ESTIMATING DYNAMICS ON-THE-FLY USING MONOCULAR VIDEO

Priyanshu Agarwal [†] **, Suren Kumar** [†]**, Jason J. Corso** [‡]**, and Venkat Krovi** [†] [∗]
Automation, Robotics, and Mechatronics Laboratory
Department of Mechanical and Aerospace Engineering [†]
Computer Science and Engineering Department [‡]
State University of New York at Buffalo
Buffalo, New York, 14260
Email: [priyansh, surenkum, jcorso, vkrovi]@buffalo.edu

ABSTRACT

We present an optimization framework to help estimate on-the-fly both the motion and physical parameters of an articulated multibody system using uncalibrated monocular image sequences. The algorithm takes video images of a physical system as input and estimates the motion together with the physical system parameters, given only the underlying articulated model topology. A valid initial pose of the system is found using a sequential optimization framework and used to bootstrap the successive pose estimation as well as estimation of physical system parameters (kinematic/geometric lengths as well as mass, inertia, damping coefficients). We also address the issue of robustly estimating a dynamically-equivalent system using partial state information (solely from noisy visual observations) and without explicit inertial parameter information. This framework results in a robust dynamically-equivalent system with good predictive capabilities when tested on a double pendulum system.

INTRODUCTION

System identification of articulated multibody systems has largely relied on existence of well-defined and persistent external excitation of the system [2]. Under the persistent excitation criterion, techniques for kinematic and inertial parameter estimation using joint sensing are well established [1]. However, estimation of inertial parameters of *articulated multi-body biological systems*, such as humans, animals or insects, has been possible only when augmented with specialized sensor systems. While kinematic identification methods for robotic systems using visual observation alone are well developed [8], they rely on calibrated camera setups. Inertial parameter estimation using visual observation alone with uncalibrated cameras still remains a challenge [9]. Nevertheless, accurate and efficient kinematic and inertial parameter estimation has many applications in telepresence [3], robot navigation [4, 5], vision-based humanoid control, multi-robot cooperation [6] and imitation based robot control [7].

Prior work has addressed the problem of articulated multi-body pose estimation using filtering techniques based on fiducials, interest-points, edges, and silhouettes. Broida *et al.* [10] used tracked feature point trajectories over multiple images and constructed a state space model incorporating both kinematic and structural states to estimate state vector using an iterated extended kalman filter (IKEF). Pathirana *et al.* [11] used robust extended Kalman filter (RKEF) and fused information from two cameras and ultrasonic sensor to estimate the kinematic states of mobile object from a moving camera. Extended Kalman Filter (EKF) has been also used by Azarbayejani *et al.* [12] to estimate motion and focal length of the camera.

In this paper, we address the issue of estimating

---

the kinematic and inertial parameters of a double-link pendulum using visual information from uncalibrated monocular video. We then seek to use these estimates to predict future poses of the double-link pendulum with no prior information about the physical system parameters or input. Well established system identification techniques fail to solve this problem due to: (i) partial state knowledge (only position information); (ii) noise due to imaging inaccuracies; (iii) unknown and insufficient input excitation (both structure and frequency); (iv) low sampling frequency; and (v) inherent dynamics nonlinearity. Further, most of the EKF/IKEF/RKEF based methods applied to motion estimation, rely on availability of point feature correspondences established over a long sequence of images, which is non-trivial for self-occluding articulated objects.

Hence, we leverage an optimization-based framework to solve the problem of parameter and input identification which resulted in a robust dynamically equivalent system. The presented methodology can use the learnt equivalent model to: (i) efficiently predict future pose; (ii) help with gap-filling when occulusions are present; and (iii) develop model-based control strategies.

PROBLEM STATEMENT

Given the information about the topology and type of body geometries in the system, estimate the camera location; orientation, size, mass, inertia of each body; and torque input/damping coefficient at each joint in the system using monocular images.

We use the optimization based framework for first estimating the camera parameters and then the initial pose of the multibody articulated system. The initial pose estimation technique is then extended to estimate the pose of the system in a video. Finally, using the estimated pose, the physical parameters of the system are estimated using dynamic-simulation-based optimization. Fig. 1 shows an outline of the overall approach.

CAMERA PARAMETER ESTIMATION

Camera Depth Estimation: We align the principle axes of the camera with the world coordinate system of the model. The camera depth (z coordinate) estimation is based on the fact that an actual body with proportional dimensions and similar orientation in space will roughly occupy a similar area in an actual image as that of the model in the synthetic image. We set up an optimization problem based on the difference in the silhouette area in
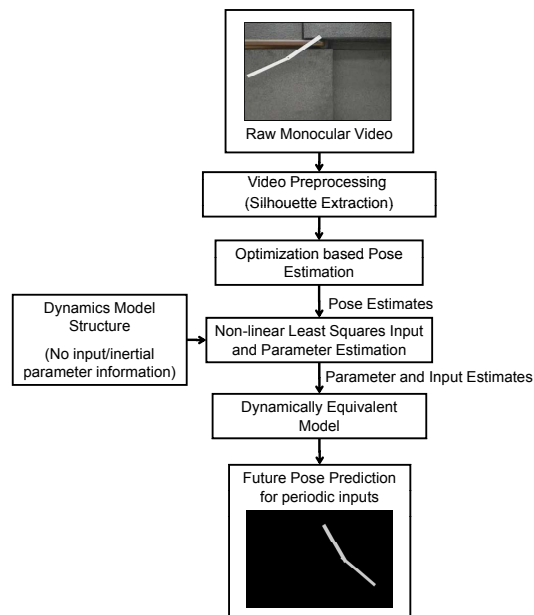

Figure 1: Overview of our approach for estimating a dynamically equivalent system.

the original image and the model generated image, and minimize the square of this difference (Eqn. (1)). We also setup an upper and a lower bound on the z coordinate to bound the search space. We specify the limit on z coordinate such that the model generates a reasonable area in the synthetic image.

$$\text{Minimize: } f_1(c_z) = (A_o - A_m(c_z))^2$$
$$\text{Subject to: } c_{z,l} \leq c_z \leq c_{z,u} \tag{1}$$

where c_z is the z coordinate of the camera in the model coordinate system, A_o and A_m is the silhouette area in the original and model generated image respectively.

Camera X,Y Coordinate Estimation: The estimation of the x, y coordinate is based on the fact that the centroid of the silhouette in the original image and the model generated image should roughly be the same for model with similar orientation. In order to estimate the x, y coordinate of the body we setup another optimization problem in which square of the distance between the centroid of the original silhouette and the model generate silhouette is minimized (Eqn. (2)). The x,y coordinates are constrained within certain bounds such that the model silhouette is within the synthetic image.

$$\text{Minimize: } f_2(c_x, c_y) = (x_{co} - x_{cm}(c_x, c_y))^2 + (y_{co} - y_{cm}(c_x, c_y))^2$$
$$\text{Subject to: } c_{x,l} \leq c_x \leq c_{x,u}, \ c_{y,l} \leq c_y \leq c_{y,u} \tag{2}$$

where (c_x, c_y) is the (x, y) coordinate of the camera in the model coordinate system, (x_{co}, y_{co}) and (x_{cm}, y_{cm}) is the centroid of the silhouette in the original and model generated image respectively.

INITIAL POSE ESTIMATION
Optimization Based Pose Estimation

The silhouette of the system is extracted from the actual image using standard image processing techniques (background subtraction and thresholding). An articulated VRML model of the system is positioned at the desired pose to capture a synthetic image (and subsequently a silhouette). These two silhouette images are then subtracted and area of absolute subtracted image is a measure of the extent of mismatch and used as the objective function for estimating the pose. The initial pose estimation relies on the fact that this objective function possesses a global minimum for correct choice of the camera and pose parameter estimates. The angular limits on the rotation of joints are imposed based on the physical constraints set by the system.

$$\text{Minimize: } f_3(\theta_1, \theta_2, ..., \theta_N) = \Sigma |I_a - I_m|$$
$$\text{Subject to: } \theta_{i,l} \leq \theta_i \leq \theta_{i,u}, i = 1, 2, ..., N \quad (3)$$

where the subscript i indicates i^{th} body in the system, N denotes the number of bodies in the articulated system, I_a and I_m denotes the actual and model generated silhouette image respectively.

Body Aspect Ratio Estimation

The aspect ratio estimation is based on the fact that once the pose of the system has been optimized the aspect ratio of the bodies in the system can be improved. In order to optimize for the aspect ratio of the bodies we minimize the objective as defined in Eqn. (4). However, now the constraints are imposed on the aspect ratio of each body.

$$\text{Minimize: } f_4(\alpha_1, \alpha_2, ..., \alpha_N) = \Sigma |I_a - I_m|$$
$$\text{Subject to: } \alpha_{i,l} \leq \alpha_i \leq \alpha_{i,u}, i = 1, 2, .., N \quad (4)$$

Optimization Framework

Each of the optimization subproblems in Eqn. (1)-(4) are highly coupled and cannot be solved independently. While a weighted/combined optimization problem may be created, it suffers from multiple local minima as well as sensitivity to weightage of each objective.
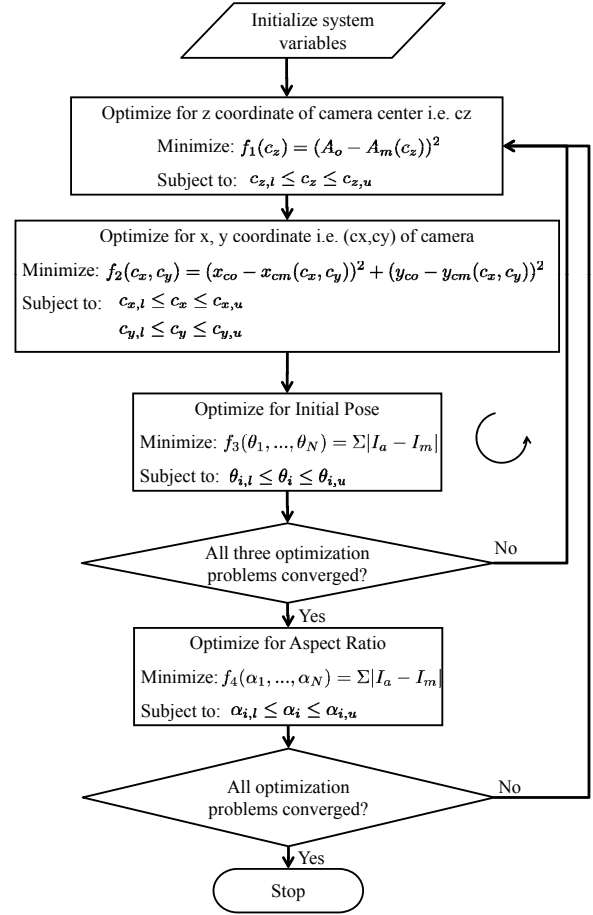

Figure 2: Summary of optimization framework implemented.

Hence, in lieu of this, we adopt a sequential optimization framework as shown in Fig. 2.

We optimize iteratively first for the camera parameters and then for the initial pose assuming an initial aspect ratio for the bodies in the system. Once this problem has converged, we optimize for the aspect ratio of the bodies in the system.

We use augmented Lagrangian method known as the method of multipliers [13] for solving the n-dimensional constrained optimization problem considering its advantage over interior/exterior penalty methods which are less robust due to sensitivity to penalty parameter chosen.

Numerical differentiation of the metrics is extremely slow due to evaluation of a computationally expensive objective several times for gradient evaluation [14]. In addition, accurate finite difference approximation is difficult due to inability to determine robust step sizes. To avoid such issues we solve the ND

unconstrained optimization subproblem using Powell's conjugate direction method [15]. For 1D optimization subproblem we employ Golden section with Swann's bounding [16]. We code all our optimization subroutines explicitly in MATLAB.

POSE ESTIMATION IN VIDEO

Once the camera position, initial pose of the system, and aspect ratio of the bodies in the system are determined, we move on to the problem of pose estimation in video. The problem of pose estimation in a video is handled by setting up an optimization problem using motion limits (move limits in optimization). The objective function considered here is defined in Eqn. (5).

$$\text{Minimize: } f_5(\theta_1, \theta_2, ..., \theta_N) = \Sigma |I_a - I_m|$$
$$\text{Subject to: } \theta_{i,l,t} \leq \theta_1 \leq \theta_{i,u,t} \tag{5}$$

The limits on the joint angles is considered as a fraction of the possible range of a joint angle such that every possible velocity can be accounted for. This can be made adaptive considering the instantaneous velocity of a joint to account for variation in the velocity which can further narrow down the search space for the optimization algorithm in each frame.

STATE FILTERING WITH $\alpha - \beta - \gamma$ Filter

Since, the state estimates are noisy due to noise in the imaging modality and the pose estimates, we first filter the estimated states using an $\alpha - \beta - \gamma$ filter. This is a suboptimal observer for data smoothing but nonetheless we use it due to its reduced computational cost (as compared to Kalman filter with commensurate filtering performance). We use the following filtering equations for smoothing the estimated states:

$$x_p(k+1) = x_s(k) + Tv_s(k) + \frac{1}{2}T^2 a_s(k)$$
$$v_p(k+1) = v_s(k) + Ta_s(k)$$
$$x_s(k) = x_p(k) + \alpha(x_o(k) - x_p(k)) \tag{6}$$
$$v_s(k) = v_p(k) + \frac{\beta}{T}(x_o(k) - x_p(k))$$
$$a_s(k) = a_p(k-1) + \frac{\gamma}{T^2}(x_o(k) - x_p(k))$$

where $x(k)$, $v(k)$, and $a(k)$ are the position, velocity and acceleration at k^{th} time step; the subscripts p, o, s stands for the predicted, observed and smoothed states after filtering respectively; α, β, γ are the filter gains; and T is the sampling time. We use the constraints (Eqn.7) on the gains obtained in [17] for optimal performance.

$$0 < \alpha \leq 1, \, 0 < \beta < 1, \, 0 < \gamma < \frac{4\alpha\beta}{2 - \alpha} \tag{7}$$
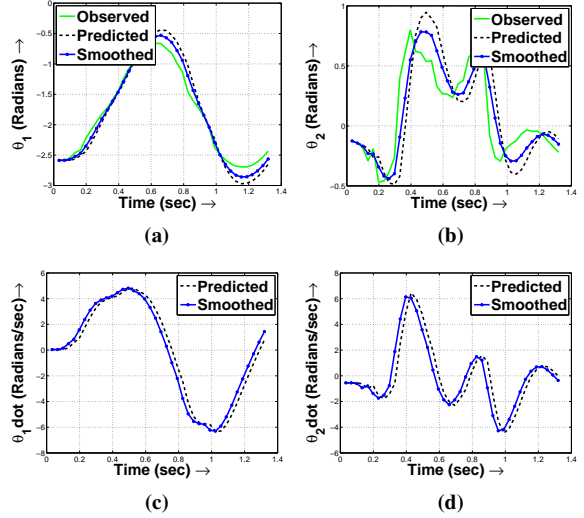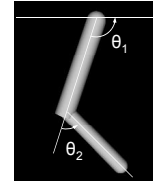

(a)   **(b)**

(c)   **(d)**

Figure 3: Filtered states of double pendulum for 40 frames, (a) θ_1 , (b) θ_2 , (c) $\dot{\theta}_1$, and (d) $\dot{\theta}_2$ ($\alpha = 0.4$, $\beta = 0.1, \gamma = 0.05$)

DYNAMIC MODELING OF 2-LINK SYSTEM

We derive the equation of the dynamics of the system incorporating the friction acting at the joints in the torque applied to the system.

(a)

Figure 4: Double pendulum model with states expressed in the counterclockwise direction as positive. θ_1 is referred from the right side of the horizontal axis and θ_2 is relative to the upper link.

$$\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix} \begin{bmatrix} \ddot{\theta}_1 \\ \ddot{\theta}_2 \end{bmatrix} + \begin{bmatrix} -h\dot{\theta}_2 & -h\dot{\theta}_1 - h\dot{\theta}_2 \\ h\dot{\theta}_1 & 0 \end{bmatrix} \begin{bmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix} \tag{8}$$

where,
$$H_{11} = m_1 l_{c1}^2 + I_1 + m_2[l_1^2 + l_{c2}^2 + 2l_1 l_{c2} \cos\theta_2] + I_2$$
$$H_{21} = m_2 l_{c2}^2 + I_2$$
$$H_{12} = H_{21} = m_2 l_1 l_{c2} \cos\theta_2 + m_2 l_{c2}^2 + I_2$$
$$h = m_2 l_1 l_{c2} \sin\theta_2$$
$$g_1 = m_1 l_{c1} g \cos\theta_1 + m_2 g[l_{c2} \cos\theta_1 + \theta_2 + l_1 \cos\theta_1]$$
$$g_2 = m_2 l_{c2} g \cos\theta_1 + \theta_2$$

We further express the masses and inertias in the system in terms of the basic system parameters assuming the density of the material and topological geometry of the links in the system to be known.

$$m_1 = \rho_1 l_1 w_1 h_1, m_2 = \rho_2 l_2 w_2 h_2, w_1 = \alpha_1 l_1, w_2 = \alpha_2 l_2$$

$$l_{c1} = \frac{l_1}{2}, l_{c2} = \frac{l_2}{2}, I_1 = \frac{m_1}{12}(l_1^2 + w_1^2), I_2 = \frac{m_2}{12}(l_2^2 + w_2^2)$$

The aspect ratio for the links (α_1, α_2) are obtained via optimization (Eqn. 4). So, the unknown parameters that need to be identified are the lengths (l_1, l_2) and the thickness (h_1, h_2). We consider various cases for estimating the parameters along with the input, relaxing the underlying assumptions gradually.

Case I: Known Friction Torque Structure

In this first case with known friction torque structure, we assume that the only torque acting on the joints is due to friction at the joints. We model this hinge friction as a rotational damping. So, the torque acting at the joints is given by

$$\tau_i = b_i \dot{\theta}_i, \ i = 1, 2 \tag{9}$$

where b_1, b_2 are the rotational viscous damping coefficients acting at joints. There are 6 unknown parameters that need to be identified in this case $(l_1, l_2, h_1, h_2, b_1, b_2)$.

Case II: Approximated Friction Torque Structure

In the second case, we consider that the output states of the double pendulum are periodic with the same period. This assumption can be justified by the fact that the friction acting at the joints is significantly low as there is no explicit damper/spring present at the joints. Furthermore, this case is applicable when the pendulum is externally excited maintaining its periodicity [2].

Considering θ_1 and θ_2 to be periodic with period being $\frac{2\pi}{\omega}$ a Fourier series expansion can be written as:

$$\theta_i = p_{10} + \sum_{j=1}^{M_i} (p_{ij} \cos j\omega t + p_{ij} \sin j\omega t), \ i = 1, 2 \tag{10}$$

Now, $\dot{\theta}_1$, $\dot{\theta}_2$, $\ddot{\theta}_1$, and $\ddot{\theta}_2$ can be expressed as

$$\dot{\theta}_i = j\omega \sum_{j=1}^{M_i} (-p_{ij} \sin j\omega t + q_{ij} \cos j\omega t), \ i = 1, 2$$

$$\ddot{\theta}_i = -(j\omega)^2 \sum_{j=1}^{M_i} (p_{ij} \cos j\omega t + q_{ij} \sin j\omega t), \ i = 1, 2$$

$$\tag{11}$$

Substituting Eqn. (10), (11) in Eqn. (8), we get

$$\widetilde{\tau}_{2 \times 1} = \widetilde{f}(\widetilde{\theta}, \dot{\widetilde{\theta}}, \ddot{\widetilde{\theta}}) \tag{12}$$

Since, the input variables to the function in Eqn. (8) are periodic with same time period, the input torque has to be periodic with the same time period. Periodic input torque can be decomposed into sum of sines and cosines using Fourier series to represent variation of torque over time (Eqn. (13)).

$$\tau_i = a_{i0} + \sum_{j=1}^{s_1} (a_{ij} \cos j\omega t + b_{ij} \sin j\omega t), \ i = 1, 2 \tag{13}$$

A pure mathematical representation for the number of terms (m) to represent torque is difficult to derive because true values of torques are not known. Variable number of terms can be selected to achieve desired level of accuracy. In fact, the selection of number of terms can be achieved as part of the optimization process - we however, explicitly specify the number of terms. There are $2s_1 + 2s_2 + 2$ unknown Fourier coefficients for the torques that together with the 4 unknown link lengths and thicknesses now need to be identified.

Case III: No Friction Torque Structure

In this more general case, we assume no prior information about the structure of the torque acting at the joints in the system. This case can handle scenarios when pendulum is subjected to arbitrary external forces/torques varying in time. We treat torques applied at the two joints at each time instant as the unknown parameters. So, for k time steps the number of unknown parameters due to torque are $2k$. The total number of unknown parameters for this case are $2k + 4$.

NON-LINEAR LEAST SQUARES PARAMETER ESTIMATION

Traditionally least squares parameter estimation employs low pass filtering before convolution [2, 18]. However, pose estimates obtained from raw images are very noisy making the derivative information obtained using an $\alpha - \beta - \gamma$ filter unreliable. Furthermore, the lack of knowledge of inputs coupled with the low visual sampling frequency prevents the use of convolution to obtain the linear system equations in terms of the unknown parameters.

Hence, we setup the parameter estimation problem using a non-linear least squares estimator noting

the peculiar challenges due to insufficient input excitation, poor state sampling frequency and missing input information.

The states considered are the pose of the bodies in the system i.e. $X = [\theta_1, \theta_2]$. The objective function is to minimize the difference between the dynamics-simulation-based states and the estimated states (Eqn. (14)).

$$\text{Minimize: } f(\Theta) = \|X_s - X_e\|$$
$$\text{Subject to: } \Theta_l \leq \Theta \leq \Theta_u \quad (14)$$

where X_s and X_e represent the simulated and estimated pose of the system respectively, and Θ represents the unknown parameters to be estimated.

We employ the active set algorithm with the Broyden-Fletcher-Goldfarb-Shanno method [19] for solving the non-linear parameter estimation problem. However, using these techniqes alone led to physically unrealizable systems during the optimization search leading to failure of the simulation. To avoid such issues, we penalize infeasible solutions and focus the optimization to search only in the feasible region.

RESULTS

The entire optimization framework was tested on a custom built double pendulum system consisting of two rods of uniform rectangular cross section made of aluminum (density $\approx 2700 kg/m^3$). The camera is placed perpendicular to the frame of the motion of the articulated system with axis aligned with the model z axis. The external forces acting on the system are: (i) gravity along negative y axis and (ii) hinge friction.
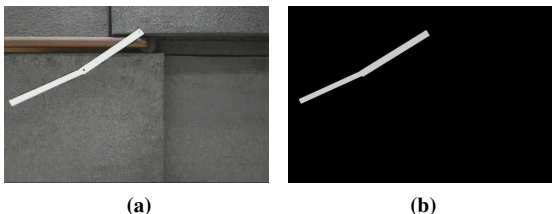

Figure 5: Actual and estimated pose of double pendulum.

For this experiment, the proposed optimization framework is applied on video obtained from an off-the-shelf camera (720x480 pixels @ 30 FPS) to estimate the states $X = [\theta_1, \theta_2]$ of the system. Constraints are applied in the optimization problem while solving the non-linear least squares estimation problem to ensure that obtained system is physically realizable by applying upper and lower bounds on all the system physical parameters.

As observed in the experiment , it is difficult to obtain precise physical dimensions because the resulting design space has multiple local minima, but an equivalent dynamical system that produces same output states with comparable physical dimensions is obtained.

From the first image, aspect ratio of both the links and location of hinge is determined. This information is propagated in subsequent images to reduce computational burden. We use the pose estimates from 40 frames ($\approx$ 1.32 secs of video) for parameter identification. Fig. 5 shows the actual and estimated model generated image of double pendulum.

Case I: Known Friction Torque Structure

Results show that when the filtered states are used for system identification the predictive performance of the system is degraded (Fig. 6). Furthermore, it turns out that the optimization based dynamically equivalent system estimation inherently filtered out the noise in the pose estimates thus, avoiding the need of an explicit filter.
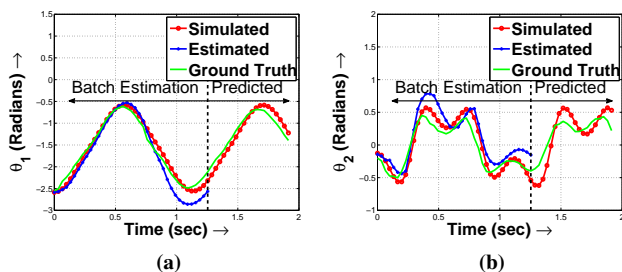

Figure 6: Simulated, estimated, and ground truth states when filtered states are used for parameter estimation.

Fig. 7 shows the simulated, estimated, and ground truth states for the pose of the double pendulum. It can be observed that the estimated pose is very close to the actual pose. The states obtained by simulating the dynamics of the pendulum are also found to be close to the actual states with accuracy improving with increasing number of frames considered during parameter estimation. In all the results below, parameter esimator were observed to be sensitive to initial conditions due to presence of local minima in the parameter space. Nevertheless, the equivalent dynamical system they form is able to simulate/predict model states satisfactorily (as shown in Fig. 7). This prediction can be further improved by

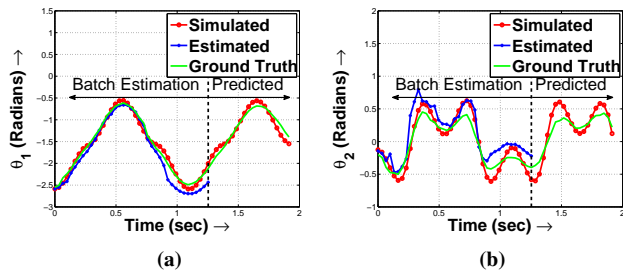considering more frames while estimating system parameters.

(a) **(b)**

Figure 7: Simulated, estimated and ground truth pose of double pendulum when friction torque was assumed to be known.

Case II: Approximated Friction Torque Structure

For approximating the friction torque we use a Fourier series expansion with 7 terms ($m = 7$) for torques at both the joints which results in total 36 unknows. Fig. 8 shows that the fourier series approximation of the torque also provides a means to estimate the system parameters. While the simulated system closely tracks the estimated pose, there is some error between both the observed and the ground truth. Further, in order to reliably predict the future poses this case requires pose estimates for multiple periods to estimate the period of the input torque.
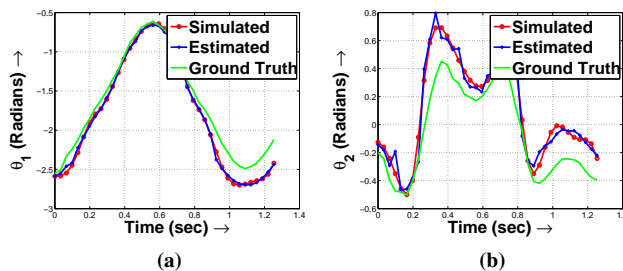

(a) **(b)**

Figure 8: Simulated, estimated and ground truth pose of double pendulum when friction torque structure was approximated using a fourier series expansion.

Case III: No Friction Torque Structure

In this case since we assume no prior information about the torques there are total 84 variables for 40 time steps. Fig. 9 shows that the simulated states exactly tracked the estimated states. However, since the estimated states are noisy there is deviation from the

ground truth. Also, in order to predict the future pose torque inputs need to be provided to the system since no structure is assumed for torque.
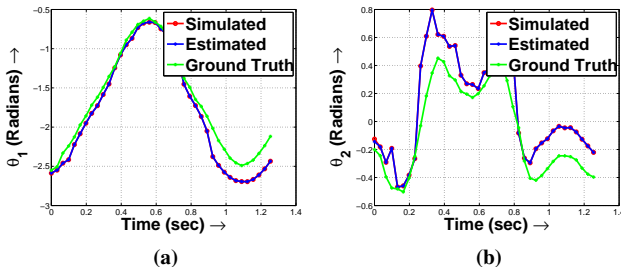

(a) **(b)**

Figure 9: Simulated, estimated and ground truth pose of double pendulum when no friction torque structure was assumed.

DISCUSSION

In this work, we proposed an optimization based framework for estimating the pose of an articulated multi-body system solely using monocular video. An optimization based inertial parameter estimation technique is then employed to realize a robust dynamically equivalent system on-the-fly using the estimated pose. Various cases are considered with gradually relaxing assumptions on the input torque structure. Results on a double pendulum system showed that the overall system is able to estimate a dynamically equivalent system robustly. The estimation of the parameter can be improved by using the information about both the joint angles and their derivatives during optimization. However, in the absence of appropriate covariance matrix on the pose estimation accuracy, weighting the cost induced due to angular position and that due to angular velocity is a challenge. Though the current methodology targets multi-body systems where the motion is restricted to a plane parallel to the camera plane, the technique is generalizable to generic multi-body systems using monocular SLAM [21] for extrinsic camera parameter estimation.

For the system under consideration any suitable initialization of the initial pose leads to a unique initial pose estimate. However, for a more complicated system intial pose estimates may affect subsequent pose estimates. In such a case an evolutionary optimization technique (e.g. genetic algorithm) can be employed to first narrow down the search space and then the convex optimization routines can be used to converge to the global minima. A similar approach has been employed for the problem of human pose estimation from monocular images [20]. There the authors tested the

algorithm on noisy silhouette sequences obtained from video data using background subtraction which resulted in good pose estimates. Also, the current optimization routines (coded in MATLAB) do not perform in real-time when executed on a standard desktop computer (Pentium(R) Dual-Core CPU E5200 @2.50 Ghz, 3.2 GB RAM). A faster implementation need to be coded in C which might result in real-time performance of the algorithm.

ACKNOWLEDGMENT

REFERENCES

[1] Siciliano, B., and Khatib, O., 2008. *Springer Handbook of Robotics*. Springer-Verlag NY Inc.

[2] Liang, Y., and Feeny, B., 2008. "Parametric identification of a chaotic base-excited double pendulum experiment". *Nonlinear Dynamics,* **52**(1), pp. 181–197.

[3] Sarkis, M., Diepold, K., and Huper, K., 2007. "Pose estimation of a moving humanoid using gauss-newton optimization on a manifold". In 7th IEEE-RAS International Conference on Humanoid Robots, pp. 228 –234.

[4] Gupta, O., and Jarvis, R., 2010. "Robust pose estimation and tracking system for a mobile robot using a panoramic camera". In IEEE Conference on Robotics Automation and Mechatronics, pp. 533–539.

[5] Bishop, B. E., and Spong, M. W., 2002. "Vision-based control of an air hockey playing robot". *Control Systems Magazine, IEEE,* **19**(3), pp. 23–32.

[6] Nogueira, M., Medeiros, A., Alsina, P., and Brazil, N., 2006. "Pose Estimation of a Humanoid Robot Using Images From a Mobile External Camera". In IFAC Workshop on Multivehicle Systems.

[7] Rubio, J., Zhou, C., and Hernández, F., 2005. "Vision-based walking parameter estimation for biped locomotion imitation". *Computational Intelligence and Bioinspired Systems*, pp. 677–684.

[8] Andreff, N., and Martinet, P., 2006. "Unifying kinematic modeling, identification, and control of a gough-stewart parallel robot into a vision-based framework". *IEEE Transactions on Robotics,* **22**(6), pp. 1077 –1086.

[9] Brubaker, M. A., Sigal, L., and Fleet, D. J., 2010. "Estimating contact dynamics". In 12th International Conference on Computer Vision, IEEE, pp. 2389–2396.

[10] Broida, T., Chandrashekhar, S., and Chellappa, R., 1990. "Recursive 3-d motion estimation from a monocular image sequence". *IEEE Transactions on Aerospace and Electronic Systems,* **26**(4), pp. 639–656.

[11] Pathirana, P., Lim, A., Savkin, A., and Hodgson, P., 2007. "Robust video/ultrasonic fusion-based estimation for automotive applications". *IEEE Transactions on Vehicular Technology,* **56**(4), pp. 1631–1639.

[12] Azarbayejani, A., and Pentland, A., 1995. "Recursive estimation of motion, structure, and focal length". *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **17**(6), pp. 562–575.

[13] Schuldt, S. B., 1975. "A method of multipliers for mathematical programming problems with equality and inequality constraints". *J. of Optimization Theory and Applications,* **17**(1), pp. 155–161.

[14] Bhat, K., Seitz, S., Popovic, J., and Khosla, P., 2002. "Computing the physical parameters of rigid-body motion from video". In *ECCV*, Vol. 2350 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, pp. 551–565.

[15] Powell, M. J. D., 1964. "An efficient method for finding the minimum of a function of several variables without calculating derivatives". *The Computer Journal,* **7**(2), pp. 155–162.

[16] Swann, W. H. "Report on the development of a new direct search method of optimization". *Research Note,* **64**(3).

[17] Tenne, D., and Singh, T., 2002. "Characterizing performance of α-β-γ filters". *IEEE Transactions on Aerospace and Electronic Systems,* **38**(3), pp. 1072–1087.

[18] Slotine, J., Li, W., et al., 1991. *Applied nonlinear control*, Vol. 461. Prentice hall, New Jersey.

[19] Nocedal, J., and Wright, S., 1999. *Numerical optimization*. Springer verlag.

[20] Agarwal, P., Corso, J., and Krovi, V., 2011. "An optimization based framework for human pose estimation in monocular images". In IEEE Conference on IROS (under Review).

[21] Chekhlov, D., Pupilli, M., Mayol-Cuevas, W., and Calway, A., 2006. "Real-time and robust monocular slam using predictive multi-resolution descriptors". *Advances in Visual Computing*, pp. 276–285.