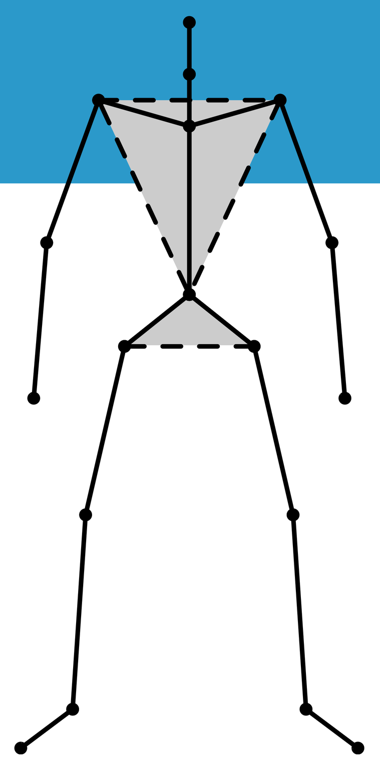
Graph Rigidity for Near-Coplanar Structure from Motion

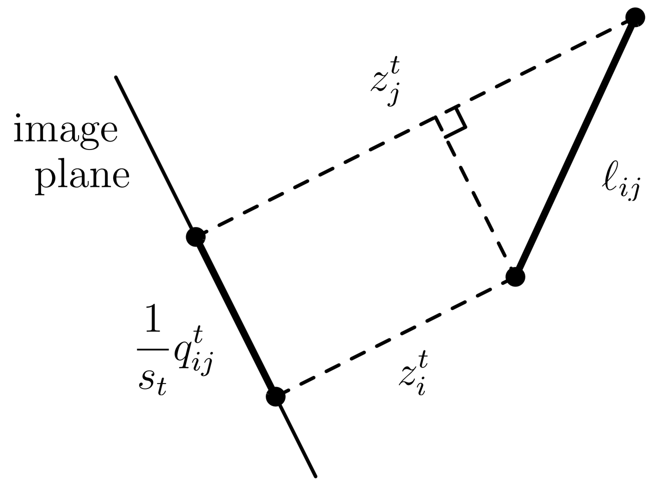
Jack Valmadre[1,2] Sridha Sridharan[2] Simon Lucey[1,2]

[1]CSIRO ICT Centre, Australia [2]Queensland University of Technology, Australia



A four-point rigid torso can be used to estimate human pose [1]. However, since it is almost coplanar, factorisation using the SVD will not give an affine reconstruction.

Weak perspective projection and Pythagoras' theorem provide a system of equations.



$$\ell_{ij}^2 = (q_{ij}^t)^2 s_t^{-2} + (z_i^t - z_j^t)^2$$

$$\underset{\boldsymbol{\ell}, \mathbf{s}, \mathbf{z}_{1..F}}{\text{minimise}} \sum_{t=1}^{F} \sum_{(i,j) \in \mathcal{E}} \left[\mathbf{b}_{ij}^T \boldsymbol{\ell} - (\mathbf{c}_{ij}^t)^T \mathbf{s} - \mathbf{z}_t^T \mathbf{D}_{ij} \mathbf{z}_t \right]^2$$
$$\text{subject to } \mathbf{1}^T \boldsymbol{\ell} = 1$$

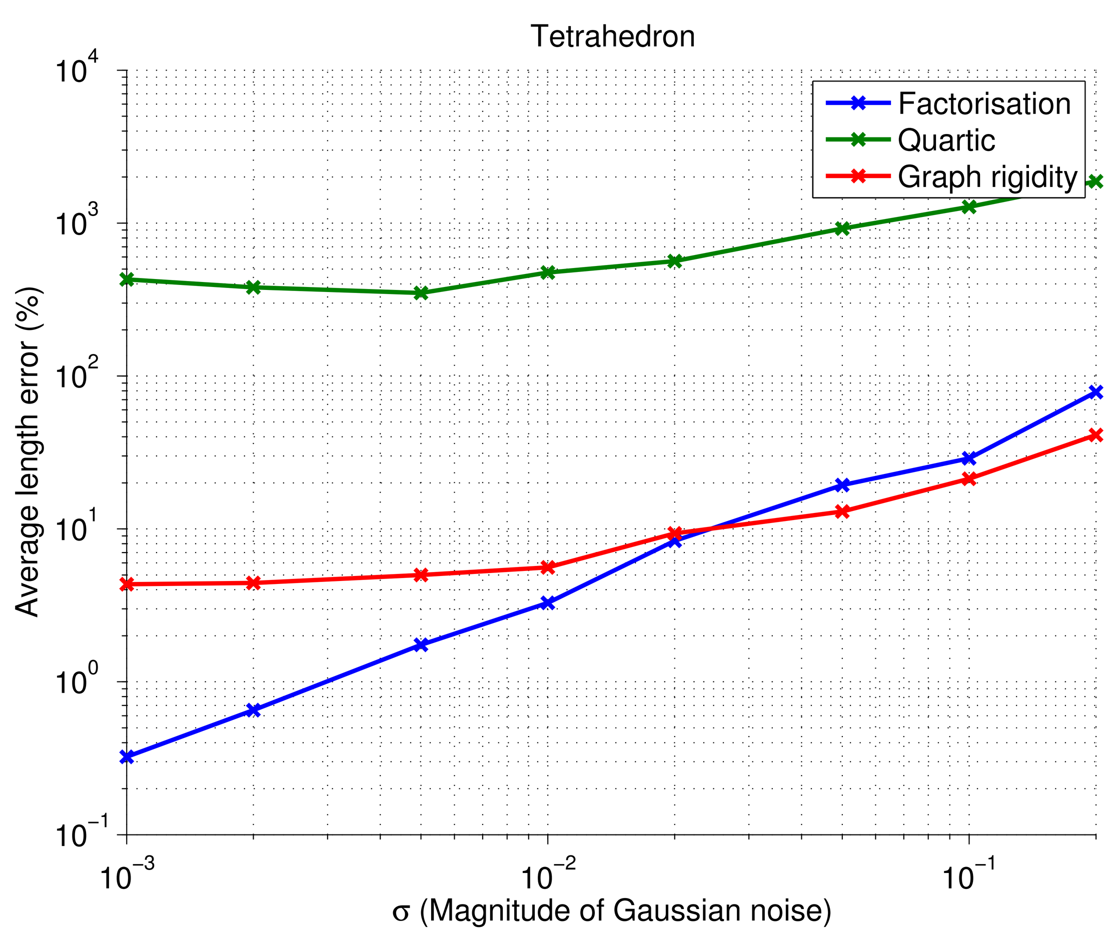
Non-convex objective.

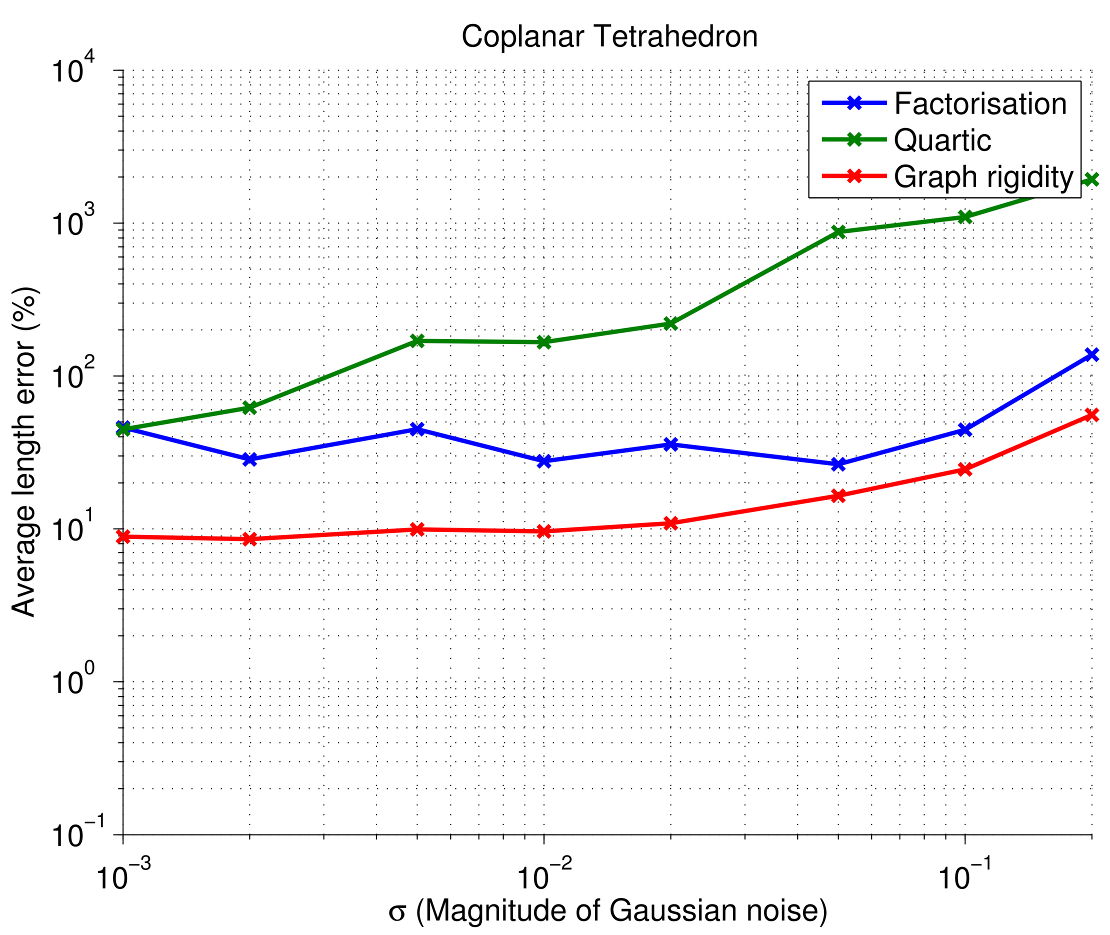
Introducing a relaxation and using the nuclear norm heursitic to minimise rank [2], a convex objective can be obtained. Results in situations with a small number of coplanar points are more accurate.

$$\mathbf{Z}_t \approx \mathbf{z}_t \mathbf{z}_t^T$$

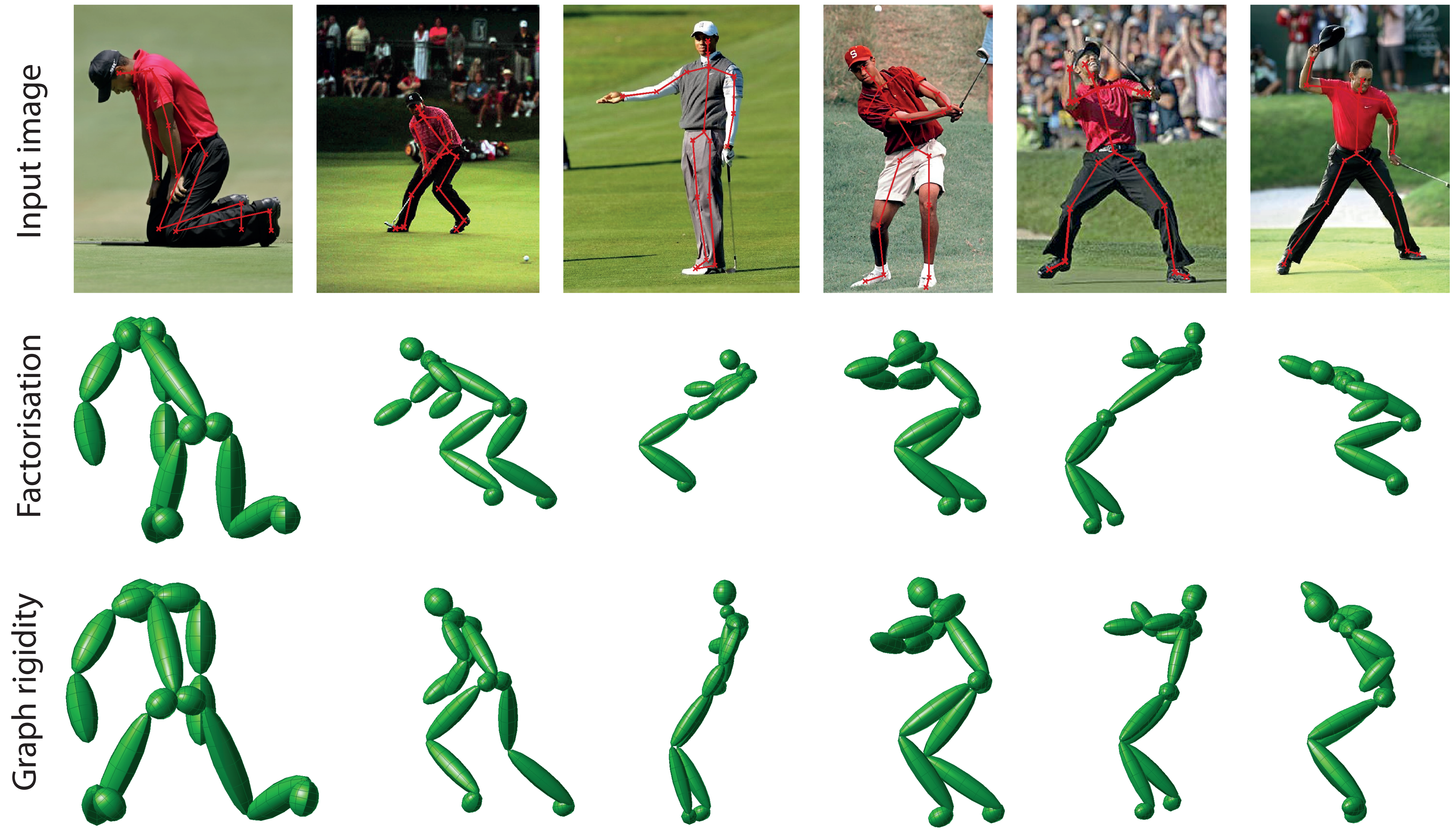
$$\mathbf{A} \succeq 0 \;\Rightarrow\; \|\mathbf{A}\|_* = \text{tr}(\mathbf{A}) = \|\boldsymbol{\lambda}(\mathbf{A})\|_1$$

$$\underset{\boldsymbol{\ell}, \mathbf{s}, \mathbf{Z}_{1..F}}{\text{minimise}} \sum_{t=1}^{F} \text{tr}(\mathbf{Z}_t)$$
$$\text{subject to } \mathbf{b}_{ij}^T \boldsymbol{\ell} - (\mathbf{c}_{ij}^t)^T \mathbf{s} - \text{tr}(\mathbf{D}_{ij} \mathbf{Z}_t) = 0 \quad \forall t, (i,j)$$
$$\mathbf{Z}_t \succeq 0 \quad \forall t, \qquad \mathbf{1}^T \boldsymbol{\ell} = 1, \qquad \boldsymbol{\ell}, \mathbf{s} \succeq 0$$

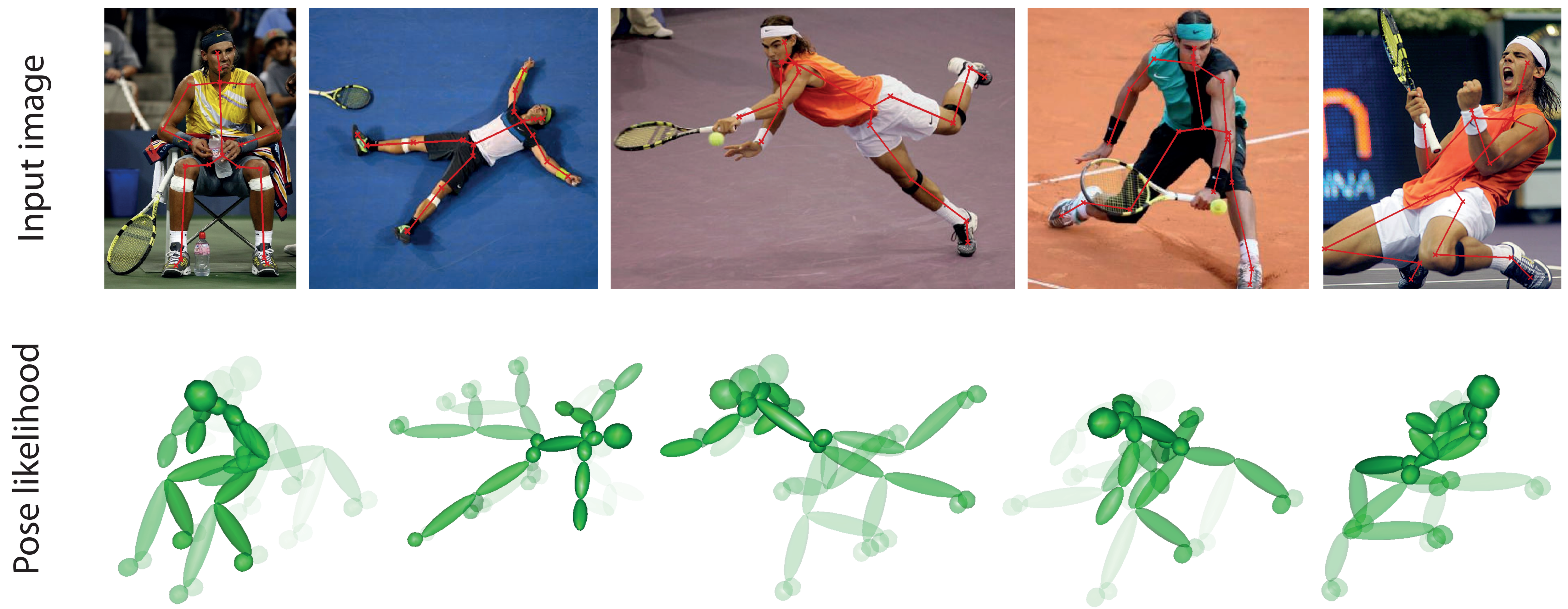




Rigid pose estimate critically affects bone length estimates through scale.



Pose likelihood learned from motion capture to reduce remaining ambiguity.



[1] X. Wei and J. Chai, *Modelling 3D human poses from uncalibrated monocular images*. ICCV, 2009.

[2] M. Fazel, *Matrix rank minimization with applications*. PhD thesis, Stanford University, 2002.