# Deterministic 3D Human Pose Estimation using Rigid Structure

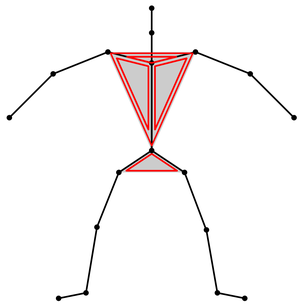
Jack Valmadre, University of Queensland, Australia
Simon Lucey, CSIRO, Australia

## The Problem

Given $F$ uncalibrated, monocular images of a person in different poses, is it possible to recover their body's 3D configuration from labelled joint positions, under an assumption of weak-perspective projection?

## Rigid Sub-Structure in Human Bodies

Wei and Chai (2009) proposed the novel constraint that the torso and hip are rigid for solving non-rigid structure from motion for human bodies.



Their model of the human skeleton consists of $B = 17$ bones, $R = 4$ rigid triangles and $M = 7$ symmetry constraints.

## Previous Solution

Applying Pythagoras' theorem to the projected image of each bone,

$$E_p(\boldsymbol{\ell}, \mathbf{s}, \mathbf{z}) = \sum_{t=1}^{F} \sum_{i=1}^{B} \left[ l_i^2 - (q_i^t)^2 s_t^{-2} - (z_i^t)^2 \right]^2$$

$l_i$ = length of bone $i$, $z_i^t$ = depth of bone $i$ out of frame $t$,
$s_t$ = scale of frame $t$, $q_i^t$ = scaled projection of bone $i$ in frame $t$.

Rigidity constraints are enforced by adding an extra hidden bone, which closes a pair of bones to form a rigid triangle.

$$E_r(\mathbf{e}, \mathbf{s}, \mathbf{z}) = \sum_{t=1}^{F} \sum_{(i,j) \in \mathbb{E}} \left\{ \left[ e_{ij}^2 - (q_{ij}^t)^2 s_t^{-2} - (z_i^t)^2 - (z_j^t)^2 \right]^2 - 4 (z_i^t)^2 (z_j^t)^2 \right\}^2$$

$e_{ij}$ = length of "extra" bone connecting bones $i$ and $j$,
$q_{ij}^t$ = scaled projection of "extra" bone in frame $t$,
$\mathbb{E}$ = set of triangle pairs $(i, j)$.

The overall objective function is quartic (in terms of the squared variables).

$$\underset{\mathbf{x}}{\text{minimize}} \;\; E_p(\boldsymbol{\ell}, \mathbf{s}, \mathbf{z}) + \lambda E_r(\mathbf{e}, \mathbf{s}, \mathbf{z})$$
$$\text{subject to } \mathbf{x} \succeq 0, \; l_1 = 1$$

$$\mathbf{x} = \begin{bmatrix} \boldsymbol{\ell} \\ \mathbf{e} \\ \mathbf{s} \\ \mathbf{z} \end{bmatrix}, \quad \boldsymbol{\ell} = \begin{bmatrix} l_1^2 \\ \vdots \\ l_B^2 \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} e_{12}^2 \\ \vdots \\ e_{ij}^2 \end{bmatrix}, \quad \mathbf{s} = \begin{bmatrix} s_1^{-2} \\ \vdots \\ s_F^{-2} \end{bmatrix}, \quad \mathbf{z} = \operatorname{vec} \left( \begin{bmatrix} (z_1^1)^2 & \cdots & (z_1^F)^2 \\ \vdots & \ddots & \vdots \\ (z_B^1)^2 & \cdots & (z_B^F)^2 \end{bmatrix} \right)$$

## Constraints versus Variables

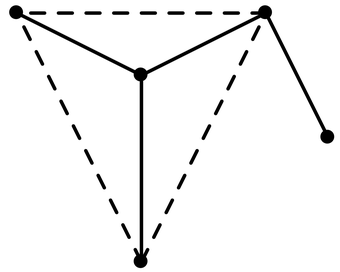
Comparing the number of unknowns and constraints, the condition on $F$ is

$$F \geq 1 + \frac{B - M}{R - 1}$$

For our skeleton model, this implies that it is always possible to find bone lengths, bone depths and camera parameters given $F \geq 5$.

## The Free Bone Counter-Example

Canonical structure from motion (Tomasi and Kanade 1992) gives a deterministic solution for a rigid tetrahedron.



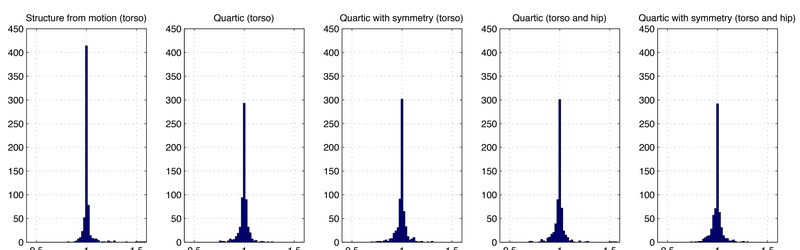
No number of frames guarantees we can find the length of the free bone. For known camera parameters, each free bone emerges as an independent, under-ranked system of linear equations.

## Solution using Structure from Motion

The advantages of using the canonical SFM method are:

– it has a deterministic, linear least-squares solution, and

– it solves for the direction of the depth across each bone, not just magnitude.



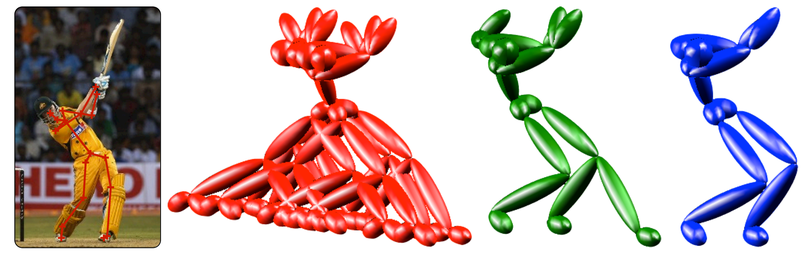

Empirical experiments on the CMU MoCap database (http://mocap.cs.cmu.edu/).

The length of each free bone is estimated by its longest observed projection.

$$l_i^2 \approx \max_{t=1}^{F} \left\{ \frac{1}{s_t^2} (q_i^t)^2 \right\}$$

## Resolving Depth Ambiguity in Joint Space

The solution is still ambiguous in the sign of each bone depth. The number of possible solutions can be reduced by eliminating solutions containing impossible joint angles.



Original image | Complete tree of solutions | Solutions which satisfy joint angle limits | Final solution, chosen manually

## Results

Australian cricket player, Michael Clarke.

Korean figure skater, Yu Na Kim.