Efficient Articulated Trajectory Reconstruction Using Dynamic Programming and Filters

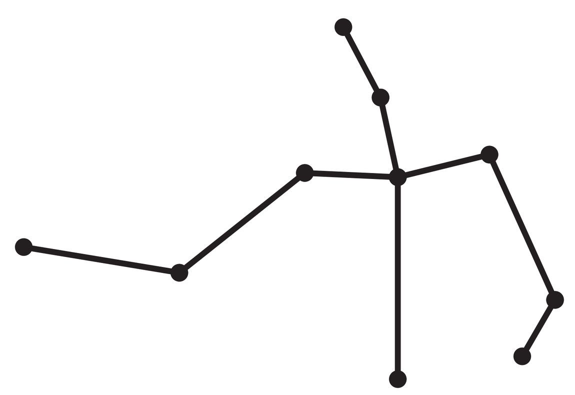
Jack Valmadre[1,3] Yingying Zhu[2,3] Sridha Sridharan[1] Simon Lucey[1,3]

Queensland University of Technology[1] University of Queensland[2] CSIRO[3] Brisbane, Australia

Problem

An articulated structure is a graph $G = (V, E)$ with edge labels $\ell : E \to \mathcal{R}$ such that

$$\{i, j\} \in E \quad \Rightarrow \quad \|\mathbf{x}_i - \mathbf{x}_j\| = \ell_{ij}$$



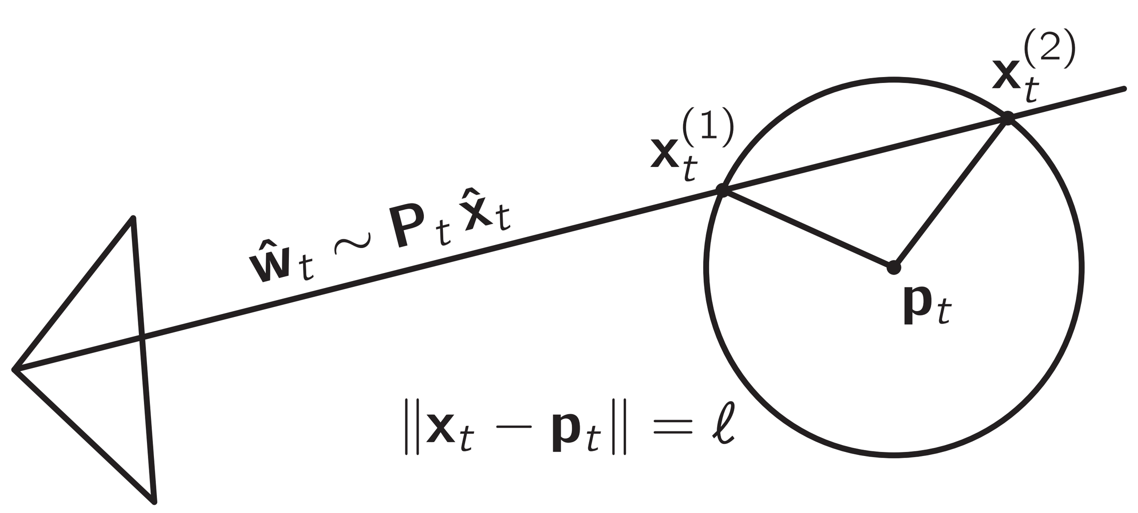
Given an articulated tree structure (G, ℓ) with, for $t = 1, \dots, n$,

- root node position $\mathbf{x}_{t1} \in \mathcal{R}^3$,
- projections of other nodes $\mathbf{w}_{ti} \in \mathcal{R}^2$ for $i = 2, \dots, |V|$,
- perspective camera $\mathbf{P}_t \in \mathcal{R}^{3 \times 4}$,

find the smoothest trajectories which satisfy projections

Single articulated trajectory

Consider single trajectory $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ with fixed distance ℓ from a known parent trajectory $\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_n)$



Each point has exactly two solutions

$$\left\{\mathbf{x}_t \,:\, \hat{\mathbf{w}}_t \sim \mathbf{P}_t \hat{\mathbf{x}}_t,\ \|\mathbf{x}_t - \mathbf{p}_t\| = \ell\right\} = \left\{\mathbf{x}_t^{(1)}, \mathbf{x}_t^{(2)}\right\}$$

2^n possible trajectories parameterised by binary variables

$$\mathbf{x}_t = (1 - y_t)\,\mathbf{x}_t^{(1)} + y_t\,\mathbf{x}_t^{(2)}, \qquad y_t \in \{0, 1\}$$

$$\mathbf{x} = \mathbf{A}\mathbf{y} + \mathbf{b}, \qquad \mathbf{y} \in \{0, 1\}^n$$

Branch and bound solution (Park and Sheikh, 2011)

Minimise component orthogonal to a low-frequency trajectory basis

$$\mathbf{y}^* = \arg\min_{\mathbf{y}} \left\|(\mathbf{I} - \mathbf{\Theta}\mathbf{\Theta}^T)(\mathbf{A}\mathbf{y} + \mathbf{b})\right\| \quad \text{subject to } \mathbf{y} \in \{0, 1\}^n$$

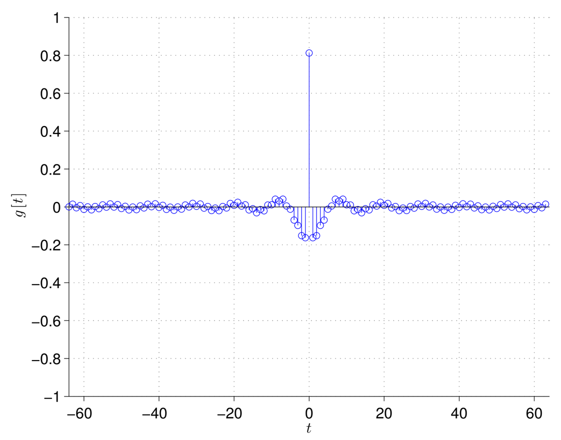
Branch and bound with the convex relaxation $0 \le \mathbf{y} \le 1$, worst case $\mathcal{O}(2^n)$ time

DCT basis and convolution (Valmadre and Lucey, 2012)

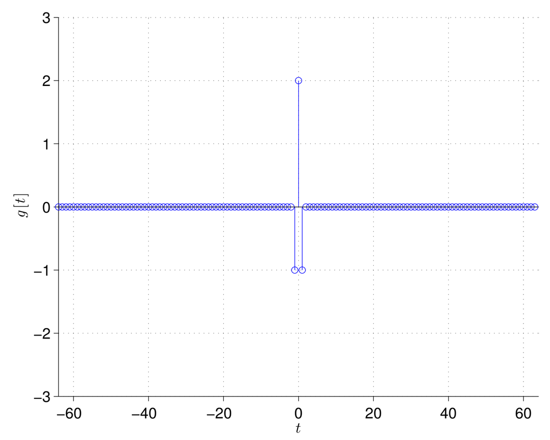
Low-frequency DCT basis $\Leftrightarrow$ symmetric convolution with a high-pass "brick wall" filter

$$\left\|(\mathbf{I} - \mathbf{\Theta}\mathbf{\Theta}^T)\mathbf{x}\right\| = \left\|\operatorname{diag}(0, \dots, 0, 1, \dots 1)\,\mathcal{C}\{\mathbf{x}\}\right\| = \|\mathbf{h} * \mathbf{x}\|$$

$\mathcal{C}\{\cdot\}$ is a Discrete Cosine Transform and $\mathbf{h}$ is the impulse response



$\mathbf{h} = \mathcal{C}^{-1}\{(0, \dots, 0, 1, \dots, 1)\}$



$\mathbf{h} = (-1, 2, -1)$

Resulting filter $\sim$ compact support

Simple filters have physical motivation:

- $(-1, 1)$ minimises kinetic energy,
- $(-1, 2, -1)$ assumes Gaussian i.i.d. forces

Dynamic programming (Felzenszwalb and Zabih, 2011)

Objectives of the form

$$f(\mathbf{y}) = \sum_{t=1}^{n-m+1} g_t(y_t, \dots, y_{t+m-1}) \qquad \mathbf{y} \in \{1, \dots, k\}^n$$

can be minimised in $\mathcal{O}(nk^m)$ time and $\mathcal{O}(nk^m)$ memory

Based on recursive partial solution

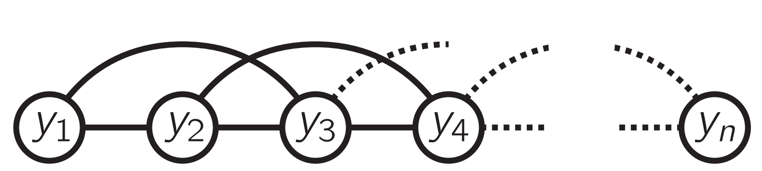
$$\begin{aligned} f_i^*(y_{i+1}, \dots, y_{i+m-1}) &= \min_{y_1, \dots, y_i} \sum_{t=1}^{i} g_t(y_t, \dots, y_{t+m-1}) \\ &= \min_{y_i} \left\{ g_i(y_i, \dots, y_{i+m-1}) + f_{i-1}^*(y_i, \dots, y_{i+m-2}) \right\} \end{aligned}$$

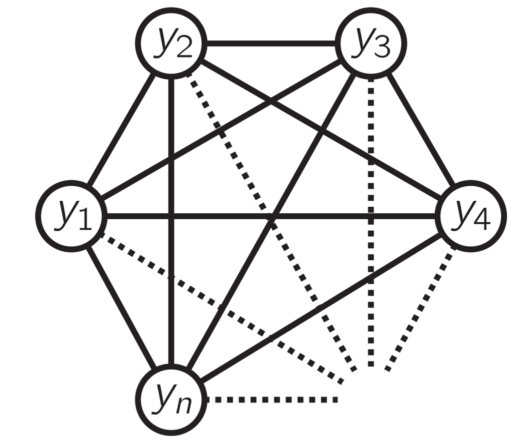
Our solution

Minimise response of feasible trajectory to compact filters ($k = 2$, $m = 3$)

$\Rightarrow \mathcal{O}(n)$ instead of $\mathcal{O}(2^n)$

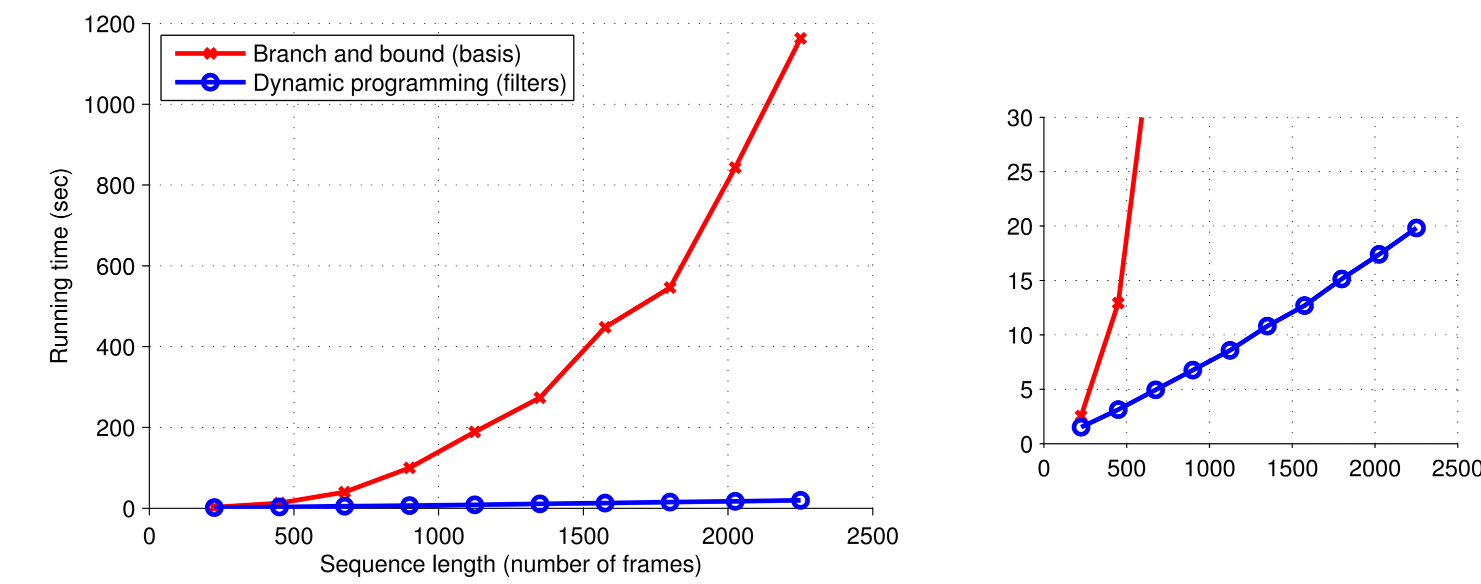
Takes advantage of the inherent locality, whereas a basis requires inner products



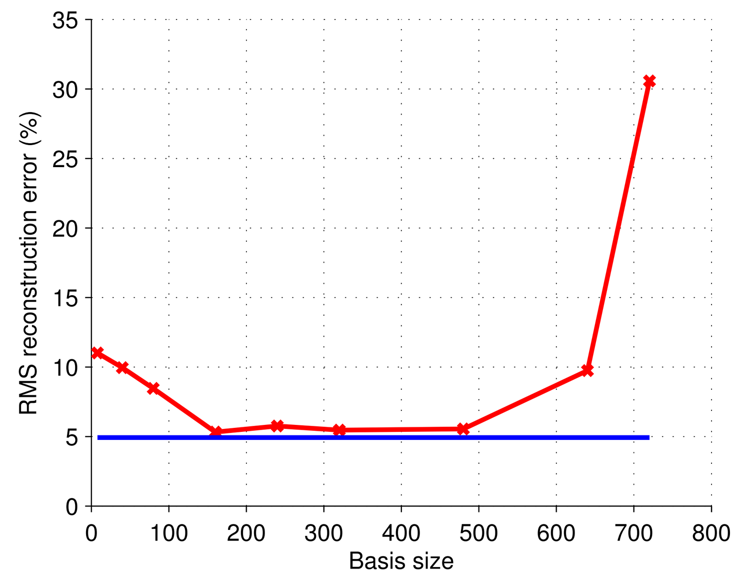


Results

Running time (synthetic sequences)



Reconstruction error (synthetic sequences)



Real sequences: see videos

Calibration-less reconstruction

Assume that

- smoothness is preserved in the camera reference frame,
- the root trajectory is smooth,
- orthographic projection is a good approximation,
- each link is observed near parallel to the image plane,

and then take

- all cameras to be identity $\mathbf{P}_t = [\mathbf{I}\ 0]$,
- the root node depth to be zero,
- the length of each edge to be its maximum projection $\ell_{ij} = \max_t \|\mathbf{w}_{ti} - \mathbf{w}_{tj}\|$

References

- H. S. Park and Y. A. Sheikh, *3D reconstruction of a smooth articulated trajectory from a monocular image sequence*, ICCV 2011
- J. Valmadre and S. Lucey, *General trajectory prior for non-rigid reconstruction*, CVPR 2012
- P. F. Felzenszwalb and R. Zabih, *Dynamic programming and graph algorithms in computer vision*, PAMI 2011

{jack.valmadre,yingying.zhu,simon.lucey}@csiro.au