How tall was Newton?

Gravity as a Reference for Estimating a Person's Height from Video

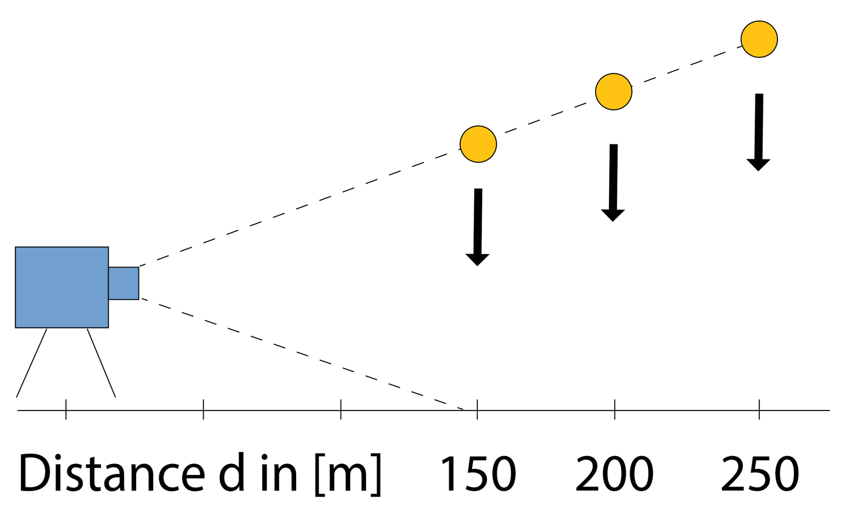
Didier Bieler, Semih Günel, Pascal Fua, and Helge Rhodin

CVLab, EPFL, Lausanne, Switzerland

UBC, Vancouver, Canada

Principles

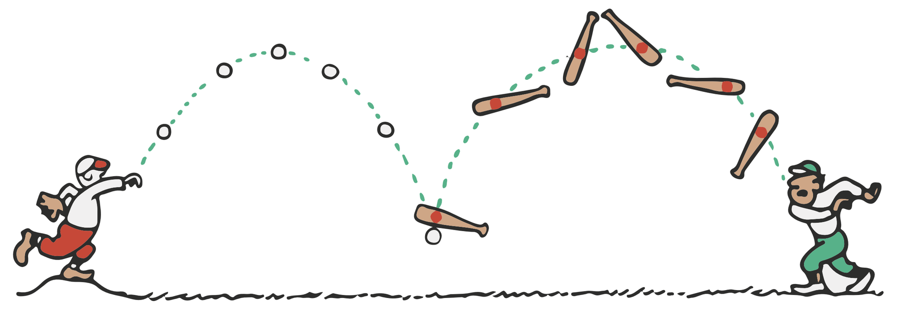
Gravity



Camera geometry

$$p(t) = p_0 + v_0 t + \frac{1}{2} a_0 t^2$$

Newton's equation of motion



Center of mass

Theory

Mathematical derivation. Absolute scales are derived by exploiting linearity in scaled-orthographic projection. We assume that perspective effects and air friction are minimal.

Projected height.

$$\begin{aligned} \mathbf{h}^{\mathrm{px}} &= \Pi(\mathbf{p}_u) - \Pi(\mathbf{p}_b) \\ &= \frac{f}{d}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}(\mathbf{p}_u - \mathbf{p}_b) \\ &= \frac{f}{d}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}\mathbf{n}h = \frac{h}{\mathbf{q}}, \end{aligned}$$

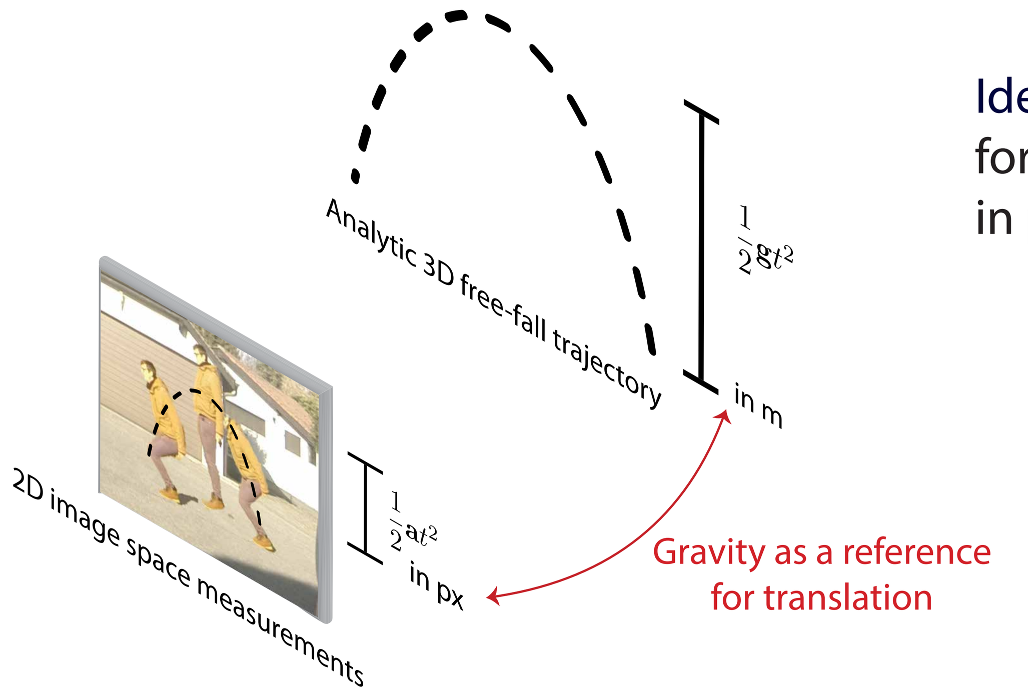
Projected free-fall motion.

$$\begin{aligned} \mathbf{p}^{\mathrm{px}}(t) &= \frac{f}{d}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}\mathbf{p}(t) \\ &= \frac{f}{d}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}\left(\frac{t^2}{2}\mathbf{g} + \mathbf{v}_0 t + \mathbf{p}_0\right) \end{aligned}$$

$$\mathbf{p}^{\mathrm{px}}(t) = \frac{1}{2}\mathbf{a}^{\mathrm{px}}t^2 + \mathbf{v}_0^{\mathrm{px}}t + \mathbf{p}_0^{\mathrm{px}} .$$

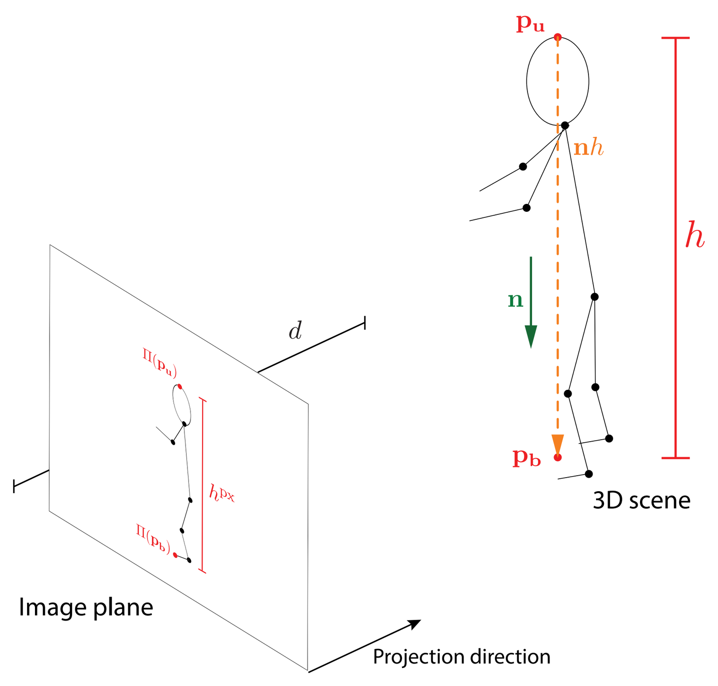
$$\Rightarrow \qquad \mathbf{a}^{\mathrm{px}} = \frac{f}{d}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}\mathbf{n}g$$

Approach



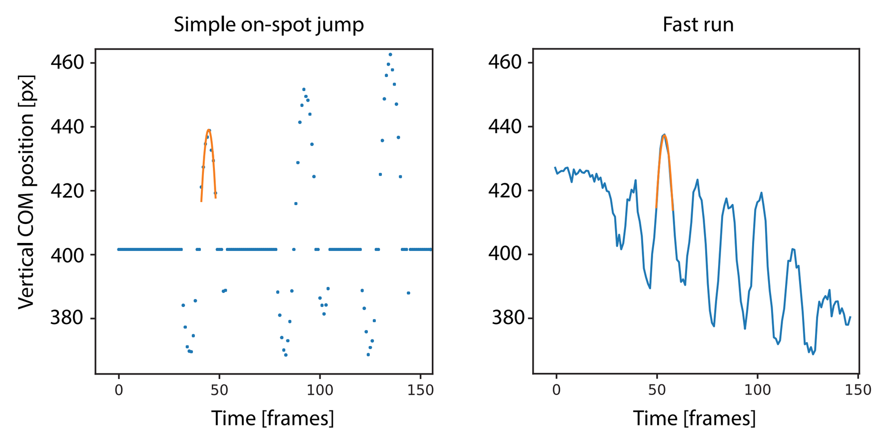
Idea. We exploit gravity as a reference *object* for mapping image height measurements in pixel [px] to metric height [m].

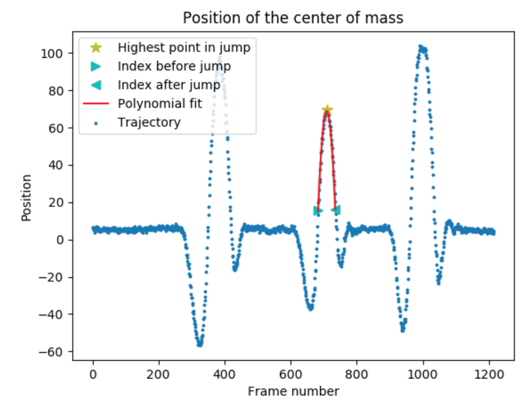
Projection of height. The projected height, h^{px} is the distance between two projected points $\Pi(p_u)$ and $\Pi(p_b)$ that are aligned with the direction of gravity *n* in 3D and span height *h*.



Motion detection. Bodypart keypoints are detected by AlphaPose, displayed as colored skeleton overlay.

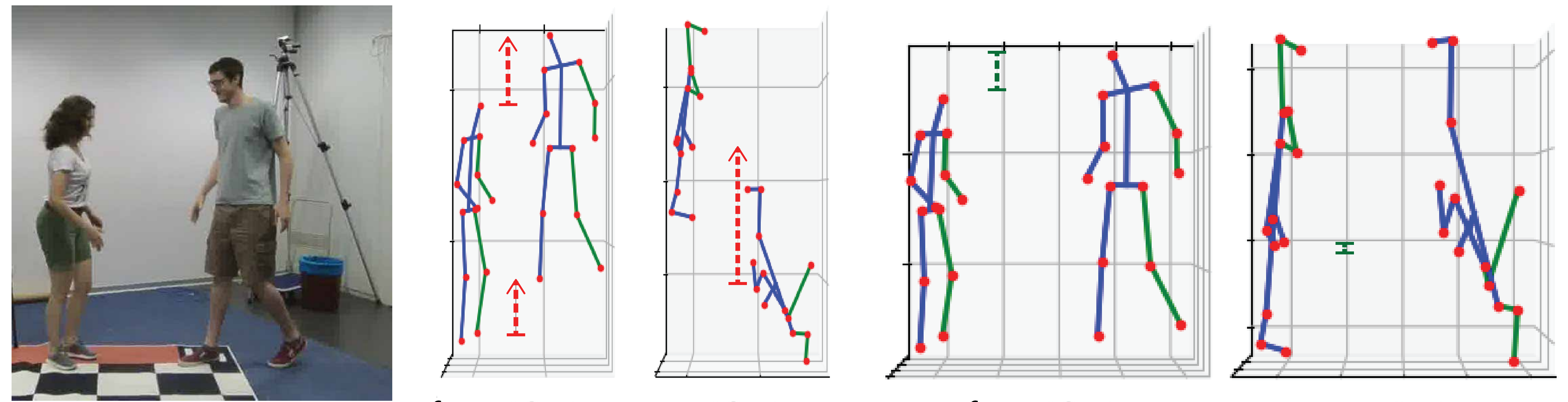
The center of mass (COM) is compued as a weighted average and acceleration through curve fitting.





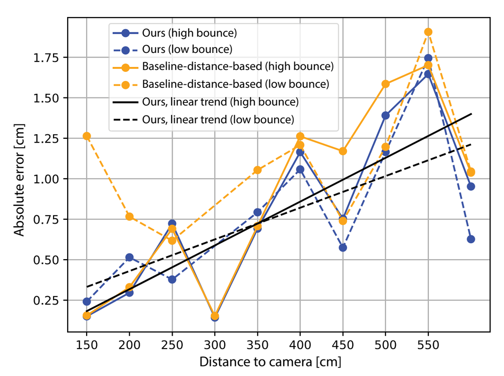
The approach assumes free-fall motion. We detect flight phases through simple heuristics on ground contact.

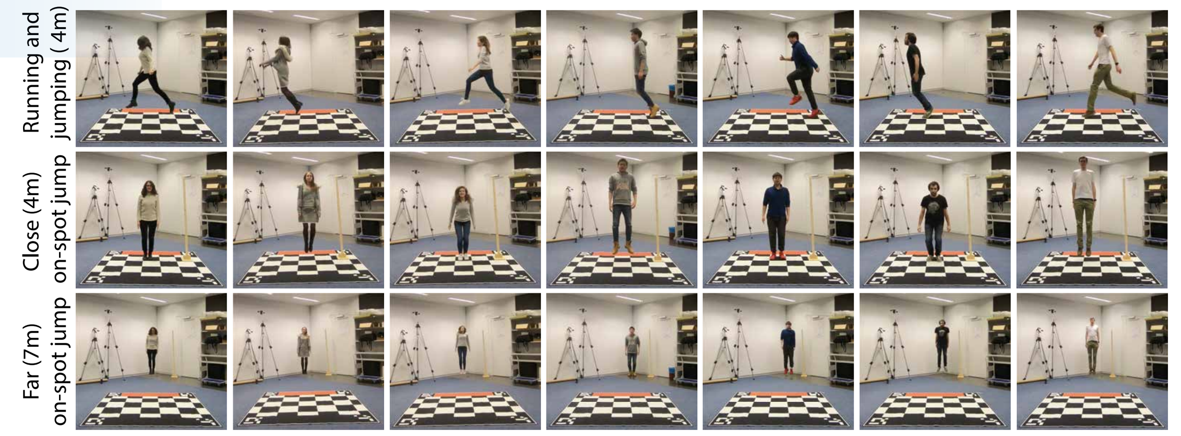
Applications



3D pose estimation methods do not recover the correct scales (marked in red). Our height from on-spot jumping recovers the scale and relative depth (in green).

Evaluation





On-spot jumps (low, medium, high)

	Error in h	
	Bias [cm] (ME±STD)	Accuracy [cm] (MAE±STD)
J1 (low)	-0.3±18.5	13.1±12.4
J2 (high)	-0.6±12.4	8.0±9.1
J3 (jack)	1.1±8	5.6±5.6
J4 (funny)	0.4±11.9	9.4±6.8

Subjects	4m distance				7m distance				Combined 4m and 7m	
ID / Height [cm] / Sex [f or m]	Error [cm]	Rel. [%]	Error [cm/px]	Rel. [%]	Error [cm]	Rel. [%]	Error [cm/px]	Rel. [%]	Error [cm]	Rel. [%]
Bias (ME±STD) across subjects:										
Ours	0.6±7.9	0.4±4.8	0.00±0.02	0±4.9	0.3±6.1	0.2±3.5	0.01±0.02	0.9±3.3	0.8±5.4	0.6±3.2
Baseline-distance-based	-7±21.3	4.3±12.5	0.02±0.05	4±13.1	-5.7±21.1	3.3±12	0.02±0.09	2.2±12.2	-7.7±16.5	4.5±9.5
Baseline-population-mean	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	-2.5±9.5	n/a
Gunel et al. [13]	3.7±6.8	2.3±4	n/a	n/a	3.6±6.9	2.3±4.1	n/a	n/a	3.7±6.7	2.3±4
Accuracy (MAE±STD) across subjects:										
Ours	4.5±6.4	2.7±3.9	0.01±0.02	3±3.8	4.6±3.8	2.7±2.1	0.02±0.01	2.9±1.6	3.9±3.7	2.3±3.2
Baseline-distance-based	18.3±11.9	10.8±7	0.05±0.03	11.7±6.4	13.6±16.7	7.8±9.5	0.05±0.07	7.2±9.9	15.5±8.8	9±5
Baseline-population-mean	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	8.1±5.0	n/a
Gunel [13]	6.5±4	3.8±2.5	n/a	n/a	6.5±3.9	3.8±2.5	n/a	n/a	6.5±3.9	3.8±2.4

Lateral motion (running and jumping)

	Jumping, 4m				Slow Run, 4m				Fast Run, 4m			
	h		q (= 0.0041)		h		q (= 0.0041)		h		q (GT: = 0.0041)	
	Error [cm]	Rel. [%]	Error [cm/px]	Rel. [%]	Error [cm]	Rel. [%]	Error [cm/px]	Rel. [%]	Error [cm]	Rel. [%]	Error [cm/px]	Rel. [%]
Bias (ME+STD)	0.4±7.5	0.2±4.4	-0.01±0.02	-1.2±4.2	11.7±19.7	6.8±11.8	-0.02±0.04	-5.4±9.7	11.4±8.5	6.6±5	-0.01±0.02	-3.2±4.5
Accuracy (MAE+STD)	6.6±3	3.9±1.8	0.01±0.01	3.3±2.8	18.9±12.2	11.2±7.3	0.03±0.03	7.2±8.3	12.3±7	7.1±4.1	0.02±0.02	4±3.7

Future work

- Perspective camera model (non-fronto-parallel motion, close distance)
- Moving and zooming cameras
- Better 2D detectors

Data and code at https://www.epfl.ch/labs/cvlab/articulated-free-fall-dataset/