Articulated Motion Estimation
From a Monocular Image Sequence
Using Spherical Tangent Bundles

Spyridon Leonardos

with Xiaowei Zhou and Kostas Daniilidis
Articulated Objects

- Why articulated objects?
  - Several objects of interest, e.g. humans and robots, are articulated.
Problem Statement

- Given the 2D perspective projections of joints over time, estimate recursively the corresponding 3D positions.
Problem Statement

- Given the 2D perspective projections of joints over time, estimate recursively the corresponding 3D positions.
Articulated Objects

- How do we describe an articulated object?
  - Graph (usually a tree) $G = (V, E)$
    - vertices ↔ joints
    - edges ↔ links between joints.
  - Articulation (constant length) constraint:
    \[
    \|x_i(t) - x_j(t)\|_2 = l_{ij} = \text{const.} \quad \forall (i, j) \in E
    \]
State space parametrization as a Riemannian manifold suggested by the articulation constraints.

Stochastic dynamical model evolving on this manifold.
  - Second order smoothness of links orientation.

Estimation using a Riemannian Extended Kalman Filter (REKF).
State Space

- Spherical tangent bundle $T\mathbb{S}^2$

  \[ T\mathbb{S}^2 = \{(x, \xi) \in \mathbb{R}^3 \times \mathbb{R}^3 : \|x\|_2 = 1, x^T \xi = 0\} \]

- State vector includes:
  - root position $x_1 \in \mathbb{R}^3$ and velocity $\dot{x}_1 \in \mathbb{R}^3$,
  - orientations of links $x_{ij} = (x_j - x_i)/\|x_j - x_i\|_2 \in \mathbb{S}^2$,
  - corresponding time derivatives $\dot{x}_{ij} \in T_{x_{ij}}\mathbb{S}^2$.

Articulation Manifold: $\mathbb{R}^6 \times (T\mathbb{S}^2)^{|E|}$
Contributions

- State space parametrization and dynamical model that incorporate articulation constraints.
- Implementation of a Riemannian Extended Kalman Filter (REKF).
- Perspective instead of orthographic camera model.
- Local observability analysis.
The tangent bundle of the manifold $\mathcal{M}$ is the set

$$\mathcal{T}\mathcal{M} = \{(x, \xi) : x \in \mathcal{M}, \xi \in T_x\mathcal{M}\}$$

and is a manifold of double dimension.

$$T\mathbb{S}^2 = \{(x, \xi) \in \mathbb{R}^3 \times \mathbb{R}^3 : \|x\|_2 = 1, x^T \xi = 0\}$$
Geometry of Tangent Bundles
Horizontal and Vertical Spaces

- Horizontal curves: \((x(t), \xi(t)) = (x(t), P_x^{t\leftarrow 0}\xi(0))\).
- Vertical curves: \((x(t), \xi(t)) = (x(0), \xi(t)), \xi(t) \in T_{x(0)}M\).
- Decomposition of tangent space in vertical and horizontal spaces:
  \[T_{(x,\xi)}TM = \mathcal{V}_{(x,\xi)} \oplus \mathcal{H}_{(x,\xi)}\]
- Each tangent vector in \(T_{(x,\xi)}TM\) can be uniquely decomposed as \(\zeta^h + \eta^v\) and thus, it is represented by the pair \((\zeta, \eta) \in T_xM \times T_xM\).
- Metric for \(TM\) from the metric of \(M\), e.g. the Sasaki metric.
Dynamical Model

- Dynamical model on $T\mathbb{S}^2$ for each link $(i, j) \in \mathcal{E}$

\[
(x_{ij}(t + 1), \dot{x}_{ij}(t + 1)) = \exp(x_{ij}(t), \dot{x}_{ij}(t)) (\dot{x}_{ij}(t), w_{ij}(t))
\] (1)

where $\exp$ denotes the exponential map of $T\mathbb{S}^2$, $(x_{ij}, \dot{x}_{ij}) \in T\mathbb{S}^2$ and $w_{ij}(t) \in T_{x_{ij}(t)}$ is zero-mean isotropic Gaussian noise (process noise).

- The exponential map of $T\mathbb{S}^2$ does not have, in general, a closed form expression.

- Perspective measurements of joints:

\[
y_i(t) = \pi(x_i(t)) + \eta_i(t), \quad \forall i \in \mathcal{V}
\] (2)
In the absence of process noise:

\[ x_{ij}(t + 1) = \exp_{x_{ij}(t)} \dot{x}_{ij}(t) \]
\[ \dot{x}_{ij}(t + 1) = P^{-1}_{x_{ij}} \dot{x}_{ij}(t) \]

\( \forall (i, j) \in E \), (3)

Example of an articulation chain with three joints. Intuitively, when \( w_{ij}(t) = 0 \), \( x_{12} \) travels on a geodesic of \( S^2 \) and \( \dot{x}_{12}(t) \) is a parallel vector field along \( x_{12} \).
Proposition 1

The differential equations of a geodesic curve on $TS^2$, equipped with the Sasaki metric, emanating from $(x, \xi) \in TS^2$ in the direction of $(\zeta, \eta) \in T_{(x,\xi)}TS^2$ read

\[ \nabla_\zeta \zeta = -R(\xi, \eta)\zeta \] (4)
\[ \nabla_\zeta \eta = 0 \] (5)

where $\nabla$ is the Levi-Civita connection and $R$ is the curvature tensor of $S^2$. The differential equations for parallel transport of a tangent vector $(\mu, \nu)$ along the same geodesic read

\[ \nabla_\zeta \mu = -\frac{1}{2}R(\xi, \nu)\zeta - \frac{1}{2}R(\xi, \eta)\mu \] (6)
\[ \nabla_\zeta \nu = \frac{1}{2}R(\zeta, \mu)\xi \] (7)
Specifcally, $\exp_{(x_0,\xi_0)}(\zeta_0, \eta_0)$ and the parallel transport of $(\mu_0, \nu_0) \in T_{(x_0,\xi_0)}S^2$ from $(x_0, \xi_0)$ to $\exp_{(x_0,\xi_0)}(\zeta_0, \eta_0)$ along the corresponding geodesic are computed using the following iterative scheme for $k = 0, 1, \ldots, N - 1$

\[
\begin{align*}
    x_{k+1} &= \exp_{x_k} \epsilon \zeta_k \\
    \xi_{k+1} &= P^{k+1 \leftarrow k}_x \xi_k + \epsilon \eta_k \\
    \zeta_{k+1} &= P^{k+1 \leftarrow k}_x \zeta_k - \epsilon R(\xi_k, \eta_k) \zeta_k \\
    \eta_{k+1} &= P^{k+1 \leftarrow k}_x \eta_k \\
    \mu_{k+1} &= P^{k+1 \leftarrow k}_x \mu_k - \frac{\epsilon}{2} (R(\xi_k, \nu_k) \zeta_k + R(\xi_k, \eta_k) \mu_k) \\
    \nu_{k+1} &= P^{k+1 \leftarrow k}_x \nu_k + \frac{\epsilon}{2} R(\zeta_k, \mu_k) \xi_k
\end{align*}
\]

where $\epsilon = 1/N$. 
Local observability

- **Local observability** at a configuration \( (x_{12}(0), \dot{x}_{12}(0)) \in T\mathbb{S}^2 \) when the ray from the camera center to the second joint is not tangential to the sphere centered at the first joint and having radius \( l_{12} \) (case (a)).

- If the ray is tangential, then \( x_{12}(0) \) has a unique global solution but the initial velocity \( \dot{x}_{12}(0) \) is not uniquely determined from \( y_2(0) = \pi(x_2(0)) \) and \( \dot{y}_2(0) \) (case (b)).
Dynamical system of interest evolves on a Riemannian manifold and is of the form

\begin{align}
    x(t + 1) &= \exp_{x(t)} \left( \log_{x(t)}(f(x(t))) + w(t) \right) \\
    y(t) &= h(x(t)) + \eta(t)
\end{align}

(8)
Riemannian Extended Kalman Filter

- **Linearization:**

\[
\begin{aligned}
F(t) &= Df(\hat{x}(t|t)) \\
C(t) &= Dh(\hat{x}(t|t-1))
\end{aligned}
\]  

- **Update:**

\[
\begin{aligned}
\hat{x}(t|t) &= \exp_{\hat{x}(t|t-1)} L(t)(y(t) - \hat{y}(t)) \\
L(t) &= \Sigma(t|t-1)C(t)^T (C(t)\Sigma(t|t-1)C(t)^T + \Sigma_\eta(t))^{-1} \\
\Sigma(t|t) &= P_{\hat{x}(t|\cdot)}^{t \leftarrow t-1}(I - L(t)C(t))\Sigma(t|t-1)
\end{aligned}
\]  

- **Prediction:**

\[
\begin{aligned}
\hat{x}(t+1|t) &= f(\hat{x}(t|t)) \\
\Sigma(t+1|t) &= F(t)\Sigma(t|t)F(t)^T + P_{\hat{x}(\cdot|t)}^{t+1 \leftarrow t}\Sigma_w(t)
\end{aligned}
\]
Reconstruction Example
Experiments

- 3D of Torso markers is provided to the algorithms.
- Error metric: 3D error for the non-rigid part of the human body.
- Actions include walking, dancing, hand signals.
Experimental Conclusions

- Considering all articulation constraints at once improves performance.
- Fast camera motion is not crucial in the proposed method.
- Proposed method depends on initialization.
References


