scientific reports

OPEN

Markerless 3D kinematics and force estimation in cheetahs

Zico da Silva[1]✉, Stacey Shield[1], Penny E. Hudson[2], Alan M. Wilson[3], Fred Nicolls[1] & Amir Patel[1]

The complex dynamics of animal manoeuvrability in the wild is extremely challenging to study. The cheetah (*Acinonyx jubatus*) is a perfect example: despite great interest in its unmatched speed and manoeuvrability, obtaining complete whole-body motion data from these animals remains an unsolved problem. This is especially difficult in wild cheetahs, where it is essential that the methods used are remote and do not constrain the animal's motion. In this work, we use data obtained from cheetahs in the wild to present a trajectory optimisation approach for estimating the 3D kinematics and joint torques of subjects remotely. We call this approach kinetic full trajectory estimation (K-FTE). We validate the method on a dataset comprising synchronised video and force plate data. We are able to reconstruct the 3D kinematics with an average reprojection error of 17.69 pixels (62.94% PCK using the nose-to-eye(s) length segment as a threshold), while the estimates produce an average root-mean-square error of 171.3N (≈17.16% of peak force during stride) for the estimated ground reaction force when compared against the force plate data. While the joint torques cannot be directly validated against ground truth data, as no such data is available for cheetahs, the estimated torques agree with previous studies of quadrupeds in controlled settings. These results will enable deeper insight into the study of animal locomotion in a more natural environment for both biologists and roboticists.

Keywords Pose estimation, Inverse dynamics, Trajectory optimisation

High-speed manoeuvrability is the "final frontier" in the study of legged locomotion. While constant-speed gait has been studied extensively, the dynamics and control of these transient movements are still sparsely investigated. This could be attributed to the complex dynamics manoeuvring entails, which require whole-body motion and force data to quantify. This data can be gathered in a lab setting[1] but does not reflect the natural locomotor conditions or animal motivations experienced in the field. Rapid manoeuvres are important to understand, however. From a biomechanics perspective, they push animals to perform at their mechanical limits, revealing what they are capable of in the most extreme circumstances. They are also interesting to the robotics field, as understanding this category of motion will be crucial for the development of more agile robotic systems that better match the capabilities of animals [2].

The cheetah (*Acinonyx jubatus*) is the perfect model for studying quadruped dynamics as it is not only the fastest terrestrial animal, but also one of the most manoeuvrable. In fact, a study which used GPS-IMU collars to investigate the behaviour of wild cheetahs revealed that it is the ability to rapidly accelerate that is most critical to their hunting success[3]. Tracking collars are, however, only able to treat the animal as one lumped rigid body and are thus unable to provide information about leg, spine or tail kinematics or joint loading.

By contrast, vision-based pose estimation methods provide a means for non-invasive full-body kinematic and kinetic estimation ("dynamic" estimation). In human research, this technique has been adopted to obtain 3D pose estimates from a single camera, i.e. monocular 3D pose estimation[4,5], focusing on full-body kinematics. While data-driven models or purely kinematic formulations are widely applied in research of human locomotion, there have been efforts to incorporate a more complete physics-based model[6,7], which often is formulated as a non-linear program (NLP) to solve for both the kinematic and kinetic parameters.

These models provide a strong prior on the motion, making it a popular choice for potentially ambiguous pose detection from a single camera setup. The internal and external torques and forces are a natural byproduct of modelling the kinetics together with the kinematics. Some researchers have used this to explicitly analyse joint torques produced by humans while performing different activities[8,9]. In conjunction, optimal-control-based approaches have been used to fit a human model to corresponding kinematic data to obtain joint torques and contact forces[10,11].

[1]Department of Electrical Engineering, University of Cape Town, Cape Town 7700, South Africa. [2]Institute of Sport Nursing and Allied Health, University of Chichester, Chichester PO19 6PE, UK. [3]Structure and Motion Laboratory, The Royal Veterinary College, London NW1 0TU, UK. ✉email: zicods7@gmail.com

For all the work done on humans, there is little to show for markerless dynamic motion estimation of animals in the wild. That said, animal 3D pose estimation (without explicit modelling of physics), for both multi-view camera and monocular systems, have been well established in recent years. The multi-view camera techniques often resort to triangulation-based methods[12–14], while the monocular systems have used the skinned multi-animal linear model (SMAL)[15,16] and pose "lifting"[17] to disambiguate 3D pose estimates from a single view. These methods have been successful at 3D pose estimation of animals in the wild, but they lack the joint torques and ground reaction forces (GRFs) that are important for biomechanic analysis and robotic design.

Researchers have estimated animal torques and GRFs using invasive methods in controlled experiments. In one study, researchers placed reflective markers on racing greyhounds and, together with a six-camera setup and force plates, were able to determine joint torque and power profiles of the hind limbs[18]. Similar work was done on the sit-to-stand dynamics of greyhounds[19]. In the same line of work, a study was done on the forelimb muscle activity of horses[20].

While there are examples of dynamic motion estimation being applied to animals, work has mostly been limited to a laboratory setting, or involved invasive markers (or trackers) being placed on the subjects in the wild. In contrast, the proposed method performs markerless full-body kinetic and kinematic estimation of animals (in this case, cheetahs) with a multi-camera setup. To the best of our knowledge, this is a novel approach to the dynamic estimation of wildlife locomotion in the wild.

In our previous work, AcinoSet [14,21], we were able to reconstruct the 3D kinematics of free-running cheetahs from multiple cameras. Here, we extend this work to include kinetic modelling for the cheetah using a complete physics-based model to describe the motion. In particular, we are interested in the joint torques during a gallop. To achieve this, we created a new dataset (denoted "kinetic dataset" in this work) that contains synchronised video and force plate data. As with AcinoSet, a trajectory optimisation problem (henceforth referred to as the kinetic full trajectory estimation (K-FTE) method is formulated and solved for the dynamic motion estimate, and then evaluated against the kinetic dataset. This differs from the purely kinematic full trajectory estimation (FTE) method developed for AcinoSet. Once the K-FTE method has been validated with the kinetic dataset, we can use this method to perform motion and torque estimation on a test set of AcinoSet. This provides valuable information about the joint loading of the limbs of the cheetah during locomotion.

Even though this work focuses on the cheetah, the proposed method can easily be generalised to work with any wild animal. This will enable biologists to understand animal locomotion in much more detail and can provide data for the design of new robotic controllers.

Results

We first present the quantitative and qualitative results of the motion and torque estimation on both datasets, and then we motivate its use using the kinetic dataset to compare estimated GRFs with the ground truth measurements from the force plates.

Motion and torque estimation

The cheetah was represented using a model consisting of 17 rigid bodies, shown in Fig. 1. The 3D pose estimation results are shown in Table 1. The mean position error (MPE) provides a relative measure of how well the reconstruction matches the baseline FTE result using our previous kinematic methods[14,21]. The MPE is minimal, suggesting that the K-FTE method produces similar 3D kinematics to the baseline FTE.

Note that we calculated an average reprojection error of 17.69 pixels, or 62.94 % percentage of keypoints (PCK) using the nose-to-eye(s) length threshold), using a subset of 540 frames of 2D hand-labelled data as ground truth from two different trials, T4 and T5. This result compliments the MPE shown in Table 1, while also providing an absolute measure of the accuracy of the 3D reconstruction.

Figure 2 shows a single pose viewed from all six camera angles. Visually, the pose estimate is reasonably accurate given the estimate is well-matched to the cheetah's skeleton in each view.

The estimated joint angles and torques for the fore and hind limbs, while in contact with the ground (stance phase of the gait), are shown in Figs. 3 and 4 for the kinetic dataset and AcinoSet respectively. The torque estimates produced for the kinetic dataset serve as a reference for what is expected given that we know the corresponding GRF. There is a good correlation between the torque estimates during the stance for both datasets.

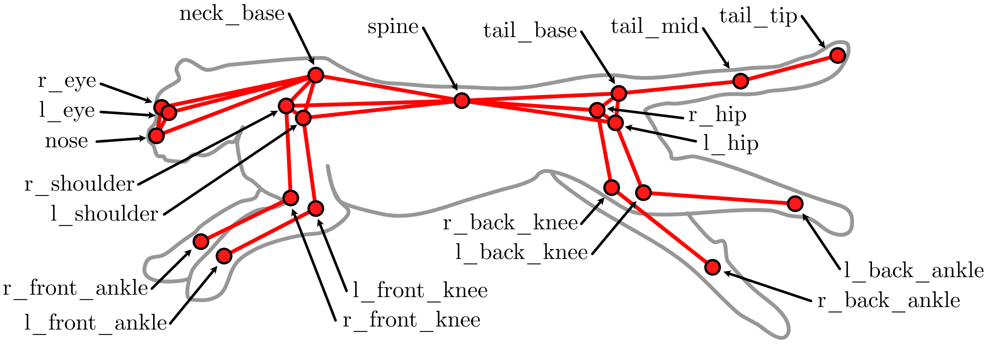

Figure 1. Multibody model of the cheetah consisting of 17 rigid links and 23 points of interest.

Test	MPE (mm)	% of deviation
T1	17.6	1.3
T2	34.0	1.4
T3	19.2	0.8
T4	20.7	2.5
T5	23.5	1.3

Table 1. Error analysis of the resultant 3D reconstruction on AcinoSet that includes a physics-based model for dynamic data estimation.

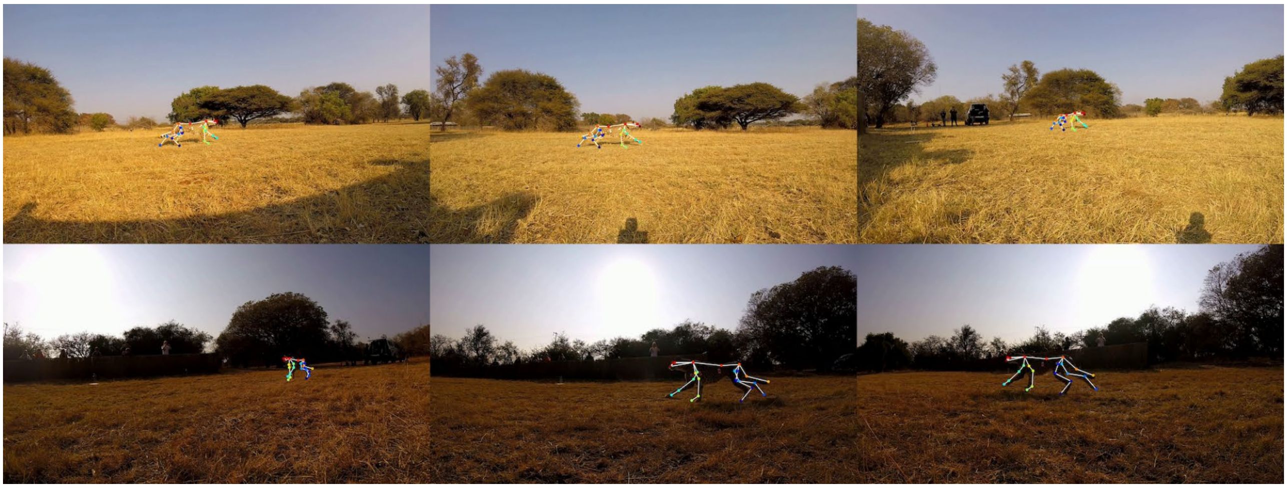

Figure 2. An visual example of a single pose viewed from all six camera angles for T1 of AcinoSet. This illustrates good 3D pose estimates of the cheetah.

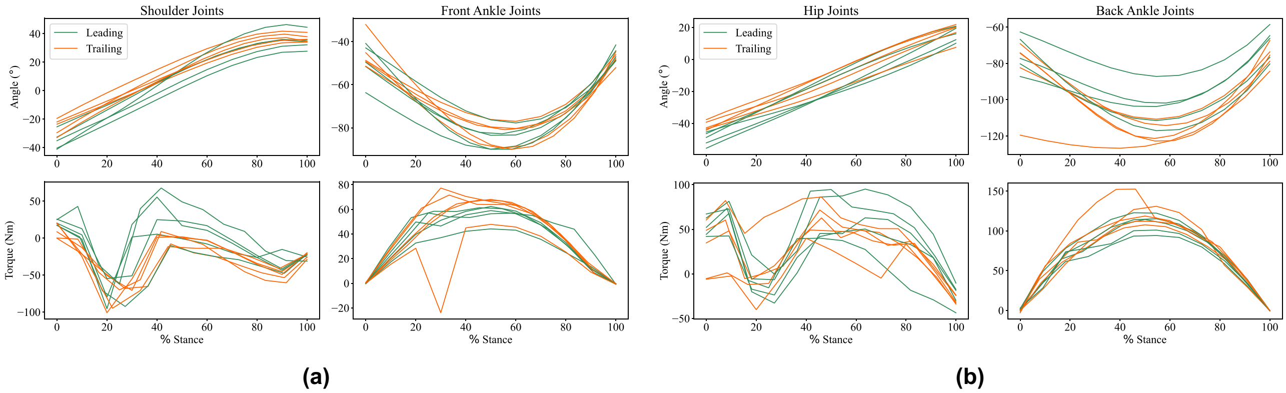

Figure 3. Joint angle and torque estimates of the forelimb (**a**) and hindlimb (**b**) hip and hock joints during stance phase of gait for the kinetic subset of the evaluation dataset. The top rows are the joint angles and bottom rows the joint torques. The joint angles and torques are consistent during the stance phase.

Based on the kinetic dataset (see Fig. 3), both the fore and hind limbs exhibit similar torque trajectories. For the front and back ankle joints, the torque profile during the stance follows a half sine wave, where the peak torque on average is around 50%. The hindlimb produces a slightly higher peak torque on average than the forelimb.

The hip torques do not conform to as consistent a profile as the ankle torques do, but they are consistently bounded between 100 and −50 N m. The trailing shoulder joint torques produce peaks at around 20 % of the stance, and quickly returns to zero thereafter. A similar pattern is found in the trailing shoulder joint, however the torque peaks on average around two points in the stance: a negative peak at 20 % and a positive peak at 40 %.

There is consistency in the joint angles for the duration of the stance. The shoulder and hip joints increase linearly from start to finish, while the front and back ankle joints display a parabola-like change during the stance.

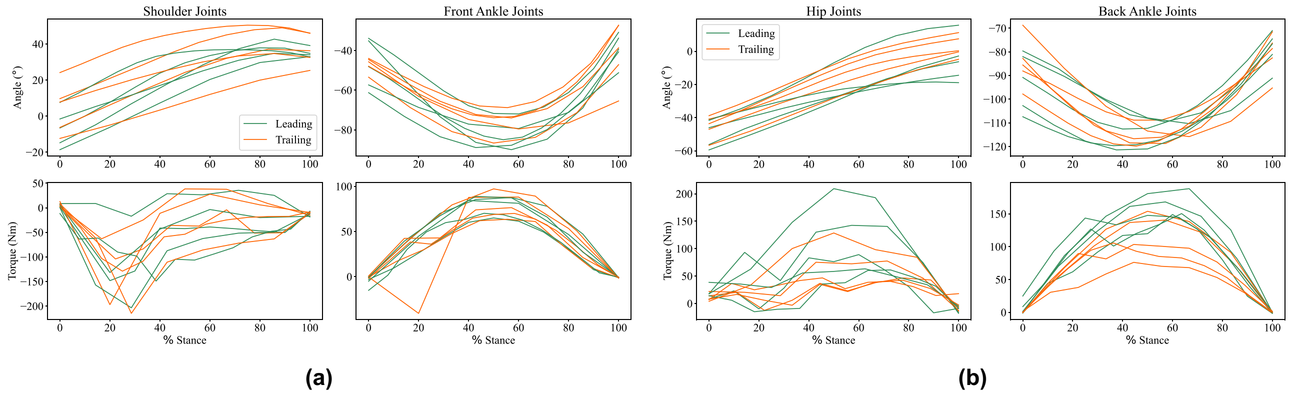

Figure 4. Joint angle and torque estimates of the forelimb (**a**) and hindlimb (**b**) hip and hock joints during stance phase of gait for AcinoSet subset of the evaluation dataset. The top rows are the joint angles and bottom rows the joint torques. The joint angles and torques are consistent during the stance phase.

Method validation

It must be stressed that the force and torque values obtained using this method are *estimates* at best—not remote measurements. If multiple feet are on the ground, the inverse dynamics problem does not have a unique solution, so the results are merely one possible combination of forces out of many that could produce the observed motion in the 3D model. Still, these estimates should give a reliable indication of the magnitudes of the forces and torques involved, even if the precise profiles cannot be confirmed.

Without ground truth joint torques, we set up an alternative evaluation procedure to determine whether the observed joint torques are plausible. We make use of the measurements from the force plate as a proxy for a "ground truth" to validate motion and torque estimates. Two K-FTE methods for estimating the ground reaction forces were compared: sinusoidal GRF and freeform GRF. The sinusoidal approach assumes a "sinusoidal" GRF profile, while the freeform approach does not make any assumptions about the GRF profile. The idea of using a sinusoidal profile for the GRF was taken from a previous study that examined the GRFs on horses during steady-state galloping[22]. In addition, the force plate obtained GRF profiles found in the kinetic dataset exhibited the familiar sinusoidal shape, and therefore, it made sense to develop a method that uses this inherent structure.

Table 2 provides the GRF estimation error for both methods, and Fig. 5 provides a visual example of the GRFs.

Method	Mean error (N)	% of peak
sinusoidal GRF	171.3	17.16
freeform GRF	471.3	47.23

Table 2. GRF estimation comparison between the estimated and the force plate measured GRF magnitudes from the kinetic dataset.

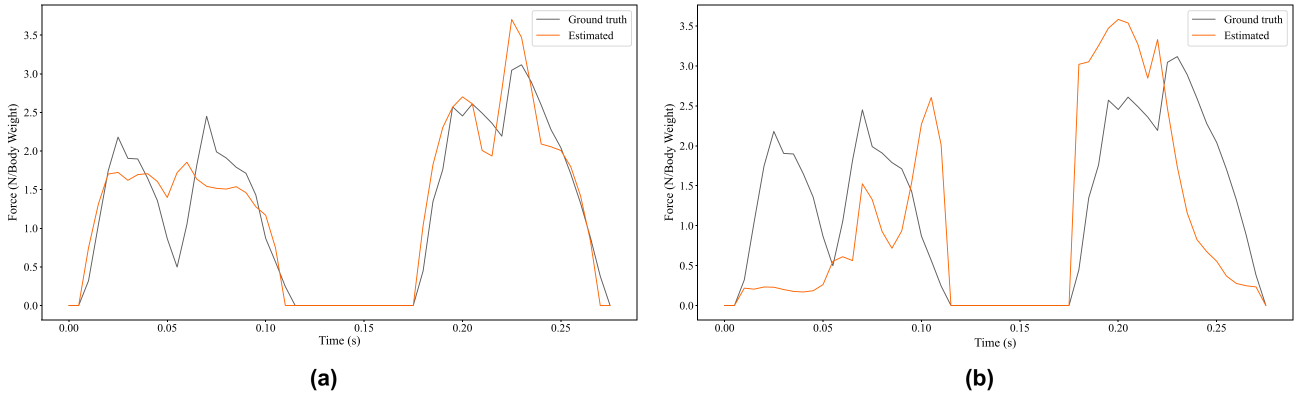

Figure 5. A representative example comparing the resultant GRF magnitudes produced by both methods: (**a**) sinusoidal GRF, and (**b**) freeform GRF. This is a result from T4 of the kinetic dataset. The force has been normalised by the body weight of the subject.

The sinusoidal K-FTE makes a drastic improvement on the GRF error that is evident in both Table 2 and Fig. 5. From the example in the figure, the sinusoidal K-FTE provides good tracking of the GRF trajectory, whereas the freeform K-FTE fails to produce a valid GRF trajectory.

Discussion

The kinetic dataset facilitated the development of a whole-body dynamic state estimator using the K-FTE method. From Table 2, it is clear that the sinusoidal K-FTE produces more accurate GRF estimates. The GRF estimate resembles that of the force measurements from the force plates (see Fig. 5a). Many of the state-of-the-art physics-based methods have validated GRF estimate errors in a similar fashion, by using existing studies on predetermined locomotion styles[7,23]. However, one study validated its GRF estimates using a parkour dataset[6], which quotes similar mean errors to our work (see Table 2).

Furthermore, this result confirms the use of the assumed sinusoidal GRF profile for steady-state galloping of quadrupeds. This was previously applied to horses, where it was also found to be a good fit for the data[22]. Limiting the GRF to a sinusoidal profile allowed for an easier optimisation problem to solve, and reduced the problem of non-unique solutions by restricting the space of possible GRFs to those matching a widely-observed form from previous quadruped studies.

Note that although the GRF error is vastly different between the two methods reported in Table 2, the resultant 3D reconstructions are similar (calculated average MPE of 13.0). Therefore, the freeform K-FTE does yield forces with plausible magnitudes even if the profile does not match the observed truth data. We compared the ground truth impulse to the estimated impulse for both methods for the same trial (T4 of the kinetic dataset). The ground truth impulse was calculated as $119.4\,kg\ \mathrm{ms}^{-1}$ and the sinusoidal K-FTE produced an impulse of $122.32\,\mathrm{kg\ ms}^{-1}$ whereas the an impulse of $97.50\,kg\ \mathrm{ms}^{-1}$ was obtained for the freeform K-FTE. Furthermore, this is confirmed by the GRF trajectory in Fig. 5a, where the area under the curve (i.e. the impulse) is similar to the ground truth ($< 20\,\%$ error).

The estimated joint torques produced a consistent torque profile on both datasets (see Figs. 3 and 4). This result is in line with the results on the hindlimbs of racing greyhounds (similar to the cheetah in that they are very fast quadrupeds)[18]. However, the quoted torques are of different magnitudes, with the greyhound's back ankle average peak roughly calculated as 40 N m, while the cheetah's average peak was calculated to be around 100 N m. This is roughly twice the torque generated during the stance when compared to the greyhound. It is known that the cheetah generates larger joint torques to resist larger GRFs than the greyhound[24]. However, it would not be advisable to place too much emphasis on this particular result given the simplified kinematic model used (in addition to inverse dynamics) to generate these quantities in this work.

The 3D pose estimation is adequate and has not drastically changed from the previous AcinoSet baseline[21], as shown in Table 1 and Fig. 2. The MPE is very low and suggests that adding a more complete physics model has not diminished the resultant kinematics. Therefore, without compromising our pose estimation accuracy, we have gained kinetic information, and with it, a more complete picture of the dynamics of the cheetah.

That said, an average PCK of 62.94 % suggests that there is room for improvement in our 2D pose estimation. This result could be attributed to the shortcomings of using a simplified rigid body model for the cheetah. The rigid body model, together with fixed joint centres, are not representative of the flexibility within the cheetah's skeleton. In addition, modelling the muscles as torque-driven actuators is clearly not characteristic of a cheetah's muscle system. Lastly, point contacts are used to model the dynamics between the foot and the ground, which completely overlooks the complexity of the cheetah's foot/paw. As highlighted in Fig. 6, there are times when a rigid body model cannot accommodate the forelimbs and neck body parts because of a lack of extension/contraction.

Nevertheless, the developed method provides a stepping stone to accurate dynamic analysis of wild quadrupeds using a non-invasive multi-camera system. Ultimately, the accuracy of the estimated 3D kinematics were adequate for our purposes.

Looking to the future, we have identified two aspects of the current implementation that we could improve: removing the explicit assumption on the GRF profile, and the rigid body model. First, we would like to investigate the incorporation of a learned prior model of the GRF profile in the cost function as a possible improvement to the existing freeform GRF method. The probability model can use examples from existing force data on quadrupeds. This would enable the study of more dynamic movements (turns and rapid acceleration or deceleration).

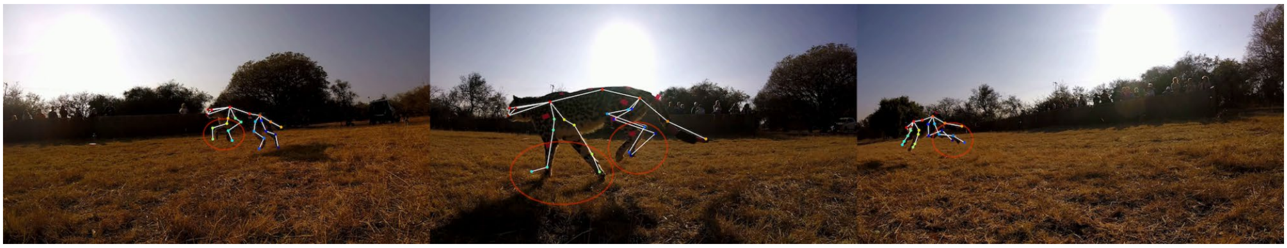

Figure 6. Visual examples of bad skeleton fits when projected onto the image plane after dynamic estimation. The red circles show clear pose estimation issues, the main contributor being the rigid body model failing to adapt to the extension/contraction of the forelimbs during locomotion.

Second, even though the overall 3D kinematic estimate was sufficient for this work, we can potentially obtain more accurate predictions by remodelling the shoulder and neck links as variable length[25] which should address the problem highlighted in Fig. 2.

Methods

We first provide background theory to multi-body dynamics and the K-FTE method used in this work. Then, we provide more details on the the kinetic and AcinoSet datasets. Lastly, we present the evaluation metrics used in this work.

Multi-body dynamics

The dynamics of the cheetah was modelled as a rigid multi-body system using absolute angle coordinates $\mathbf{q}(t)$ [26], often expressed as

$$\mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} = \boldsymbol{G}(\mathbf{q}) + \mathbf{Bu} + \mathbf{J}_g^T(\mathbf{q})\boldsymbol{\lambda}_g + \mathbf{J}_c^T(\mathbf{q})\boldsymbol{\lambda}_c, \tag{1}$$

where $\mathbf{M}(\mathbf{q})$ represents the inertia matrix, $\mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})$ captures Coriolis and centrifugal terms, $\mathbf{G}(q)$ is the gravity vector, $\mathbf{B}$ maps $\mathbf{u}$ inputs to generalised forces, $\boldsymbol{\lambda}_g$ are the ground constraint forces, and $\boldsymbol{\lambda}_c$ are the constraint torques that enforce the geometry of the model. These constraint forces are mapped to the generalised coordinates of the model through the Jacobians, $\mathbf{J}_g^T(\mathbf{q})$ and $\mathbf{J}_c^T(\mathbf{q})$. For brevity, we will combine these contact forces into a single constraint force vector $\boldsymbol{\lambda}$, mapped by a single Jacobian $\mathbf{J}$. In this work, $\mathbf{Bu}$ represents the joint torques produced by the cheetah during locomotion, henceforth denoted by $\boldsymbol{\tau}$.

The 17-link rigid multi-body model of the cheetah results in a state vector $\mathbf{q}$ of size 54 (3 angles for each link and an additional 3 for the position of the base link in the inertia frame).

Each link was modelled as a cylinder described by three parameters, length, mass, and radius. These parameters were determined per subject, using the dimensions and masses quoted in multiple resources produced Hudson et al.[24,27,28]. Note that the length parameters were fine-tuned by projecting the cheetah skeleton into one of the cameras and validating that the links match the limb lengths of the cheetah.

Trajectory optimisation

In general, the trajectory optimisation problem can be described as a non-linear program (NLP). Here, we formulate a NLP with the following constraints:

$$\underset{\mathbf{q},\boldsymbol{\tau},\boldsymbol{\lambda},\boldsymbol{T_c}}{min} \quad \sum_{k=0}^{N} g(\mathbf{q}_k, \boldsymbol{\tau}_k) \tag{2}$$

subject to

$$\dot{\mathbf{q}}_k = f(\mathbf{q}_k, \boldsymbol{\tau}_k, \boldsymbol{\lambda}_k) \qquad \forall k \in [0, N], \tag{3}$$

$$C(\mathbf{q}_k, \boldsymbol{\lambda}_k) = 0 \qquad \forall k \in [0, N], \tag{4}$$

where $f(\cdot)$ denotes the multi-body dynamics (Equation (1)) and $C(\cdot)$ are the path and contact constraints. The cost function to optimise is $g(\cdot)$. The index k indicates the discrete instant in time (or *node*) within the trajectory. The state variables at each node are related to those at the previous node by numerical integration constraints

$$\mathbf{q_k} = \mathbf{q_{k-1}} + h\dot{\mathbf{q}}_k \qquad \forall k \in [1, N], \tag{5}$$

$$\dot{\mathbf{q}}_k = \dot{\mathbf{q}}_{k-1} + h\ddot{\mathbf{q}}_k \qquad \forall k \in [1, N], \tag{6}$$

where h is the timestep between nodes.

An equality constraint is used to enforce the dynamics of the system (i.e. $f(\mathbf{q}_k, \boldsymbol{\tau}_k, \boldsymbol{\lambda}_k)$, see Equation (1)):

$$\mathbf{M}_k\ddot{\mathbf{q}}_k + \mathbf{C}_k\dot{\mathbf{q}}_k - (\mathbf{G}_k + \boldsymbol{\tau}_k + \mathbf{J}_k^T\boldsymbol{\lambda}_k) - \mathbf{w}_k = \mathbf{0} \qquad \forall k \in [1, N] \tag{7}$$

The function notation has been omitted for brevity (e.g. $\mathbf{M}(\mathbf{q})$ is simply stated as $\mathbf{M}$). The additional variable $\mathbf{w_k}$ added to this equation accounts for unknown disturbances to the model, as might be introduced by factors such as variation in the animal subjects and ground conditions.

In addition to the motion constraint above, two constraints to incorporate the measurement data from the cameras are specified:

1. A constraint that relates the current pose $\mathbf{q}$ to a set of 3D marker positions.
2. A constraint that relates reprojection of the 3D marker positions to the 2D pixel coordinates for each camera.

Constraint 1 can be determined using the pose equations that relate the generalised coordinates to positions. This is defined as a general function $g(\mathbf{q}_k) = \mathbf{x}_k$ made into an equality constraint, $g(\mathbf{q}_k) - \mathbf{x}_k = \mathbf{0}$, at finite time k. Constraint 2 uses a camera projection matrix, $\mathbf{P}$, to transform a 3D marker from world coordinates to 2D image space. The aim is to minimise the error between the ground truth and reprojected pixel locations; therefore,

$\mathbf{v}$ is used to capture this error so that it can be added to the cost function. The resultant equality constraint is defined as

$$\mathbf{y}_{k,c,m} - \mathbf{P}_c \mathbf{x}_{k,m} - \mathbf{v}_{k,c,m} = \mathbf{0}, \tag{8}$$

where c selects a specific camera to project to, m selects a specific marker on the cheetah, and k is the finite time that this operation is performed at.

The constraints used in the optimisation include not only the measurement data but also two additional types of constraints: contact constraints to enforce the correct GRF at nodes where a foot has been determined to be grounded, represented by $\boldsymbol{\lambda}_{g,k}$, and joint constraints to ensure that constraint torques, $\lambda_{c,k}$, to prevent motion in forbidden directions (which is necessary when using absolute angle coordinates).

The ground contact constraints are enforced only at those nodes where a foot has been determined to be grounded. We will refer to this subset of nodes as D, and use the notation $\hat{\boldsymbol{\lambda}}_{\boldsymbol{g}}$ to refer to the contact force vector acting on the grounded foot. A no-slip model of contact is developed using the foot velocity $\mathbf{v}_f$ and contact force $\boldsymbol{\lambda}_{\boldsymbol{g}}$ to construct the following constraints,

$$\hat{\boldsymbol{\lambda}}_{g,k} > 0 \qquad k \in D \tag{9}$$

$$|\mathbf{v}_{f,k}| \leq \epsilon \qquad k \in D, \tag{10}$$

The first constraint enables the generation of a GRF vector, while the second constraint ensures the foot is approximately fixed ($\epsilon \leq 1$) to the ground during the contact phase. The contact constraints are relaxed to within a small penalty variable, ϵ, which is minimised in the cost function. This penalty approach makes the optimisation problem easier to solve.

Since sliding contact is not considered, the contact force $\boldsymbol{\lambda}_g$ vector has to satisfy friction cone constraints[29] to estimate valid forces. To obtain a simpler model of contact, linearised friction cone constraints are formulated (resulting in a friction pyramid):

$$|\hat{\lambda}_{x,k}| \leq \mu_s \hat{\lambda}_{z,k} \qquad k \in D, \tag{11}$$

$$|\hat{\lambda}_{y,k}| \leq \mu_s \hat{\lambda}_{z,k} \qquad k \in D \tag{12}$$

where $\hat{\lambda}_x$ and $\hat{\lambda}_y$ are the tangential components of the contact force, $\hat{\lambda}_z$ is the vertical component ($\hat{\lambda}_z > 0$), and μ_s is the friction coefficient (and was set to 1.3 for all experiments[3]). The tangential components of the contact force are decomposed into two positive variables in implementation, i.e. $\hat{\lambda}_x = \hat{\lambda}_x^+ - \hat{\lambda}_x^-$ where $\hat{\lambda}_x^+, \hat{\lambda}_x^- \geq 0$. This facilitates the optimisation by replacing the non-linear absolute value function with a linear version (e.g. $|\hat{\lambda}_x| \equiv \hat{\lambda}_x^+ + \hat{\lambda}_x^-$) in the above friction cone constraints. (Note that this is only true if either $\hat{\lambda}_x^+$ or $\hat{\lambda}_x^-$ is zero. IPOPT ensured that this condition was true.)

There is a requirement for explicit modelling of joints to remove redundant degrees of freedom introduced by the absolute angle coordinates. Two different types of joints are supported: revolute and universal. The former is created by the constraints

$$\mathbf{r}_{i,y} \cdot \mathbf{r}_{i+1,x} = 0, \tag{13}$$

$$\mathbf{r}_{i,y} \cdot \mathbf{r}_{i+1,z} = 0, \tag{14}$$

where the vector $\mathbf{r}_i i, y$ is aligned with the y axis of the first link, and $\mathbf{r}_i i+1, x$ and $\mathbf{r}_i i+1, y$ are aligned with the x and z axes of the subsequent link. We have left out the index k for clarity, as these constraints are applied at all nodes. When these constraints are met, the joint is restricted to rotate about the y-axis of the first link only, removing two degrees of freedom. This type of joint is applied at the cheetah's elbows and knees. Note that $\mathbf{r}_{i,y}$ is the column vector from the rotation matrix $\mathbf{R}$ that associates rotations about the y-axis.

A universal joint is represented by the constraint

$$\mathbf{r}_{i,y} \cdot \mathbf{r}_{i+1,x} = 0, \tag{15}$$

preventing a rotation about the z-axis and removing a single degree of freedom. This type of joint is used in the middle of the model's spine.

Finally, the cost function to be minimised is defined as

$$g(\mathbf{q}, \boldsymbol{\tau}) = \alpha_1 e_{\text{meas}} + \alpha_2 e_{\text{model}} + \alpha_3 e_{\text{smooth}}, \tag{16}$$

where $\alpha_1 = 1, \alpha_2 = 10000$, and $\alpha_3 = 1$. The e_{meas} term is defined as

$$e_{\text{meas}} = \sum_{k=1}^{N} \sum_{j=1}^{c} \sum_{i=1}^{m} \sum_{l=1}^{2} C\left(\frac{\mathbf{v}_{k,j,i,l}}{\sigma_i}\right), \tag{17}$$

where $C(\cdot)$ is the redescending robust cost function[30] and σ_i is used to normalise the measurement error. This uncertainty parameter is derived from the predicted 2D keypoint error distribution. The uncertainty parameters were obtained in a similar fashion to our previous work[14]. In addition, e_{model} now represents the error in the dynamic equation of motion instead of the assumed constant acceleration motion:

$$e_{\text{model}} = \sum_{i=1}^{n}\sum_{j=1}^{p} \mathbf{w}_{i,j}^2. \tag{18}$$

Note that the above error did not require normalisation.

Lastly, the e_{smooth} term is defined as:

$$e_{\text{smooth}} = \sum_{k=1}^{n}\sum_{l=1}^{t} 10(\boldsymbol{\tau}_{k,l})^2 + 0.1h^2 \sum_{k=1}^{n}\sum_{i=1}^{m} (\ddot{\mathbf{x}}_{k,i})^2, \tag{19}$$

where h is the frame rate. The first term penalises the joint torques to ensure a minimum energy solution. The literature suggests that it is a common assumption that humans and animals conserve energy during movement[31]. The 10 weight is used to increase its contribution during minimisation.

During experiments, it was noted that the optimal solutions would produce jittery motion. For this reason, the second term in Equation (19) was added to the cost function. This term penalises the acceleration of the 3D markers, favouring solutions that have smooth trajectories for all 3D markers. The h^2 weighting term performs normalisation, and the 0.1 weight is used to decrease its contribution during minimisation. Note that through normalisation, each term is assumed to be equally weighted in Equation (16). However, the goal is for the resultant motion estimate to be as physically plausible as possible, and therefore an enormous weight is placed on the e_{model} to direct the optimiser to find solutions that essentially force the model noise to be zero, i.e. $\mathbf{w} = \mathbf{0}$. Note that IPOPT[32] was used and configured with the MA97 linear solver[33] to implement the optimisation.

K-FTE

Trajectory optimsiation was used to investigate two K-FTE methods: sinusoidal GRF and freeform GRF. The sinusoidal K-FTE assumes a "sinusoidal" GRF profile that essentially reduces the optimisation problem to solely estimate the joint torques (we allow for 20 % variation from this assumption). While the freeform K-FTE does not make any assumptions about the GRF profile, it instead performs a complete joint estimation of the torques and GRFs. Both methods set up the same optimisation problem (as outlined above), but the sinusoidal K-FTE has an extra pre-processing step that generates the initial GRF profile. An example of the expected profile is shown in Fig. 7. We assume contact timing is known and therefore did not require an automatic contact detection algorithm.

In Fig. 7, the F_z component is created using a half sine wave with an amplitude that is determined through a linear model for each limb. The linear model relates the running speed of the cheetah with the peak vertical force[27]. The F_x component is approximated by a spline using 5 control points: zero points for the start, midpoint, and end of the stance, $\frac{F_z}{2}$ for the deceleration peak, and $\frac{F_z}{4}$ for the acceleration peak. We assume that there is no F_y component present in the straight-line steady-state gallops that are included in the datasets (in every trial, the cheetah runs along the x-axis).

Datasets

Two cheetah locomotion datasets were used to develop and evaluate the proposed approach. Neither was a perfect test case – one (AcinoSet) lacked force plate data, while the other used a non-standard camera calibration that resulted in slightly higher uncertainty in the animal's pose – however, the combination of the two with these

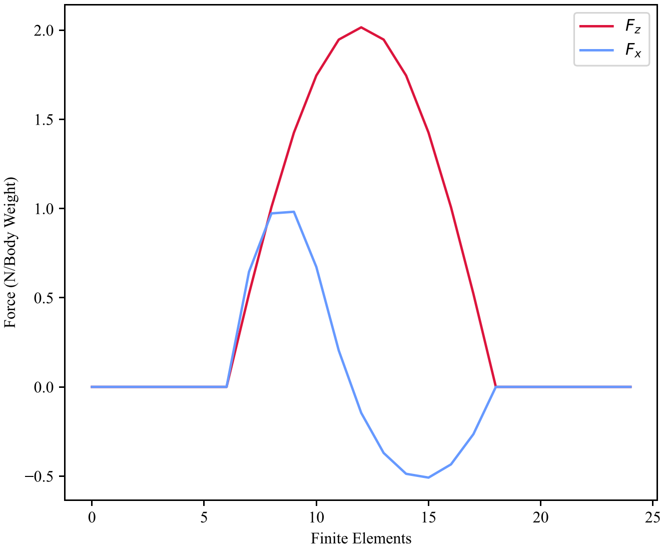

Figure 7. An example of the synthesis of the sinusoidal GRF profile.

complimentary deficiencies was sufficient to prove the concept. The availability of more complete and accurate locomotion data will likely improve this method in future.

AcinoSet

AcinoSet [14] is a cheetah-running dataset that is used for the evaluation of the motion and torque estimation. The dataset contains 90 running videos with six different views from low-cost GoPro cameras (2704 x 1520 resolution filmed at 120 frames per second (FPS) and 1920 x 1080 resolution, filmed at 90 FPS). There are 7588 human-annotated frames and the average video length is approximately 2 s. Furthermore, the dataset includes DeepLabCut [34] 2D pose estimates for each camera view for each video set, as well as the corresponding 3D reconstruction data, in addition to both the intrinsic and extrinsic calibration data. The intrinsic camera calibration data included fisheye lens distortion parameters.

Here, we use the DeepLabCut network that was developed in our previous work[21] and a subset of five trials from AcinoSet for evaluation (see Table 3). Each trial consists of the cheetah doing a steady-state gallop for two different subjects.

Kinetic dataset

This dataset was generated for this work to validate the proposed method of dynamic data estimation through the use of force plate measurements. The dataset was acquired from The Royal Veterinary College, and methods used to capture the data are described in their previous work[27].

The dataset contains grayscale videos (including calibration targets) filmed at 1000 FPS (1280 x 560 and 800 x 600 resolutions) with synchronised force plate data sampled at 3500 Hz (see Fig. 8). From the original dataset, a subset of five trials of two different subjects was selected (*Cheetah 1* and *Cheetah 2*, referred to in Table 1 of the paper by Hudson et al. [27])o. This included only those strides where each foot strike landed on a single force plate, to reduce ambiguity in the data. We denote this subset the ''kinetic dataset" in this work (see Table 4).

The original work gathered trials of the cheetahs chasing a mechanical lure across eight Kistler force plates (model 9287A and 9287B) and in front of four cameras[27]. Both the video and the force plate data were resampled to a 200 Hz sample rate for the kinetic dataset. This allows for a more tractable K-FTE problem formulation.

As the dataset only consisted of grayscale videos and synchronised force plate measurements, 2D keypoint inference and a scene calibration are required to enable 3D reconstructions of the cheetah.

2D pose estimation

The DeepLabCut [34] network trained for AcinoSet (RMSE of 8.38 pixels on the test set) was first used to obtain 2D keypoints on the cheetah in this dataset. However, to quantify the networks ability to infer 2D keypoints on an entirely new dataset, we needed to produce a new test set from the kinetic dataset. A subset of 260 frames of the new data was hand-labelled to form a test set, and the RMSE between ground truth and inferred 2D keypoints was 10.85 pixels. In order to reduce the error on the test set, we decided to train a new network to improve inference on the kinetic dataset. An additional 95 frames were manually labelled from the kinetic dataset to improve the training dataset for AcinoSet. This improved performance on the kinetic test set by reducing the RMSE to 6.10 pixels.

Camera calibration

The calibration techniques developed for AcinoSet were used to perform intrinsic and extrinsic calibration for the kinetic dataset. There were some hurdles to overcome, in that the dataset did not contain a standard checkerboard calibration target that was compatible with OpenCV's calibration library[35] (see Fig. 8b). This meant that the automatic corner finder could not be used. Instead, multiple frames of the calibration target from each camera were gathered and the points of interest were hand-labelled. There were nine points of interest that were chosen on the calibration target shown in Fig. 8b: the four inside corners of the black border, four centre points of the outer squares, and the single centre point of the centre square. This is a smaller number of points than typically used for intrinsic calibration, which could decrease the accuracy of the reconstruction of the fish-eye distortion model (particularly components that are orthogonal to the relatively narrow horizontal band that the calibration points were concentrated in), however, as these components are not likely to have a large effect on the parameters being estimated, we decided to proceed with this data set in spite of these deficiencies.

Test	Trial	Length (s)
T1	2017_08_29/top/phantom/run1_1	44/90
T2	2017_08_29/top/jules/run1_1	30/90
T3	2017_09_02/top/jules/run1	30/90
T4	2019_03_07/phantom/run	57/120
T5	2017_09_02/bottom/jules/run2	33/90

Table 3. The test dataset selected from the AcinoSet. The length is presented as the number of frames in the trajectory divided by the video frame rate. Each trajectory length is equivalent to one stride performed by the cheetah.

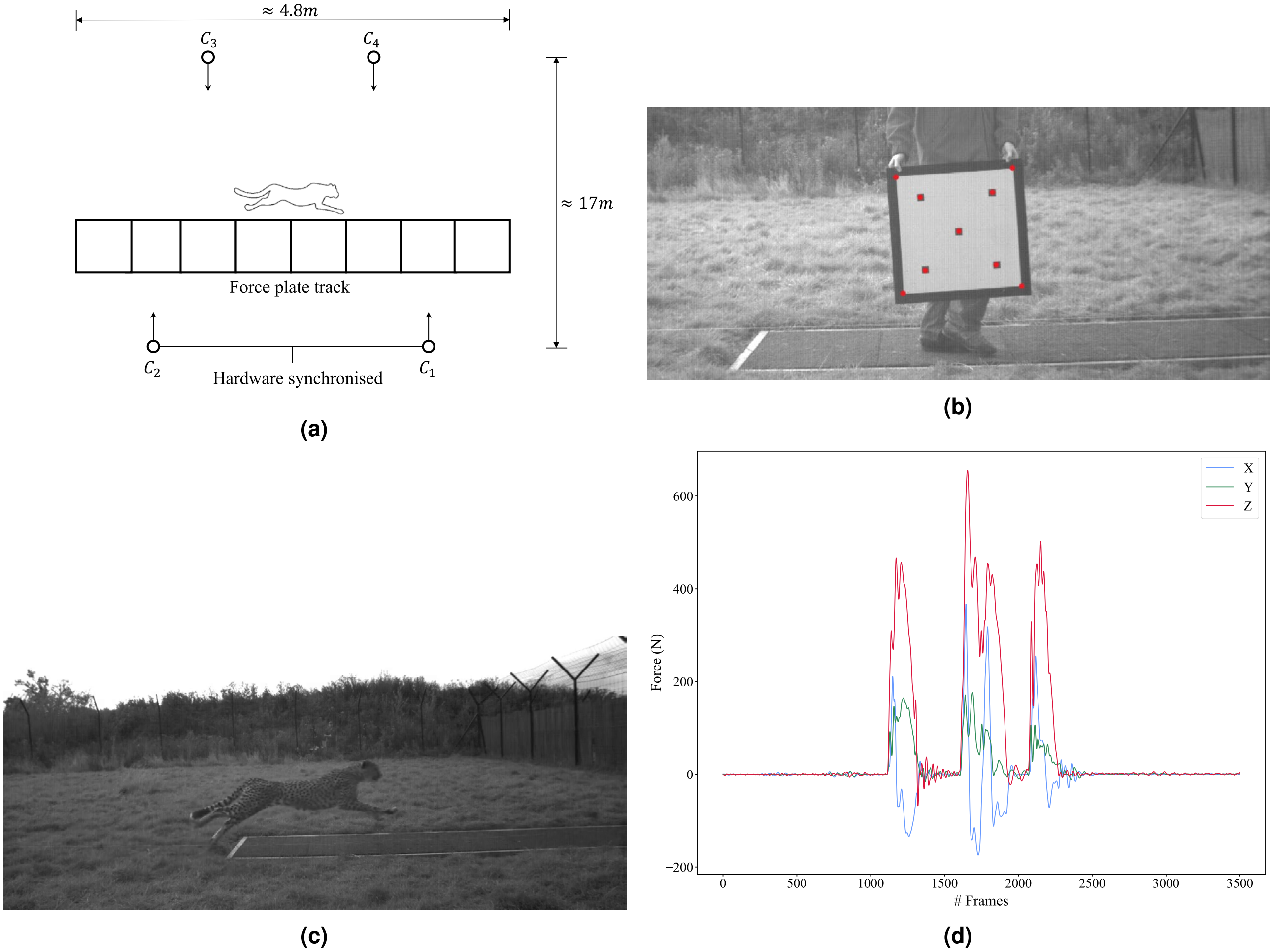

Figure 8. Example data from the kinetic dataset: (**a**) is a diagram of the scene setup, (**b**) provides an example of the calibration target used to perform intrinsic and extrinsic calibration, with the calibration points indicated in red, (**c**) shows a frame from the captured video, and (**d**) shows the measurement data from the force plates.

Test	Trial	Length (s)
T1	2009_09_07/arabia/trial06	49/200
T2	2009_09_07/shiraz/trial04	52/200
T3	2009_09_08/shiraz/trial04	52/200
T4	2009_09_11/shiraz/trial01	55/200
T5	2009_09_11/shiraz/trial02	48/200

Table 4. The test dataset selected from the kinetic dataset. The length is presented as the number of frames in the trajectory divided by the video frame rate. Each trajectory length is equivalent to one stride performed by the cheetah.

The hand labelling of the calibration target was done for roughly 16 frames across the camera's field of view, and for each trial included in the dataset. Once the hand labelling was done, a standard intrinsic calibration was performed to obtain the camera intrinsic parameters. Note that radial distortion parameters were estimated during the intrinsic camera calibration process.

The extrinsic calibration of all four cameras simultaneously proved difficult due to the lack of shared scenes between the cameras and the calibration target. In particular, the kinetic dataset used in the calibration process contained videos of the calibration target that were not seen in more than two cameras at the same time, and the calibration target was not viewed at the same time across the force plate track. This meant that there was no shared scene data between the first stereo pair (C_1 and C_2) and the second stereo pair (C_3 and C_4) shown in Fig. 8b. Based on the distance from the force plate track, cameras C_1 and C_2 are referred to as the near side stereo pair, and cameras C_3 and C_4 are referred to as the far side stereo pair.

To overcome this challenge, a stereo calibration was performed between cameras C_1 and C_2 to obtain the pose of camera C_2 relative to camara C_1. The world frame was set to the first force plate, and the pose of camera C_1 was determined using the perspective-n-point (PnP) algorithm[35]. This algorithm uses a set of 3D-2D point correspondences on the force plates, viewed by camera C_1, to compute the pose of the camera relative to the world frame.

After this stereo calibration, the extrinsic parameters for the near side stereo pair had been estimated. However, it was not possible to determine the pose of the far side cameras relative to the near side due to the lack of shared scenes of the calibration target. To overcome this limitation, a joint optimisation process was utilised to obtain the pose of the cameras on the far side. By using the points on the cheetah, and assuming the cheetah runs in a straight line, it was possible to set up an optimisation problem that jointly estimated the cheetah's 3D pose and the pose of one of the far side cameras. The stereo calibration of the near side cameras was enough to seed the optimisation for roughly 30 frames of an example trial. The optimisation cost function was defined to minimise reprojection error. This process was performed for each camera on the far side separately. The optimal solutions from the optimisation were found to be reasonable by comparing them with the rough measurements that were taken of the scene by the original researchers[27].

It is worth noting that this non-standard approach added uncertainty to the observations made by the far side cameras. Therefore, when using K-FTE for 3D reconstructions of the kinetic dataset, an extra 40 % (i.e. multiply observation errors by 0.6) of uncertainty was added to the observation errors from the far side cameras. This figure was found to provide good solutions during experimentation.

Evaluation

Root-mean-square error

The RMSE is a general error metric used to quantify the performance of a model by measuring the differences between estimated $\hat{\mathbf{x}}$ and ground truth $\mathbf{x}$ values in $\mathbb{R}^n$. It is defined as

$$e_{\text{RMSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}, \tag{20}$$

where x_i is the ith element of the vector $\mathbf{x}$.

2D metrics

To assess pose estimation methods in image space, two reprojection error metrics are used: ℓ^2 norm in pixels and percentage correct keypoints (PCK)[36]. The ℓ^2 norm of a vector $\mathbf{x} \in \mathbb{R}^n$ is defined as

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n}|x_i|^2}. \tag{21}$$

PCK is a metric that considers the percentage of estimated keypoints that are correctly located within a certain threshold distance of the ground truth keypoints. In this project, a matching threshold of the nose-to-eye(s) segment length was used, similar to what was done in a study by Mathis et al. [37]. This value is a normalised error metric that is calculated in pixels from the ground truth keypoints. Note that the PCK metric is 'strict', in that the segment length is small relative to the body and therefore considers a small region for a correct keypoint. In addition, this metric is only evaluated on images where the nose and one of the eyes are clearly visible. This allows the threshold to be determined.

Unless stated otherwise, the ℓ^2 norm is used to denote a 'general error' value for prediction performance.

3D metrics

To assess 3D pose estimation methods, a 3D position error metric is used: global mean position error (MPE) over the full trajectory. The MPE accounts for the absolute position in 3D space and is defined as

$$e_{\text{MPE}} = \frac{1}{NM}\sum_{k=1}^{N}\sum_{j=1}^{M}\|\mathbf{x}_{k,j} - \hat{\mathbf{x}}_{k,j}\|, \tag{22}$$

where N is the length (in frames) of the trajectory, M is the number of markers (placed on the joints of the cheetah), $\mathbf{x}_{k,j}$ is the position of marker j in $\mathbb{R}^3$ at time k, and $\hat{\mathbf{x}}_{k,j}$ is an estimate of $\mathbf{x}_{k,j}$.

Conclusions

In this paper, we present kinetic full trajectory estimation: a model-based nonlinear optimisation approach that incorporates estimation of the ground reaction forces to improve markerless 3D pose estimation in wild animals. Although this cannot be regarded as a remote force measurement method, the GRF profiles obtained using the sinusoidal model agree with those observed in this study from force plate data obtained from running cheetahs, as well as other studies of running quadrupeds. The approach produced promising results when tested on running cheetahs. In future, we hope that access to more complete truth data, and the incorporation of more advanced ground reaction force and skeletal models will improve the approach sufficiently to support its use in the study of fast, dynamic manoeuvres.

Data availibility

All data and code are available at https://github.com/African-Robotics-Unit/torque-estimationhttps://github.com/African-Robotics-Unit/torque-estimation.

Received: 10 December 2023; Accepted: 26 April 2024
Published online: 08 May 2024

References

1. Robertson, D. G. E., Caldwell, G. E., Hamill, J., Kamen, G. & Whittlesey, S. *Research Methods in Biomechanics* (Human Kinetics, 2013).
2. Daley, M. A. Non-steady locomotion. In *Understanding Mammalian Locomotion: Concepts and Applications*. 277–306 (2016).
3. Wilson, A. M. *et al.* Locomotion dynamics of hunting in wild cheetahs. *Nature* **498**, 185–189 (2013).
4. Mehta, D. *et al.* Monocular 3D human pose estimation in the wild using improved CNN supervision. In *2017 International Conference on 3D Vision (3DV)*. 506–516 (IEEE, 2017).
5. Martinez, J. Hossain, R. Romero, J. & Little, J. J. A simple yet effective baseline for 3d human pose estimation. In *Proceedings of the IEEE International Conference on Computer Vision*. 2640–2649 (2017).
6. Li, Z. *et al.* Estimating 3D motion and forces of human–object interactions from internet videos. *Int. J. Comput. Vis.* **130**, 363–383 (2022).
7. Shimada, S., Golyanik, V., Xu, W. & Theobalt, C. Physcap: Physically plausible monocular 3D motion capture in real time. *ACM Trans. Graph. (ToG)* **39**, 1–16 (2020).
8. Riemer, R. & Hsiao-Wecksler, E. T. Improving joint torque calculations: Optimization-based inverse dynamics to reduce the effect of motion errors. *J. Biomech.* **41**, 1503–1509 (2008).
9. Zell, P. Wandt, B. & Rosenhahn, B. Joint 3D human motion capture and physical analysis from monocular videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 17–26 (2017).
10. Felis, M. L., Mombaur, K. & Berthoz, A. An optimal control approach to reconstruct human gait dynamics from kinematic data. In *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*. 1044–1051 (IEEE, 2015).
11. Schemschat, R. M. *et al.* Joint torque analysis of push recovery motions during human walking. In *2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. 133–139 (IEEE, 2016).
12. Nath, T. *et al.* Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat. Protoc.* **14**, 2152–2176. https://doi.org/10.1038/s41596-019-0176-0 (2019).
13. Karashchuk, P. *et al.* Anipose: A toolkit for robust markerless 3D pose estimation. https://doi.org/10.1101/2020.05.26.117325 (2021).
14. Joska, D. *et al.* Acinoset: A 3D pose estimation dataset and baseline models for cheetahs in the wild. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. 13901–13908 (IEEE, 2021).
15. Zuffi, S. Kanazawa, A. Jacobs, D. W. & Black, M. J. 3D menagerie: Modeling the 3D shape and pose of animals. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 6365–6373 (2017).
16. Biggs, B. Roddick, T. Fitzgibbon, A. & Cipolla, R. Creatures great and small: Recovering the shape and motion of animals from video. In *Asian Conference on Computer Vision*. 3–19 (Springer, 2018).
17. Gosztolai, A. *et al.* Liftpose3d, a deep learning-based approach for transforming two-dimensional to three-dimensional poses in laboratory animals. *Nat. Methods* **18**, 975–981 (2021).
18. Williams, S., Usherwood, J., Jespers, K., Channon, A. & Wilson, A. Exploring the mechanical basis for acceleration: Pelvic limb locomotor function during accelerations in racing greyhounds (*Canis familiaris*). *J. Exp. Biol.* **212**, 550–565 (2009).
19. Ellis, R. G., Rankin, J. W. & Hutchinson, J. R. Limb kinematics, kinetics and muscle dynamics during the sit-to-stand transition in greyhounds. *Front. Bioeng. Biotechnol.* **6**, 162 (2018).
20. Harrison, S. M. *et al.* Forelimb muscle activity during equine locomotion. *J. Exp. Biol.* **215**, 2980–2991 (2012).
21. Muramatsu, N. da Silva, Z. Joska, D. Nicolls, F. & Patel, A. Improving 3D markerless pose estimation of animals in the wild using low-cost cameras. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 3770–3776 (IEEE, 2022).
22. Witte, T., Knill, K. & Wilson, A. Determination of peak vertical ground reaction force from duty factor in the horse (*Equus caballus*). *J. Exp. Biol.* **207**, 3639–3648 (2004).
23. Rempe, D. *et al.* Contact and human dynamics from monocular video. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*. 71–87 (Springer, 2020).
24. Hudson, P. E. *et al.* Functional anatomy of the cheetah (*Acinonyx jubatus*) hindlimb. *J. Anat.* **218**, 363–374 (2011).
25. Fukuhara, A., Gunji, M., Masuda, Y., Tadakuma, K. & Ishiguro, A. A bio-inspired quadruped robot exploiting flexible shoulder for stable and efficient walking. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 7832–7839 (IEEE, 2020).
26. Knemeyer, A., Shield, S. & Patel, A. Minor change, major gains: The effect of orientation formulation on solving time for multi-body trajectory optimization. *IEEE Robot. Autom. Lett.* **5**, 5331–5338 (2020).
27. Hudson, P. E., Corr, S. A. & Wilson, A. M. High speed galloping in the cheetah (*Acinonyx jubatus*) and the racing greyhound (*Canis familiaris*): Spatio-temporal and kinetic characteristics. *J. Exp. Biol.* **215**, 2425–2434 (2012).
28. Hudson, P. E. *et al.* Functional anatomy of the cheetah (*Acinonyx jubatus*) forelimb. *J. Anat.* **218**, 375–385 (2011).
29. Trinkle, J. C., Pang, J.-S., Sudarsky, S. & Lo, G. On dynamic multi-rigid-body contact problems with coulomb friction. *ZAMM-J. Appl. Math. Mech./Z. Angew. Math. Mech.* **77**, 267–279 (1997).
30. Nicholson, B., Lopez-Negrete, R. & Biegler, L. On-line state estimation of nonlinear dynamic systems with gross errors. *Comput. Chem. Eng.* **70**, 149–159. https://doi.org/10.1016/j.compchemeng.2013.11.018 (2014).
31. Brubaker, M. A., Sigal, L. & Fleet, D. J. Physics-based human motion modeling for people tracking: A short tutorial. *Image (Rochester, NY)* 1–48 (2009).
32. Wächter, A. & Biegler, L. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math. Program.* **106**, 25–57 https://doi.org/10.1007/s10107-004-0559-y (2006).
33. HSL. *A Collection of Fortran Codes for Large Scale Scientific Computation*. https://www.hsl.rl.ac.uk/ (2007).
34. Mathis, A. *et al.* Deeplabcut: Markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).
35. Bradski, G. The OpenCV library. *Dr. Dobb's J. Softw. Tools* (2000).
36. Andriluka, M. Pishchulin, L. Gehler, P. & Schiele, B. 2D human pose estimation: New benchmark and state of the art analysis. In *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*. 3686–3693 (2014).
37. Mathis, A. *et al.* Pretraining boosts out-of-domain robustness for pose estimation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 1859–1868 (2021).

Author contributions

Z.D.S. performed algorithm design, experiments and primary writing of the manuscript. S.S. assisted in algorithm design and review/revision of the manuscript. A.P. conceived the study and revised the manuscript. P.E.H. and A.M.W. collected the cheetah kinetic data, and F.N. provided technical inputs on the image processing and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Z.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.