SEMANTIC PARAMETRIC BODY SHAPE ESTIMATION FROM NOISY DEPTH SEQUENCES

Alexandru E. Ichim

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Federico Tombari

TECHNISCHE UNIVERSITÄT MÜNCHEN

CV LAB Computer Vision Laboratory

SOME OF MY OTHER WORK I

Static

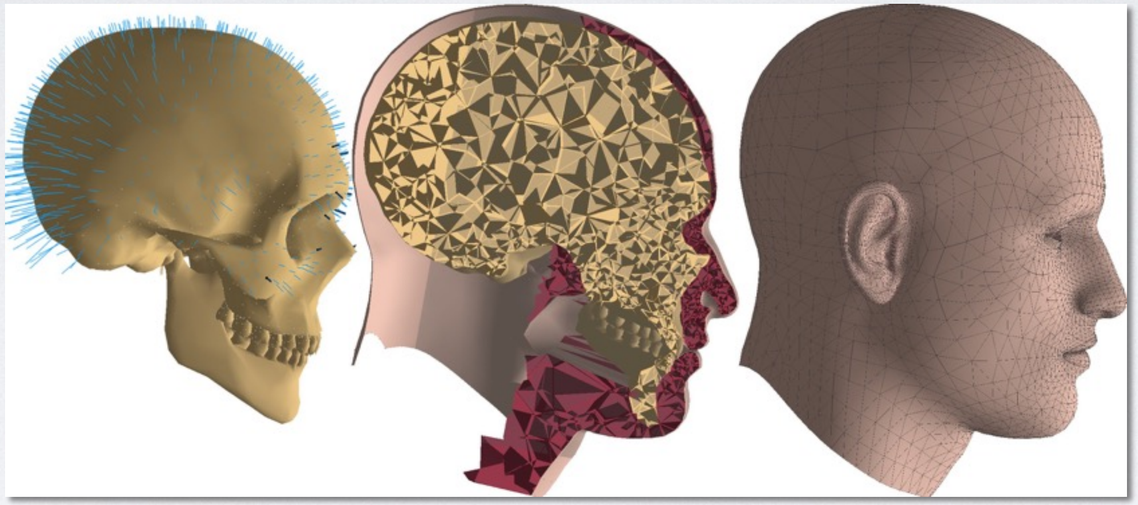
8 MP, ~100 images

Dynamic

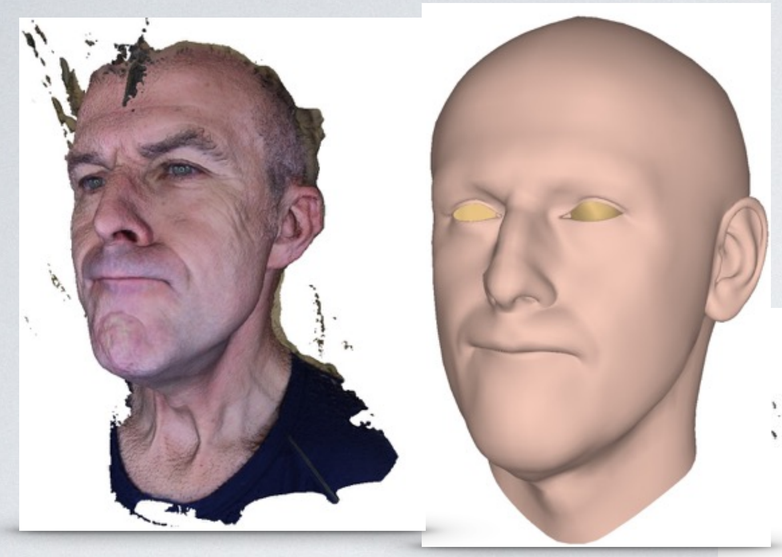
HD, ~2 min

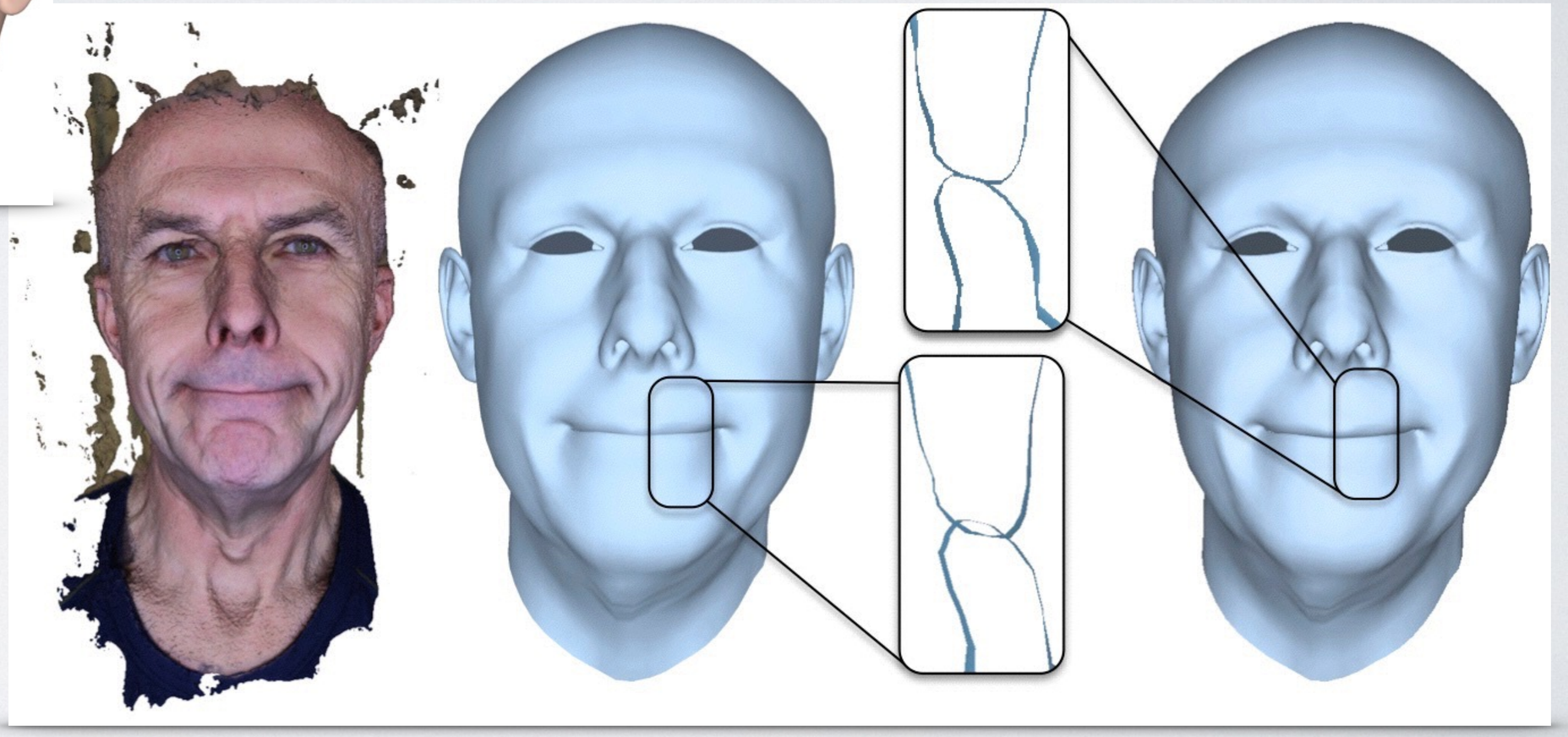
SOME OF MY OTHER WORK I

SOME OF MY OTHER WORK 2

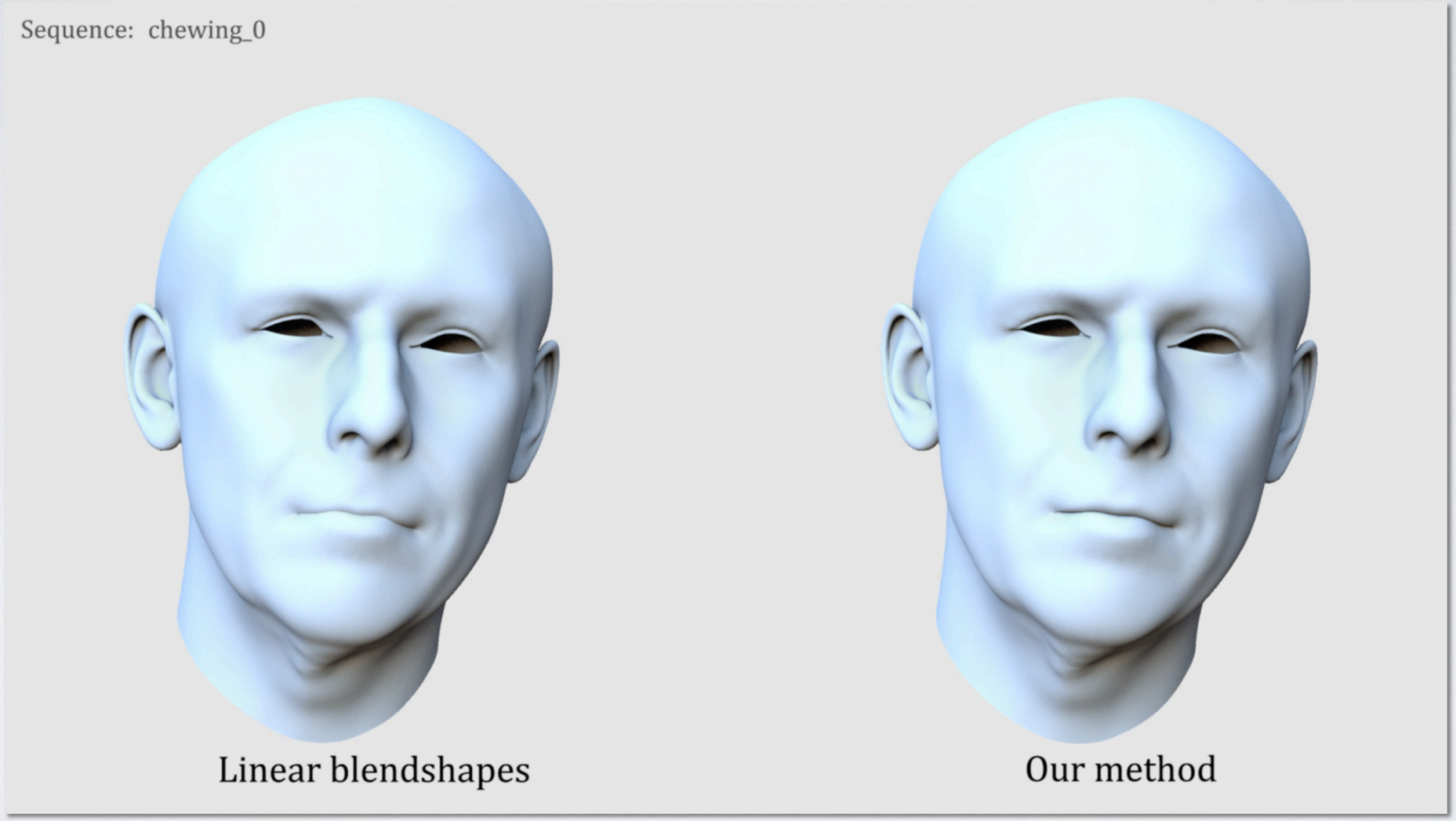


SOME OF MY OTHER WORK 2



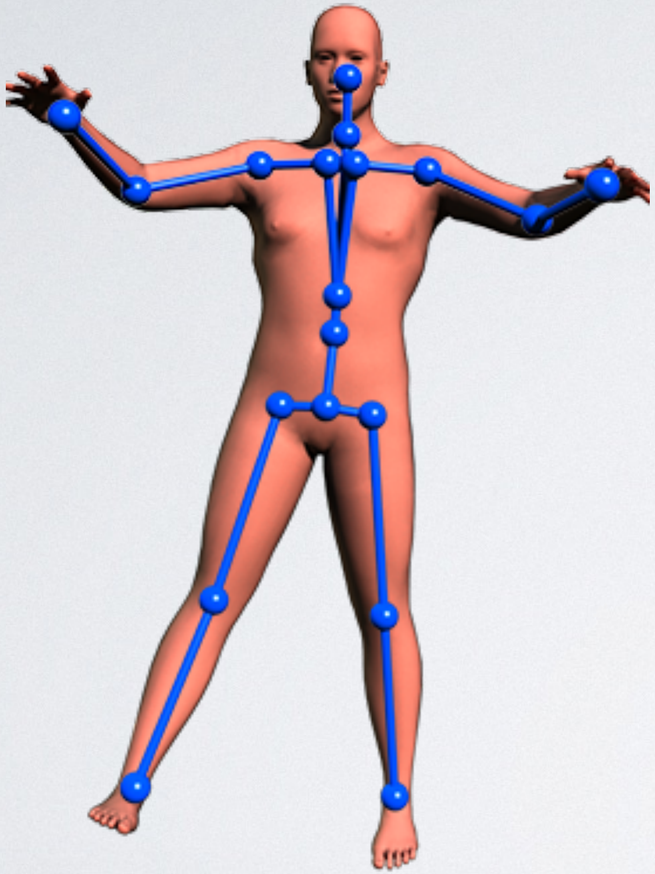


SOME OF MY OTHER WORK 2



MOTIVATION

Generic framework for tracking and modeling articulated bodies



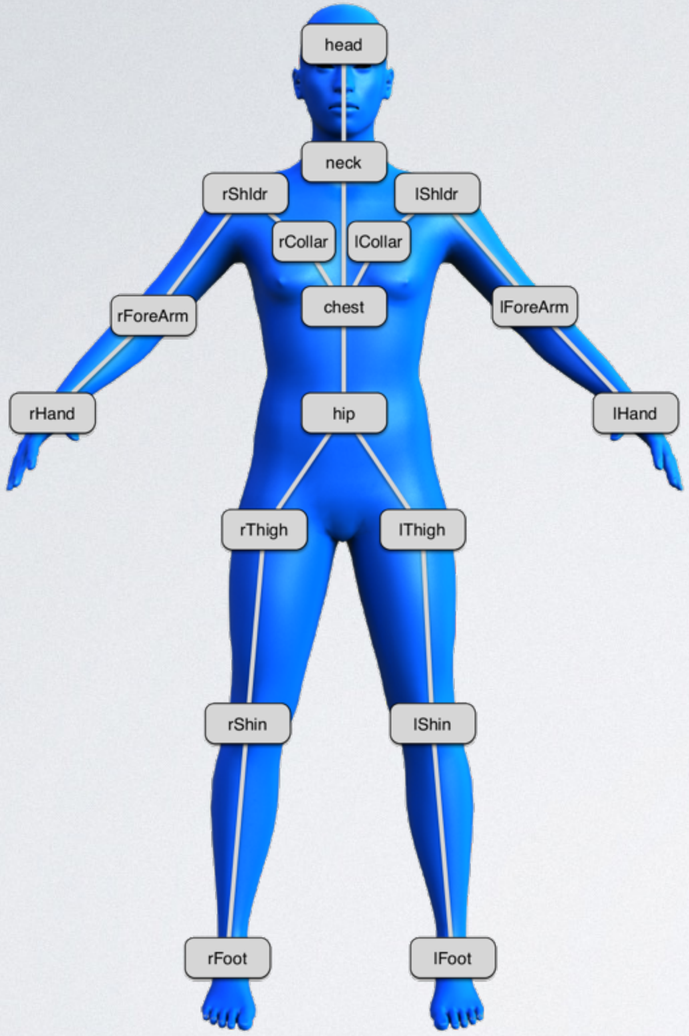
in our experiments:

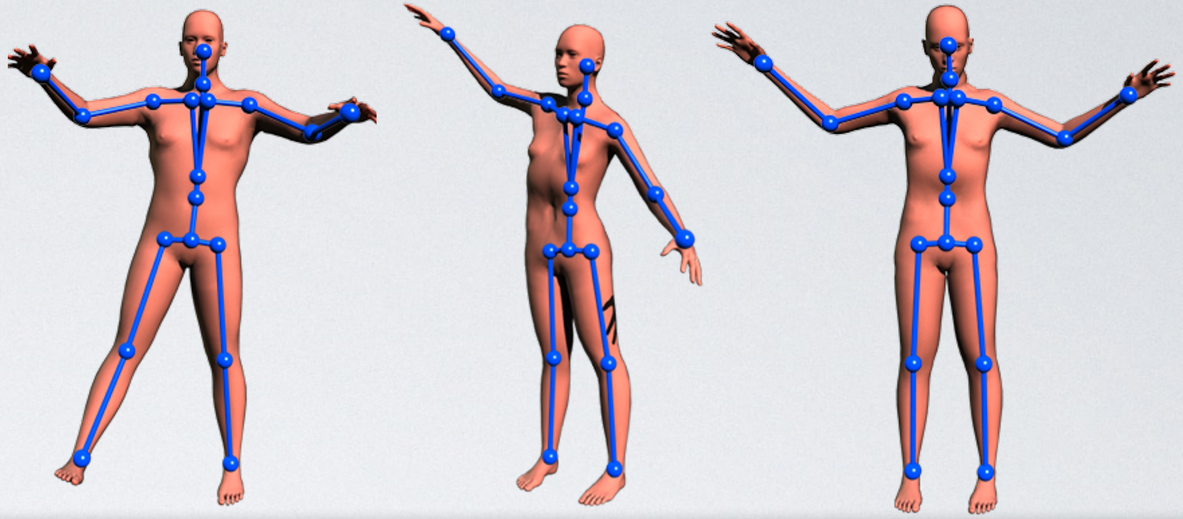
- human bodies
- RGB-D input

can be applied to:

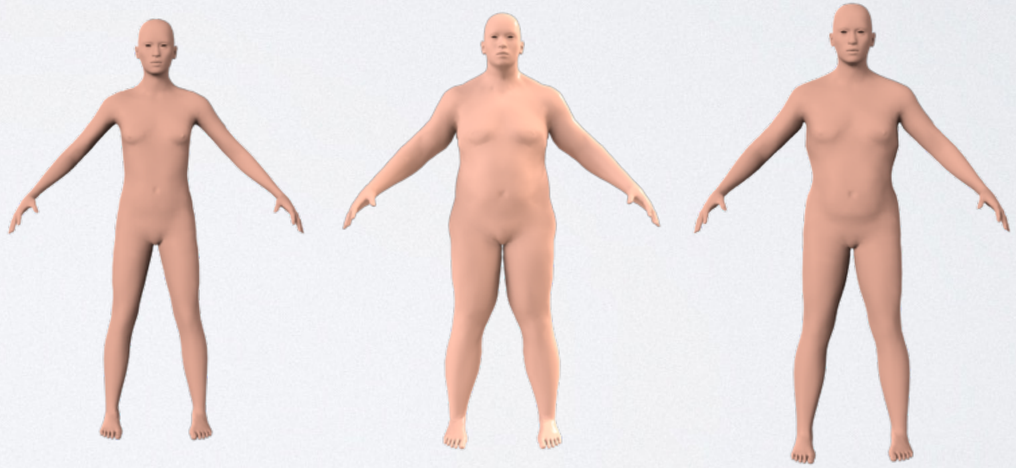
- hand tracking
- animal tracking

BODY REPRESENTATION



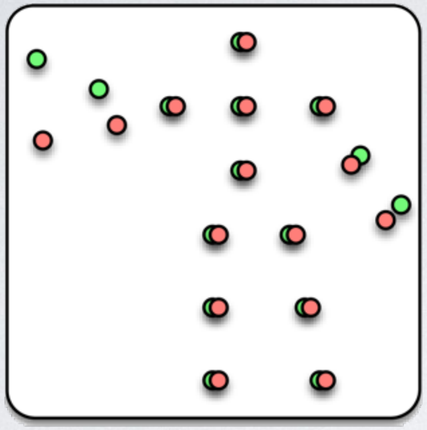


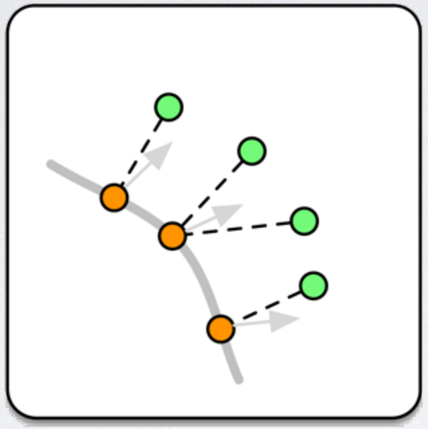
Pose variations - skeleton + linear blend skinning

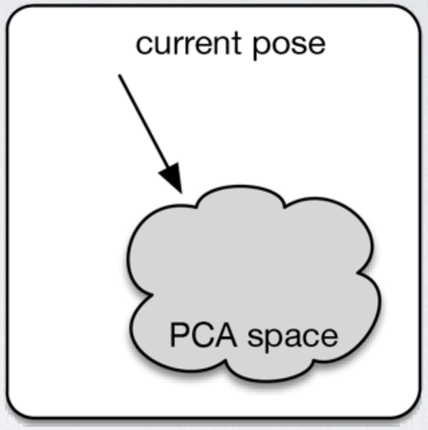


Rest pose variations - blendshapes + bone lengths

REGISTRATION ENERGY







$$E_{total} = \lambda_1 E_{features} + \lambda_2 E_{surface} + \lambda_3 E_{contour} + E_{prior} + \lambda_4 E_{smoothness}$$

REGISTRATION ENERGY

$$E_{total} = \lambda_1 E_{features} + \lambda_2 E_{surface} + \lambda_3 E_{contour} + E_{prior} + \lambda_4 E_{smoothness}$$

$$E_{features} = \sum_j \left|\left| \boldsymbol{t}_j^{pose} - \boldsymbol{\chi}_j \right|\right|_2^2$$

Features

NiTE landmarks - replaceable

$$E_{surface} = \sum_i \left|\left| \boldsymbol{n}_i^T (\boldsymbol{x}_i - \boldsymbol{v}_i) \right|\right|_2^2$$

Point-to-plane dense closeness

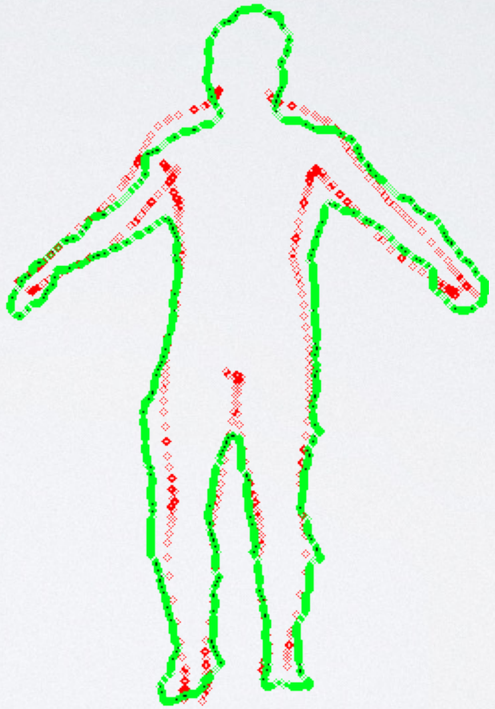
ICP iterations

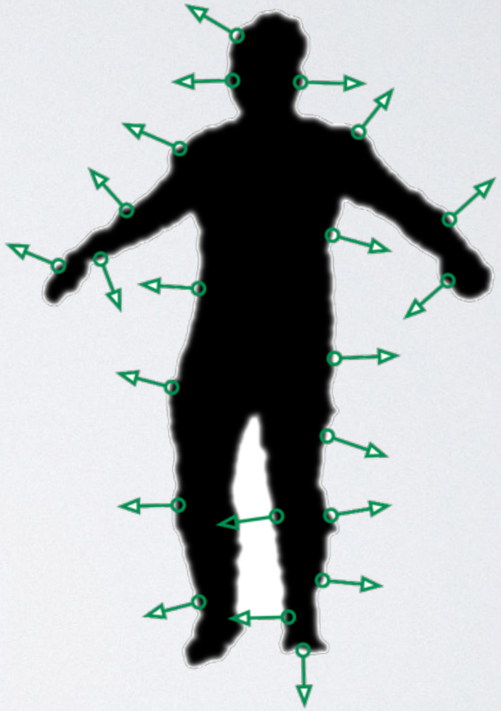
REGISTRATION ENERGY

$$E_{total} = \lambda_1 E_{features} + \lambda_2 E_{surface} + \lambda_3 E_{contour} + E_{prior} + \lambda_4 E_{smoothness}$$

$$\boldsymbol{n} = \frac{(\nabla x \bar{\boldsymbol{I}}_{contour}, \nabla y \bar{\boldsymbol{I}}_{contour}, 0)^T}{||(\nabla x \bar{\boldsymbol{I}}_{contour}, \nabla y \bar{\boldsymbol{I}}_{contour}, 0)||}$$

Contour





extracted from the 2D depth image

can be extended to RGB-segmentation

REGISTRATION ENERGY

$$E_{total} = \lambda_1 E_{features} + \lambda_2 E_{surface} + \lambda_3 E_{contour} + E_{prior} + \lambda_4 E_{smoothness}$$

$$E_{prior} = \lambda_5 E_{prior\_proj} + \lambda_6 E_{prior\_dev}$$

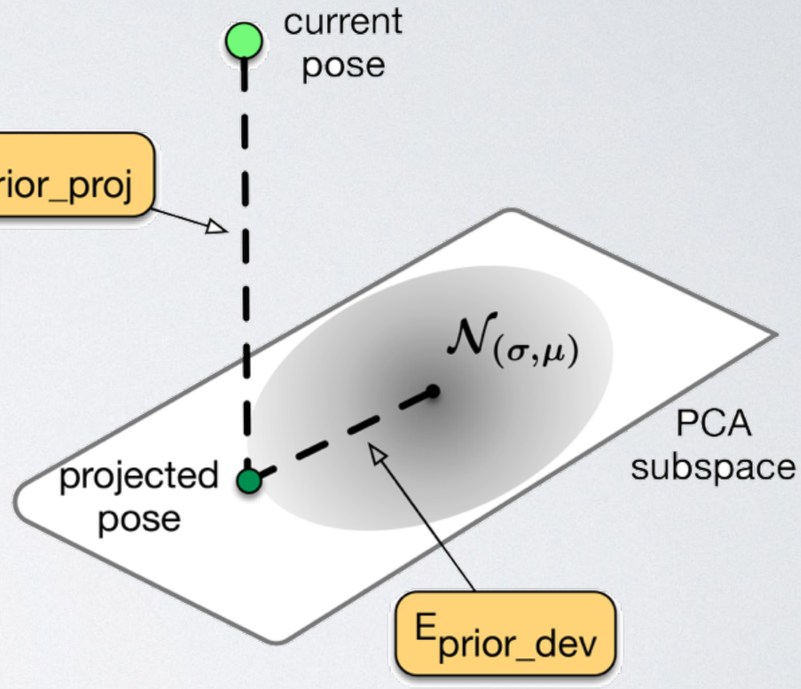
$$E_{prior\_proj} = \left|\left|(\boldsymbol{\beta} - \boldsymbol{\mu}) - \boldsymbol{M}\boldsymbol{M}^T(\boldsymbol{\beta} - \boldsymbol{\mu})\right|\right|_2^2$$

distance to projection

$$E_{prior\_dev} = \left|\left|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{M}^T(\boldsymbol{\beta} - \boldsymbol{\mu})\right|\right|_2^2$$

std-dev-weighting of the projection

Statistical Pose Prior



REGISTRATION ENERGY

$$E_{total} = \lambda_1 E_{features} + \lambda_2 E_{surface} + \lambda_3 E_{contour} + E_{prior} + \lambda_4 E_{smoothness}$$

$$E_{smoothness} = \left|\left|\boldsymbol{\beta}_{(t)} - \boldsymbol{\beta}_{(t-1)}\right|\right|_2^2$$

Smoothness term

MODELING ENERGY

$$E_{modeling} = \gamma_1 E_{surface} + \gamma_2 E_{contour} + \gamma_3 E_{bs\_reg}$$

$$E_{bs\_reg\_L2} = ||\boldsymbol{s}||_2^2$$

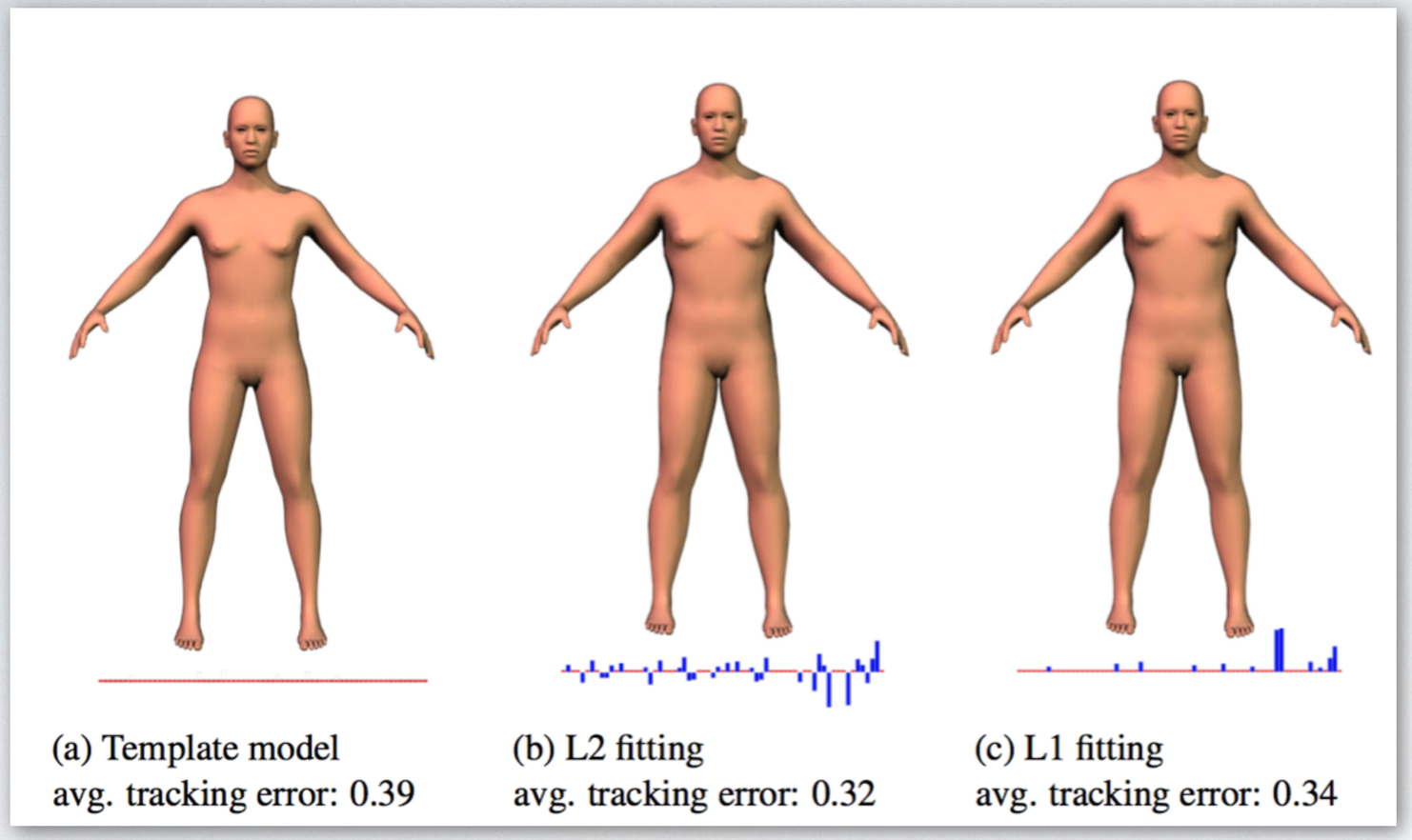
$$E_{bs\_reg\_L1} = ||\boldsymbol{s}||_1$$

Regularization terms

accumulate constraints over windows of frames - performance trade-off

solve and update the body template shape

MODELING ENERGY



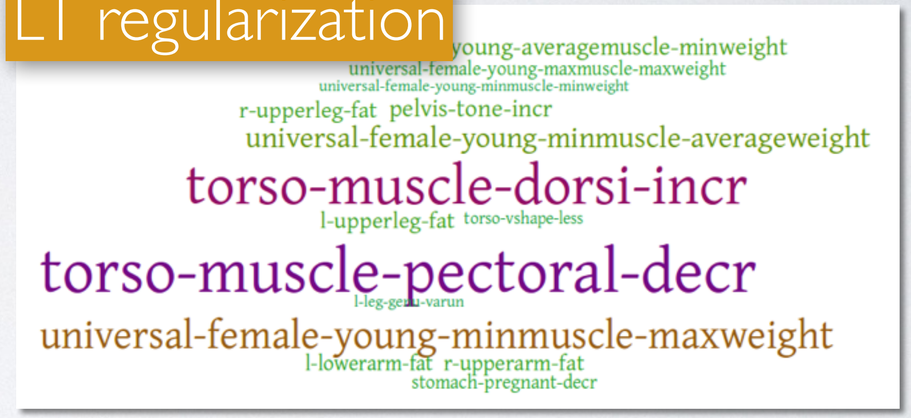
(a) Template model
avg. tracking error: 0.39

(b) L2 fitting
avg. tracking error: 0.32

(c) L1 fitting
avg. tracking error: 0.34

L2 regularization

L1 regularization



word cloud visualizations

RESULTS

Semantic Parametric Body Shape Estimation from Noisy Depth Sequences

Alexandru E. Ichim Federico Tombari

project webpage: http://lgg.epfl.ch/~ichim/bodies_ras_2015/

Supplementary video

Thank you!