig. 1. Until now, the location of the landmarks in the sequences is carried out by human experts. However, due to the sheer magnitude of the data, this is a tremendously time-consuming task. Even though thousands of videos have been recorded to date, the bottleneck of manual labeling has prevented the evaluation of large amounts of data. Therefore, there is a urgent need to automate the task of landmark tracking in x-ray locomotion sequences. Unfortunately, this is a difficult task due to several characteristics of the data. As can be seen in Fig. 1, the contrast in the images is very low. However, the biggest problem is the presence of occlusions of various body parts resulting from the movement of the bird and the x-ray acquisition. As a consequence, many prominent tracking algorithms fail in this setting.

The goal of this paper is to present a new robust tracking method which can deal with the challenges occurring in the data and substantially reduce the human effort spent in landmark labeling. In Sect. 2, related work and recent approaches are discussed and the motivation of our own approach is presented. Our robust tracking algorithm is described in Sect. 3 in a general manner. Afterwards, the application to x-ray locomotion data is examined in Sect. 4. We then compare the performance of our algorithm with respect to several other robust state-of-the-art tracking approaches based on a variety of real-life datasets in Sect. 5.

2. Related Work and Motivation

As landmark and object tracking is a crucial part of computer vision, numerous approaches and variations have been developed over time. A comprehensive overview of state-of-the-art methods can be found in the survey of [YJS06]. Generally, we can distinguish between appearance based and feature based tracking approaches. For the former, typical examples are template matching, optical flow/KLT tracking [LK81, TS94], and region tracking [HB98, JD02a]. A performance evaluation of these methods is given in [DGBD05]. One example for feature based approaches is SIFT keypoint

tracking [Low04]. The main drawback of all aforementioned methods is the lack of robustness with respect to severe occlusions in the image data.

For the template matching approach which is also pursued in this paper, various adaptations and extensions exist. To name but a few, these include drift-free template updates [MIB04, Sch07] and the dynamic template extension [HIN10]. Most importantly, however, several approaches exist for occlusion handling. In [NWvdB01], a Kalman filter framework [WB95] is used to model the template appearance and short-term occlusions on a per-pixel basis. A similar approach is presented in [PHZ08], where regions of the template are classified as occluded and non-occluded. This is closely related to the idea of dividing a template into sub-templates and neglecting occluded parts from the estimation of the new position as proposed in [IMB02, JD02b].

For the fusion of the results of the sub-templates, [IMB02] assume a fixed amount of them to be non-occluded and therefore only regard the p best matching percent of sub-templates. While this approach works well in many cases, it does neither allow full-occlusions nor does it use all available information in case of no occlusions. The sub-template fusion employed in [JD02b] is identical to the method presented in [Jur99]. Here, each sub-template supports one hypothesis, and all hypotheses are combined using a Gaussian uncertainty model in the template pose space. The drawback with both approaches is that each sub-template can only support one hypothesis and thus makes a hard decision. However, especially for x-ray occlusions, a part of the original image structure may still be observable. This fact is the main motivation for our approach, in which we aim to allow soft decisions for the fusion of sub-template hypotheses.

In the field of anatomical landmark tracking in x-ray locomotion sequences, active appearance models (AAMs) [CET01] have proven to be well-suited for tracking certain landmarks [HD11a]. For the case of bird locomotion, “inner-torso” landmarks such as hip and the knee joints can be tracked reliably. However, due to the nature of the data and the AAMs, this approach is not suitable for landmarks that

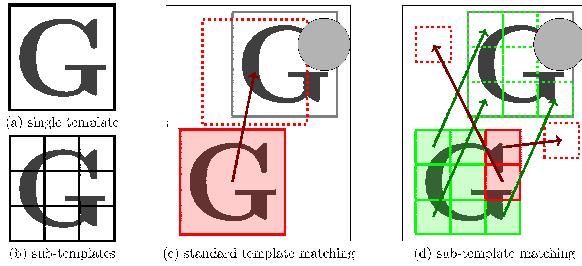


Figure 2: Comparison of standard and sub-template based template matching under occlusion. When using template matching based on a single template (a), occlusions lead to inaccurate results (c). By dividing the template into sub-templates (b), the correct position can still be obtained based on the non-occluded parts (d).

have a relatively large distance to the torso, such as the outer leg landmarks. Therefore, we mainly focus on “outer-torso” landmarks such as legs and feet in this work. Note, however, that our method is not limited to this kind of landmarks.

3. Landmark Tracking under Occlusions

This section describes our approach to robust landmark tracking in a general manner. After a short overview, various adaptations with respect to standard tracking methods will be presented. The application of our algorithm for landmark tracking in x-ray locomotion sequences is described in detail in Sect. 4.

The main idea of the tracking approach proposed in this paper is to develop an improved version of template tracking of a region around a landmark. Motivated by the application scenario of anatomical landmark tracking in x-ray locomotion sequences, our goal is to obtain a method which has the properties of

- invariance against **rotation and scaling**,
- robustness against partial and full **occlusions**,
- optional inclusion of **global context**, and
- fast computation times, allowing the evaluation of **large amounts of data**.

Each of the topics mentioned above will be described in detail in the following subsections.

3.1. Standard Template Tracking

Given an image sequence $I_1, \dots, I_N \in \mathcal{I}$ and a template image $T \in \mathcal{T}$, the general goal of *template tracking* is to find the location of T in every frame of the sequence [YJS06]. This matching is based on a particular score function $R: \mathcal{I} \times \mathcal{T} \rightarrow \mathcal{I}$, which maps each pixel of a search image I_n to a scalar indicating the similarity of I_n with the template T located at this particular pixel. The optimal position (\hat{x}_n, \hat{y}_n) of T in the image I_n is then obtained by using

the position providing the best match, *i.e.*

$$(\hat{x}_n, \hat{y}_n) = \operatorname{argmax}_{x,y} R(I_n, T)[x, y]. \quad (1)$$

Prominent score functions are the *sum of squared differences* (SSD), the *cross-covariance* (CCOV), and the *cross-correlation* (CCOR) [BK08, RN88]. Exact definitions of R_{SSD} , R_{CCOV} , and R_{CCOR} are listed in Table 1. Here, $I_n^{(0)}$ and $T^{(0)}$ are the mean centered versions of I_n and T with

$$I_n^{(0)}[x, y] = I_n[x, y] - \frac{1}{|T|} \sum_{x', y'} I[x + x', y + y'] \quad (2)$$

and

$$T^{(0)}[x, y] = T[x, y] - \frac{1}{|T|} \sum_{x', y'} T[x', y'], \quad (3)$$

with $|T|$ being the number of pixels in T .

While SSD is widely used for gradient descent based approaches such as [HB98], [JD02a], and [MIB04], cross-covariance and cross-correlation generally tend to give better results because of their invariance to additive and—in the case of correlation—multiplicative influences.

For the template tracking approach described above, invariance with respect to rotation or scaling can easily be achieved [KdA07]. Instead of matching only the original template T against the search image I_n , rotated versions $\tilde{T}_{\theta_1}, \dots, \tilde{T}_{\theta_M}$ of T with a sufficiently accurate angular resolution are used. Similar to Eq. 1, the optimal template location (\hat{x}_n, \hat{y}_n) and rotation $\hat{\theta}$ is then given by

$$(\hat{x}_n, \hat{y}_n, \hat{\theta}) = \operatorname{argmax}_{x,y,\theta} R(I_n, \tilde{T}_\theta)[x, y]. \quad (4)$$

Above extension can be directly modified to also include scale invariance and shall therefore not be considered separately at this point. Note that for the sake of simplicity, in all following discussions we will use the notation of pure translation—even for cases of template matching with rotation and scaling.

3.2. Handling Partial and Full Occlusions

One major problem of standard template tracking is its failure in the case of occlusions [JD02b, IMB02]. As can be seen in Eq. 1 and the definitions of the score functions in Table 1, even partially corrupted image data I_n can tremendously bias the values of $R(I_n, T)$ and therefore change the matching results for the worse.

In the case of partial occlusions, however, certain parts of the image still allow a correct estimation of the real template position. Therefore, by considering parts of the original template independently of each other as suggested in [JD02b] and [IMB02], partial occlusions can be resolved. This idea is schematically shown in Fig. 2, where standard template matching Fig. 2(a, b) and sub-template based matching Fig. 2(b, d) are compared for the case of partial

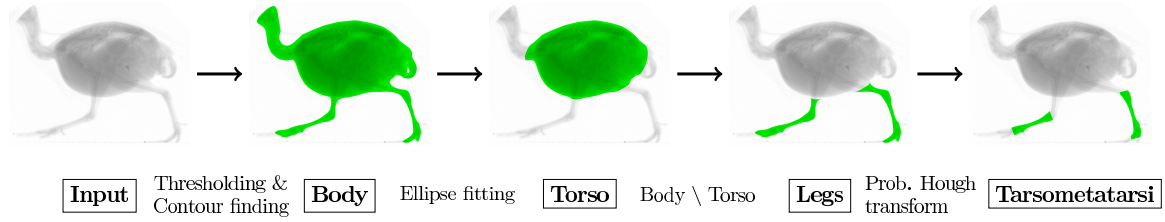


Figure 3: Schematic process of object mask determination for various types of anatomical landmarks in x-ray locomotion data. Anatomically motivated object masks can be obtained in a fast and robust way based on four basic steps. The masks obtained after each step are shown in green color. Especially for the landmarks located on the joints of the tarsometatarsi, the usage of such masks can lead to a vastly improved tracking quality.

occlusions. Where our approach differs from [JD02b] and [IMB02] is the way in which the final estimation for the template movement is obtained from the sub-template results. While in [JD02b] and [IMB02] each sub-template votes for only one final template transition and thus makes a hard decision, we aim to allow a soft decision for each sub-template. Our approach is motivated by three main observations:

- The matching results $R(I_n, T)$ of all non-occluded sub-templates have a peak at the correct position.
- Even for partially occluded sub-templates, there is usually a local peak around the correct position.
- Fully occluded sub-templates give somewhat random results, *i.e.* each occlusion leads to peaks at different positions.

As a consequence, averaging over the scores of all sub-templates has the effect of removing peaks with small occurrences and preserving peaks with high support. Note that with this method it is possible for one sub-template to vote for multiple peaks—in fact, each sub-template gives a rating for *every possible template transition*, weighted by its actual matching quality.

If our original template T is split into K sub-templates S_1, \dots, S_K , the final template position for our proposed method is obtained by finding the best match in the averaged score \bar{R} , *i.e.*

$$(\hat{x}_n, \hat{y}_n) = \operatorname{argmax}_{x,y} \bar{R}[x,y] \quad (5)$$

with

$$\bar{R} = \frac{1}{K} \sum_{k=1}^K R(I_n, S_k). \quad (6)$$

In the case of a full occlusion, *i.e.* an occlusion of all K sub-templates, it is not possible to obtain an accurate estimation based on image information. One approach to deal with the case of short-term full occlusions is to limit the velocity of template movement to prevent the template from drifting away too quickly [YJS06]. However, this limitation restricts the method to small template movements and fast objects will be lost. Another idea is to apply a linear motion model to maintain the current movement when a full

occlusion occurs [PHZ08]. Because we do not make a hard decision between occluded and non-occluded templates, this method is not applicable here. To avoid the drawbacks stated above, in our approach we employ a Kalman filter [WB95] framework and combine both techniques. The Kalman filter is used to estimate and predict the current state of motion and template position. After each prediction step, however, the actual measurement is limited to a small region around the prediction. While this approach enables template motions at arbitrary speed, the *difference* of template speed, *i.e.* the template acceleration is limited to prevent large drifts under full occlusions.

3.3. Usage and Refinement of Object Masks

The template matching approach in its initial form makes use of a rectangle to cover the landmark or object of interest. Therefore, the template usually contains background or other inappropriate parts. To improve the performance of our method we adjust the template to the actual object which is to be tracked. This can easily be done by laying out all initial sub-templates S_k based on a given region of interest (ROI) describing the object. Hence, the amount of neighboring objects and background can be substantially reduced and thus leads to an improved robustness. In Subsect. 4.2 it is described how anatomical object masks can be easily computed for the case of x-ray locomotion sequences.

Additionally, the set of sub-templates can be refined to only contain parts that can be tracked reliably, *e.g.* by using methods such as [TS94].

3.4. Performance Aspects

The naive implementation of the template matching algorithm described above leads to unnecessary slow computation times, as described in [KdA07]. A large performance gain can be achieved by using the convolution theorem. Also note that for template matching with rotation and scale invariance, the transformed—possibly rotated and scaled—sub-patches can be precomputed. With this technique, only the Fourier transform of the image has to be computed for each frame to be tracked.







DATASETS						
	Q1	Q2	Q3	B1	L1	L2
						
SPECIES	Quail	Quail	Quail	Bantam	Lapwing	Lapwing
FRAMES (LABELED)	2245 (91)	1841 (32)	1372 (81)	1024 (78)	1551 (236)	2301 (264)
STEPS (LABELED)	11 (11)	23 (4)	9 (9)	6 (5)	12 (2)	16 (2)

Table 2: Overview of the six x-ray bird locomotion datasets used for the experimental evaluations.

4. Application to X-ray Locomotion Sequences

The tracking approach presented in Sect. 3 is described in a generic manner and is applicable for a wide variety of tracking tasks and conditions. While the general appliance of the proposed algorithm to a certain scenario is straightforward, specific aspects should be considered to match the needs of the underlying data. Therefore, this section discusses the characteristics of the x-ray locomotion sequences used in this paper and describes according adaptations.

4.1. General Aspects

Due to the constrained movement on a treadmill and the specific camera setup, one very beneficial characteristic of the x-ray sequences at hand is the fact that distance and orientation of the animal with respect to the camera is approximately constant. As one consequence, scale effects can be neglected in this scenario and only template translation and rotation are of interest. Furthermore, as all landmarks to be tracked are anatomical keypoints, the appearance of a small region around a landmark does not change over time if occlusions are disregarded. Therefore, unlike applications presented in [MIB04] or [PHZ08], in our case there is no need for updating template appearances while tracking.

4.2. Obtaining Object Masks

As described in Subsect. 3.3, our approach is capable of using object masks, *i.e.* arbitrarily shaped regions used for tracking. While for many standard tracking scenarios a simple rectangular region around a landmark is sufficient, in the present application we can include anatomical knowledge into the selection of object masks.

Good examples for cases in which such additional knowledge is advantageous are the landmarks located at the distal and proximal ends of the *tarsometatarsi* (see Fig. 3, right image). A simple rectangular template around any of these landmarks includes parts of two bones and their connecting joint. As the angle between these bones varies throughout one sequence, this is only a suboptimal choice. Taking anatomical knowledge into account, a better solution for the

landmarks named above is to use the *tarsometatarsi* as object masks. As will be shown in the experiments in Sect. 5, this data-specific adaption can lead to vastly improved tracking results.

To obtain anatomical masks for an image taken from an x-ray locomotion sequence, we employ the following scheme:

1. **Body mask:** apply global threshold to image and find largest contour [SA85].
2. **Torso mask:** fit ellipse on the body mask and intersect it with the body mask.
3. **Leg mask:** subtract the torso mask from the body mask and find contours below the centroid of the torso mask.
4. **Tarsometatarsus mask:** apply progressive probabilistic Hough transform [MGK00] on the leg mask.

Due to the homogeneity of the underlying x-ray images, the usage of these relatively simple steps allows for a fully automatic fast and robust mask determination. An example application of this algorithm to a typical x-ray locomotion image is shown in Fig. 3. The masks obtained after each step are shown in green color.

5. Experiments and Results

To analyze the performance of our proposed landmark tracking approach, numerous experiments on real-world datasets were conducted. A total of six x-ray bird locomotion sequences are employed, of which three show quails, one shows a bantam chicken, and two show lapwings. Each dataset was recorded at a frame rate of 1000 Hz and a resolution of 1536×1024 pixels. The length of the sequences varies between 1024 and 2245 frames at a total of 10,334 frames covering 77 steps. Ground-truth landmark positions provided by human experts (zoologists) are available for 782 frames. An overview of the used datasets is shown in Table 2. As pointed out in Sect. 2, tracking approaches for inner-torso landmarks exist already. For this reason, we focus mainly on outer-torso landmarks (*cf.* Fig. 1) in the following experiments.

In our experiments, we wish to analyze the following questions:

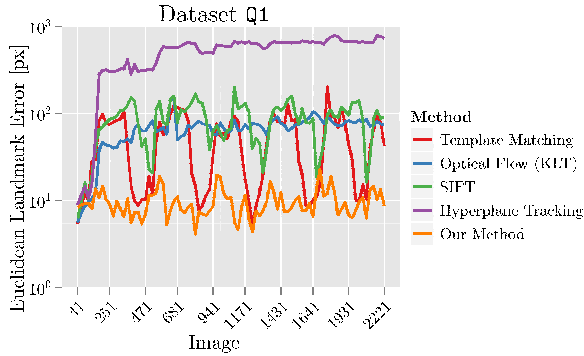


Figure 4: Comparison of standard tracking approaches to our proposed method for x-ray locomotion data. The lines show the median Euclidean landmark distances to ground-truth data provided by human experts. Due to the numerous occlusions, the standard approaches quickly drift away, while our robust approach succeeds in tracking the landmarks over the entire sequence.

1. Is our approach capable of overcoming difficulties in the data which can not be dealt with using **standard approaches**?
2. How does our proposed template motion estimation perform compared to other state-of-the-art approaches for **sub-template fusion**?
3. Is the *constrained* Kalman filter based **motion model** better than other proposed ways of full-occlusion handling?
4. Is it beneficial to use customized **object masks** rather than a simple rectangular region around a landmark?
5. Do the **computation times** allow for an evaluation of large amounts of data?

If not stated otherwise, for the experiments we used tarsometatarsus object masks as described in Subsect. 4.2. The sub-templates were chosen to be circular with a radius of 15 pixels. The distance between the centers of two sub-templates was set to be 8 pixels.

5.1. Comparison to Standard Tracking Approaches

To verify that the occlusions occurring in the x-ray locomotion datasets are too severe to use standard tracking approaches, we tested several prominent methods and compared the results to our final approach. The employed methods are standard template matching, KLT [LK81, TS94], SIFT [Low04] and hyperplane region tracking [JD02a].

The results of the experiments are shown in Fig. 4. Note that as the results are very similar for all six datasets, only those of the Q1 dataset are presented here in detail. The figure shows the median Euclidean errors of the various approaches with respect to the ground-truth labelings. It can be stated that all standard tracking approaches fail to track most landmarks and drift away quickly, even for relatively few occlusions. Our approach, on the contrary, is able to overcome the severe occlusions and the low contrast in the data and

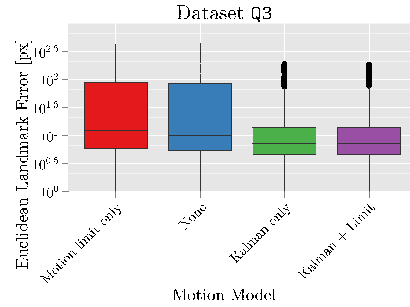


Figure 6: Influence of the motion model on the tracking robustness. Each boxplot shows the Euclidean landmark errors made over the entire sequence. The usage of a Kalman filter has the biggest influence on the tracking performance.

succeeds in tracking the landmarks over the entire sequence at an average error of 10 pixels. Taking into account the image size of 1536×1024 pixels, this is an excellent result and comparable to human tracking results [HD11b].

5.2. Obtaining Template Motion from Sub-Templates

As stated in Subsect. 3.2, there exist several ways to estimate the final template transformation based on all sub-template hypotheses of one tracking step. As opposed to a fusion based on hard decisions [JD02b, IMB02], we introduced a new method allowing multiple hypotheses per sub-template. In this subsection, our proposed method is compared to the existing fusion methods of [JD02b] and [IMB02]. For the experiments, our final tracking algorithm was used and only the fusion method was varied to ensure a fair comparison.

In Fig. 5, the according results for all six datasets are shown. First of all it can be stated that the method of [JD02b] has the worst performance in this scenario, as for all datasets the tracked landmarks quickly drift away. The method of [IMB02] is more stable than [JD02b], as for four out of six sequences no landmark is lost. Our approach, however, achieves the best performance, as all landmarks are tracked correctly for each dataset. Note that for the cases in which [IMB02] does not drift away, the results are comparable to our approach. This leads to the conclusion that for x-ray sequences our approach is as accurate as the one of [IMB02] while being more robust to occlusions.

5.3. Influence of Motion Model

Another adaption we suggested was the usage of a Kalman filter with constrained acceleration to tolerate short-term full-occlusions. To determine the performance of this approach, we compare it to three other methods of motion/full-occlusion modeling. Namely, those methods are no motion model at all, motion vector truncation [YJS06], and a standard Kalman filter [PHZ08].

The results of the experiments are presented in Fig. 6.

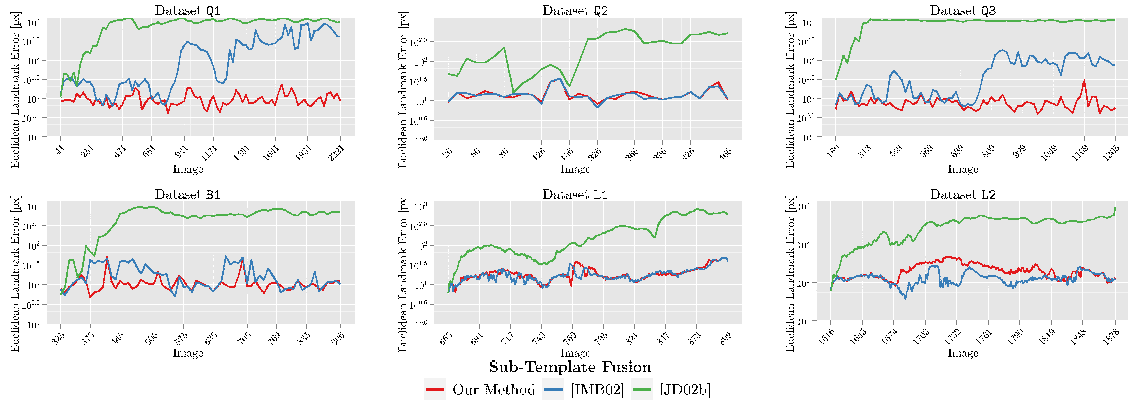


Figure 5: Comparison of fusion methods for sub-template based tracking approaches. Each line shows the median Euclidean error of the outer leg landmarks for a particular method. The method of [JD02b] (green) has the worst results. Our approach (red) tends to be slightly more robust than the method used in [IMB02] and has the best tracking performance.

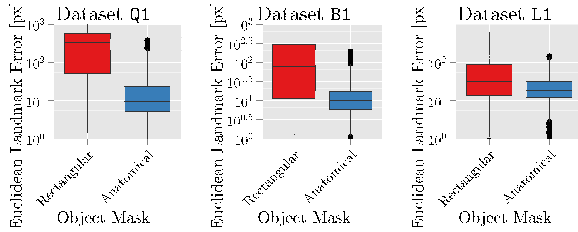


Figure 7: Influence of the usage of anatomical object masks instead of rectangles. The results show that specific anatomical masks drastically improve the tracking results.

The most striking result is that the usage of a Kalman filter vastly improves the tracking performance. Truncating the motion vector as suggested in [YJS06], on the other hand, seems to have no substantial influence on accuracy and stability. While the first observation is fairly reasonable, the latter is a bit surprising. This suggests that the occurring full-occlusions in the data are short enough to cause no observable disturbances. However, as limiting the motion vector does not worsen the tracking performance, it is possible to use it together with the Kalman filter framework for the case of longer full-occlusions.

5.4. Usage of Anatomical Object Masks

In Subsect. 4.2 it was shown how anatomical object masks can be obtained for x-ray locomotion landmarks in an easy and robust manner. To answer the question whether the usage of such specific masks rather than a rectangular region around a landmark is able to improve the tracking performance, we tested both cases using our algorithm. Note, however, that also for the case of rectangular regions, homogeneous sub-templates were removed as suggested in Subsect. 3.3.

In Fig. 7, the results are exemplarily shown for the three datasets Q1, B1, and L1. It can be seen that in any case the usage of anatomical object masks substantially improves the performance. While for the Q1 dataset the median error for

the case of simple rectangular tracking regions is approximately 300 pixels, the identical setup with object masks leads to a 10 pixel median error. Note, however, how the difference between rectangular and anatomical masks decreases as the leg length of the analyzed bird increases (see Table 2 for an overview of the datasets). This interesting effect can be explained by the fact that for short legged birds such as quails, occlusions with parts of the torso and the feet are more likely and will cover a larger relative area. Summing up, we can verify our assumption that specific anatomical masks drastically improve the tracking results in the scenario at hand and can state that they are an important part of our tracking approach.

5.5. Computation Times

Our algorithm was solely implemented in C/C++ using the OpenCV library [BK08]. As suggested in Subsect. 3.4, pre-computation steps and Fourier transform were used to speed up the most time-consuming steps of our algorithm. The experiments were carried out on a standard desktop computer with an Intel® Core™ i7-2600 CPU @3.40GHz. The average tracking speed achieved for the presented experiments was between 7.51 fps (object masks implying many sub-patches) and 13.73 fps (object masks implying few sub-patches) per object mask to be tracked. Note that the runtime of our algorithm heavily depends on the used number and size of the sub-templates. As for the x-ray locomotion scenario tracking quality is more important than real-time tracking, the parameters were selected very defensively. However, for applications requiring real-time processing, the parameters can easily be adopted to allow 30 or more frames per second, while giving only slightly worse results.

6. Conclusions and Outlook

We have presented a fast and robust method for landmark tracking in x-ray locomotion sequences. By extending robust

state-of-the-art sub-template based matching approaches, we are able to successfully track landmarks undergoing severe occlusions. One main advantage of our proposed approach is that the tracking hypotheses generated by sub-templates are fused in a soft manner, whereas recent approaches assume a hard decision for each sub-template. Also, we showed how the usage of additional—in our case anatomical—knowledge can be used to further improve the tracking results. Based on diverse experiments on real-life x-ray locomotion datasets, we showed that our approach is superior to other robust approaches.

An interesting point for further work is the inclusion of context knowledge, e.g. by exploiting the relative positions of other landmarks for the tracking process.

Acknowledgements

The authors would like to thank Alexander Stöbel for providing the Q1 dataset and John Nyakatura from the Institute of Systematic Zoology at the Friedrich Schiller University of Jena for providing the Q2, Q3, B1, L1, and L2 dataset.

This research was supported by grant DE 735/8-1 of the German Research Foundation (DFG).

References

- [BBG*10] 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 (2010), 262–279. 1
- [BK08] BRADSKI G., KAEHLER A.: *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly, Cambridge, MA, 2008. 3, 7
- [CET01] COOTES T. F., EDWARDS G. J., TAYLOR C. J.: Active appearance models. *IEEE T. Pattern Anal.* 23, 6 (2001), 681–685. 2
- [DGBD05] DEUTSCH B., GRÄSSL C., BAJRAMOVIC F., DENZLER J.: A comparative evaluation of template and histogram based 2d tracking algorithms. In *Proceedings of the 27th DAGM Symposium (DAGM 2005), Vienna, Austria, August 31 - September 2, 2005* (2005), pp. 269–276. 2
- [FL11] FISCHER M., LILJE K.: *Dogs in Motion*. VDH, 2011. 1
- [HB98] HAGER G. D., BELHUMEUR P. N.: Efficient region tracking with parametric models of geometry and illumination. *IEEE T. Pattern Anal.* 20, 10 (1998), 1025–1039. 2, 3
- [HD11a] HAASE D., DENZLER J.: Anatomical landmark tracking for the analysis of animal locomotion in x-ray videos using active appearance models. In *Proceedings of the 17th Scandinavian Conference on Image Analysis (SCIA 2011)* (2011), Heyden A., Kahl F., (Eds.), no. 6688 in LNCS, Springer, pp. 604–615. 2
- [HD11b] HAASE D., DENZLER J.: Comparative evaluation of human and active appearance model based tracking performance of anatomical landmarks in locomotion analysis. In *Proceedings of the 8th Open German-Russian Workshop Pattern Recognition and Image Understanding (OGRW-8-2011)* (2011), pp. 96–99. 6
- [HIN10] HOLZER S., ILIC S., NAVAB N.: Adaptive linear predictors for real-time tracking. In *23rd IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13-18 June 2010* (2010), pp. 1807–1814. 2
- [IMB02] ISHIKAWA T., MATTHEWS I., BAKER S.: *Efficient Image Alignment with Outlier Rejection*. Tech. Rep. CMU-RI-TR-02-27, Carnegie Mellon University Robotics Institute, 2002. 2, 3, 4, 6, 7
- [JD02a] JURIE F., DHOME M.: Hyperplane approximation for template matching. *IEEE T. Pattern Anal.* 24, 7 (2002), 996–1000. 2, 3, 6
- [JD02b] JURIE F., DHOME M.: Real time robust template matching. In *Proceedings of the British Machine Vision Conference 2002, (BMVC), Cardiff, UK, 2-5 September 2002* (2002). 2, 3, 4, 6, 7
- [Jur99] JURIE F.: Solution of the simultaneous pose and correspondence problem using gaussian error model. *Computer Vision and Image Understanding* 73, 3 (1999), 357–373. 2
- [KdA07] KIM H. Y., DE ARAÚJO S. A.: Grayscale template-matching invariant to rotation, scale, translation, brightness and contrast. In *Proceedings of the 2nd Pacific Rim Symposium on Advances in Image and Video Technology (PSIVT), Santiago, Chile, December 17-19, 2007* (2007), pp. 100–113. 3, 4
- [LK81] LUCAS B. D., KANADE T.: An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81), Vancouver, BC, Canada, August 1981* (1981), pp. 674–679. 2, 6
- [Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* 60, 2 (2004), 91–110. 2, 6
- [MGK00] MATAS J., GALAMBOS C., KITTLER J.: Robust detection of lines using the progressive probabilistic hough transform. *Computer Vision and Image Understanding* 78, 1 (2000), 119–137. 5
- [MIB04] MATTHEWS I., ISHIKAWA T., BAKER S.: The template update problem. *IEEE Trans. Pattern Anal. Mach. Intell.* 26, 6 (2004), 810–815. 2, 3, 5
- [NWvdB01] NGUYEN H. T., WORRING M., VAN DEN BOOMGAARD R.: Occlusion robust adaptive template tracking. In *8th IEEE Conference of Computer Vision (ICCV)* (2001), pp. 678–683. 2
- [PHZ08] PAN J., HU B., ZHANG J. Q.: Robust and accurate object tracking under various types of occlusions. *IEEE Trans. Circuits Syst. Video Techn.* 18, 2 (2008), 223–236. 2, 4, 5, 6
- [RN88] RODGERS J. L., NICEWANDER W. A.: Thirteen ways to look at the correlation coefficient. *The American Statistician* 42, 1 (1988), 59–66. 3
- [SA85] SUZUKI S., ABE K.: Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing* 30, 1 (1985), 32–46. 5
- [Sch07] SCHREIBER D.: Robust template tracking with drift correction. *Pattern Recogn Lett* 28, 12 (2007), 1483–1491. 2
- [TS94] TOMASI C., SHI J.: Good features to track. *IEEE Conference on Computer Vision and Pattern Recognition* (1994), 593–600. 2, 4, 6
- [WB95] WELCH G., BISHOP G.: *An Introduction to the Kalman Filter*. Tech. rep., Chapel Hill, NC, USA, 1995. 2, 4
- [YJS06] YILMAZ A., JAVED O., SHAH M.: Object tracking: A survey. *ACM Comput. Surv.* 38, 4 (2006). 2, 3, 4, 6, 7