Robust Estimation of 3D Human Poses from a Single Image

RAO Anyi

ESE

Overview

- Sparse coding to to reduce the ambiguities in the 3D pose estimation and is robust to occlusions (e.g. missing joints).
- Enforcing human-portion constrasints to get a minimization problem with constraints.

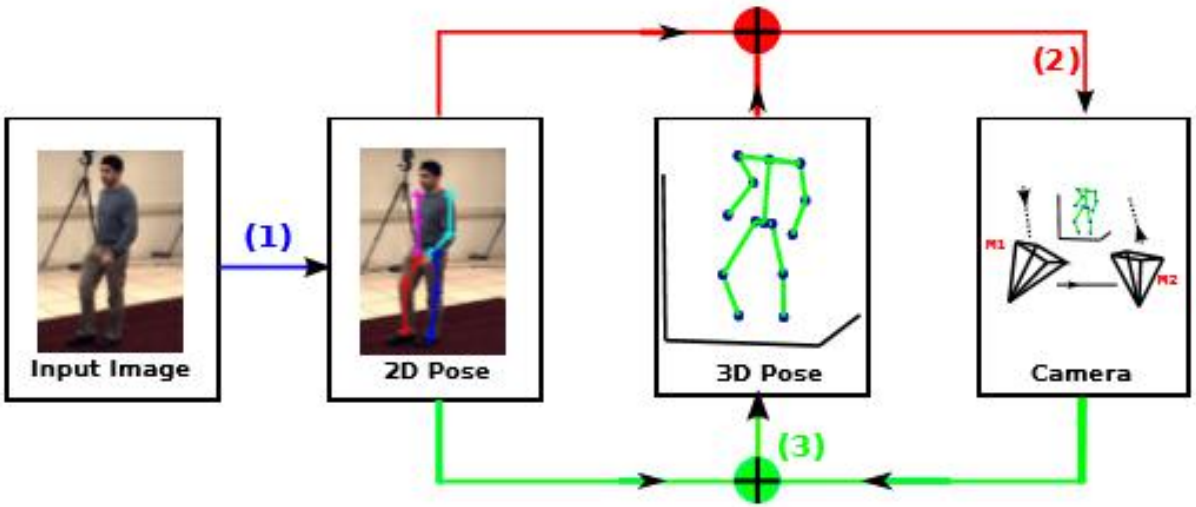


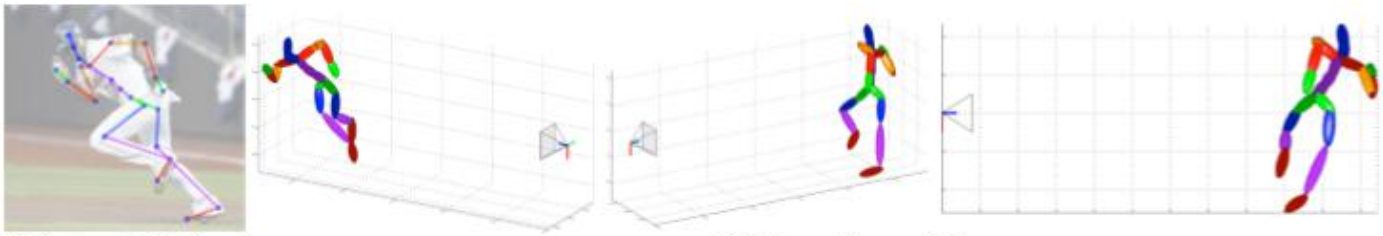
Figure 1. **Method overview**. (1) On a test image, we first estimate the 2D joint locations and initialize a 3D pose. (2) Then camera parameters are estimated from the 2D and 3D poses. (3) Next we update the 3D pose with the current camera parameters and the 2D pose. We repeat steps (2) and (3) until convergence.

Origin of the Idea

Reconstructing 3D Human Pose from 2D Image Landmarks

Varun Ramakrishna, Takeo Kanade, Yaser Sheikh

Robotics Institute, Carnegie Mellon University



2D Anatomical Landmarks

3D Human Pose and Camera

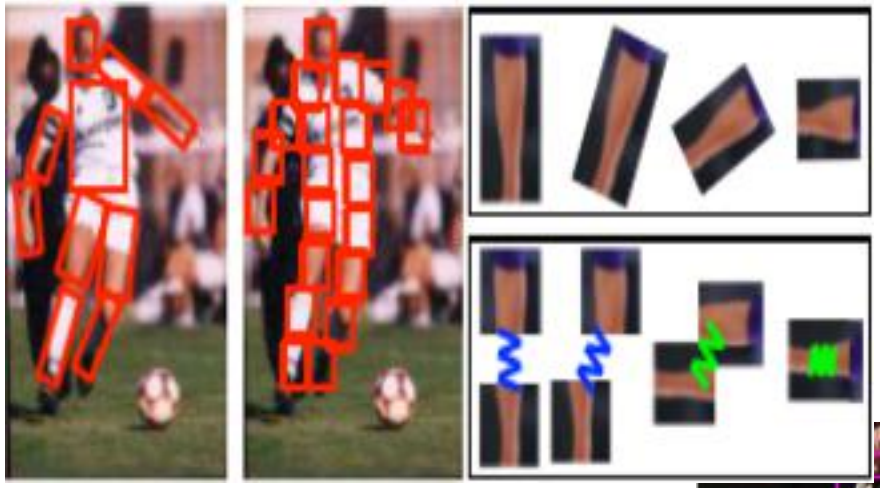
Fig. 1. Given the 2D location of anatomical landmarks on an image, we estimate the 3D configuration of the human as well as the relative pose of the camera.

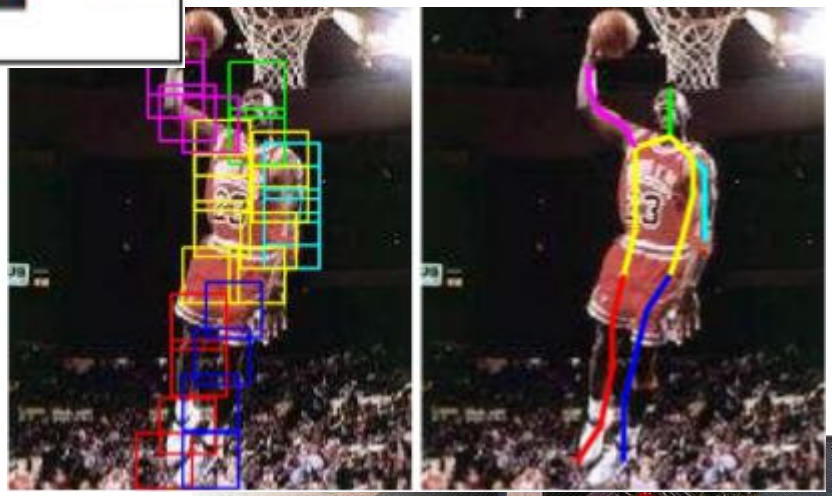
Detector rather than Manual L abeling

Manual labeling

Articulated Human Detection with Flexible Mixtures-of-Parts

Yi Yang, *Member, IEEE,* and Deva Ramanan, *Member, IEEE*





Eight limb length constraints rather than sum of limb length constraints

Sum or Parts ?

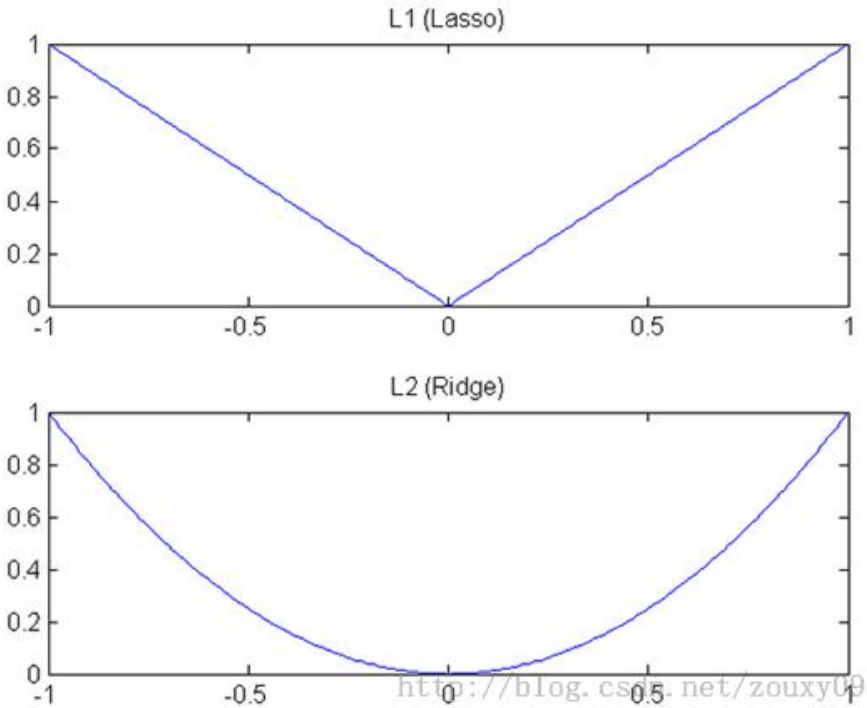
Enforcing anthropometric regularity through strict limb length constraints is intractable because satisfying multiple quadratic equality constraints on a least squares system is nonconvex. Instead, we encourage anthropometric regularity by enforcing a necessary condition (i.e., an equality constraint on the sum of squared lengths) as a constraint that is applied.

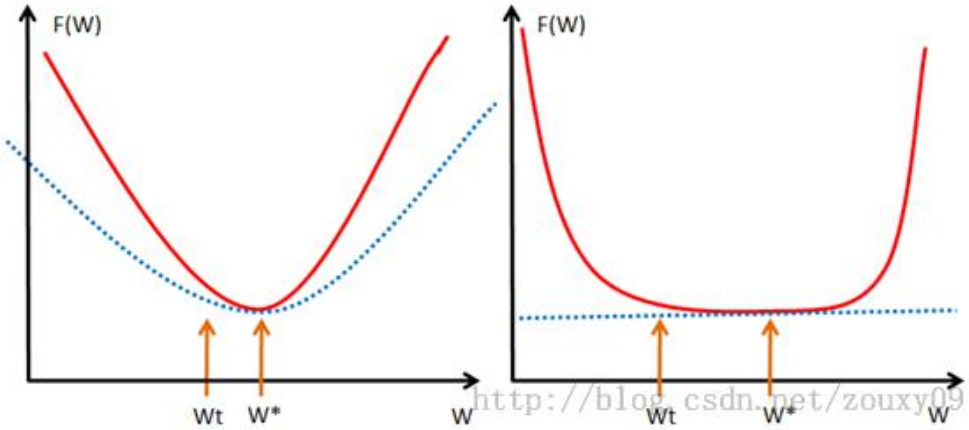
The eight limbs are left/rightupper/lower-arm/leg. Normalizing the length of the right lower leg to one and compute the squared lengths of other limbs (say Li) according to the limb proportions.

L1-norm rather than L2-norm

Refer to

Convergence Rate and Accuracy

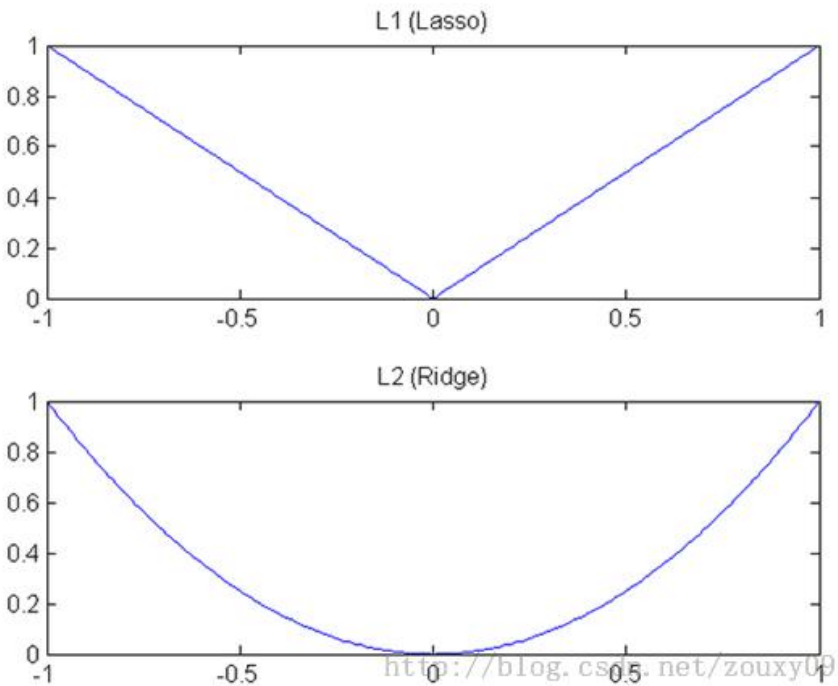


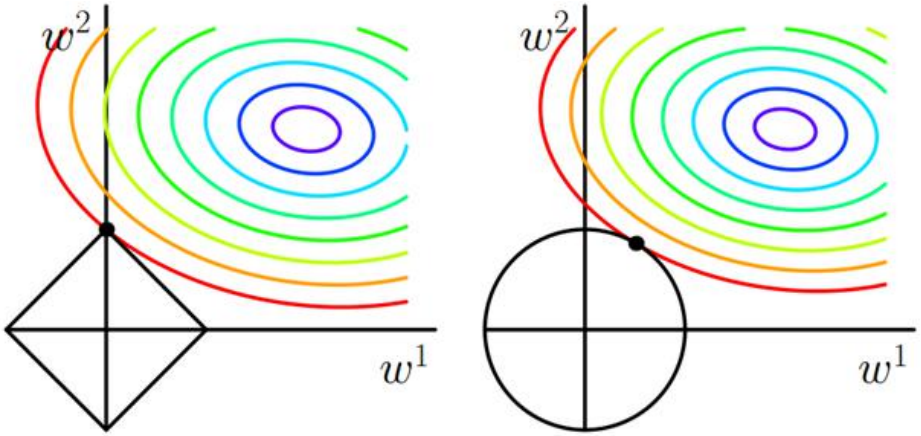


L1-norm rather than L2-norm

Refer to

Sparsity





(a) ℓ_1-ball meets quadratic function. ℓ_1-ball has corners. It's very likely that the meet-point is at one of the corners.

(b) ℓ_2-ball meets quadratic function. ℓ_2-ball has no corner. It is very unlikely that the meet-point is on any of axes.

L2-norm is sensitive to inaccuracies in 2D pose estimation, which are usually caused by failures in feature detections and other factors, because it tends to distribute errors uniformly.

Approach-Robust 3D Pose Estimation

$$y = \sum_{i=1}^{k} \alpha_i \cdot b_i + \mu$$

3D pose

$$\|x - M(B\alpha + \mu)\|_1$$

Estimating the coefficients *alpha* by minimizing an L1-norm error between the projection of the estimated 3D pose and the 2D pose

$$\min_{\alpha} \quad \|x - M(B\alpha + \mu)\|_1 + \theta \|\alpha\|_1$$

The sparsity can be induced by minimizing the L1-norm of *alpha*

- Removing incorrect one
- Preventing overfitting to (inaccurate) 2D pose

$$\min_{\alpha} \quad \|x - M(B\alpha + \mu)\|_1 + \theta \|\alpha\|_1$$

$$\text{s.t.} \quad \|C_i(B\alpha + \mu)\|_2^2 = L_i, i = 1, \cdots, t$$

Anthropomorphic Constraints

Approach-Camera Parameter Estimation

Camera $M_0 = \begin{pmatrix} m_1^T \\ m_2^T \end{pmatrix}$ s.t. $m_1^T m_2 = 0.$

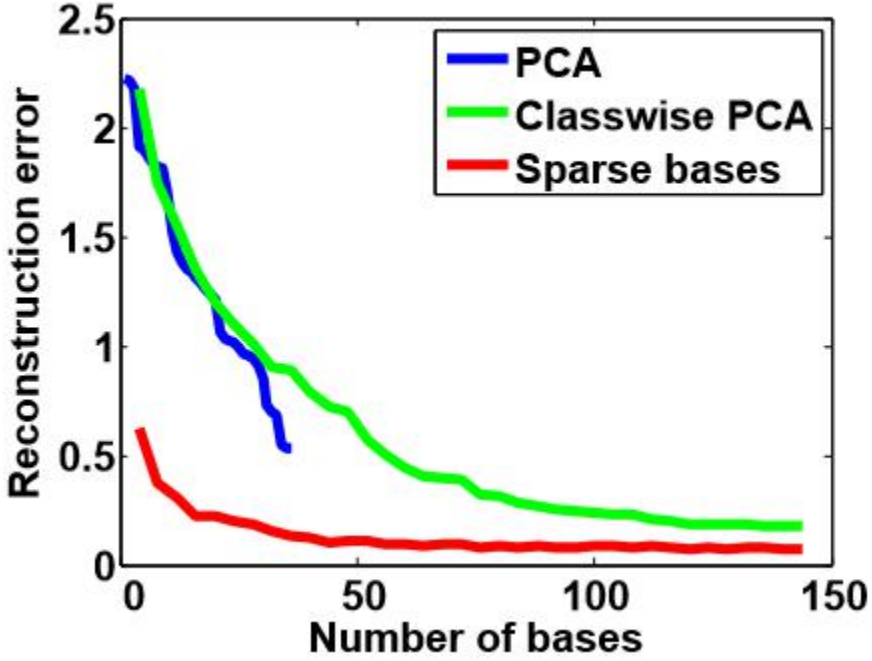
2D pose → $X = M_0 Y$ ← 3D pose

$$\min_{m_1, m_2} \left\| X - \begin{pmatrix} m_1^T \\ m_2^T \end{pmatrix} Y \right\|_1, \quad \text{s.t.} \quad m_1^T m_2 = 0.$$

Experimental Results

- Controlled experiments (assuming that the 2D joint locations are known)
- Real experiments

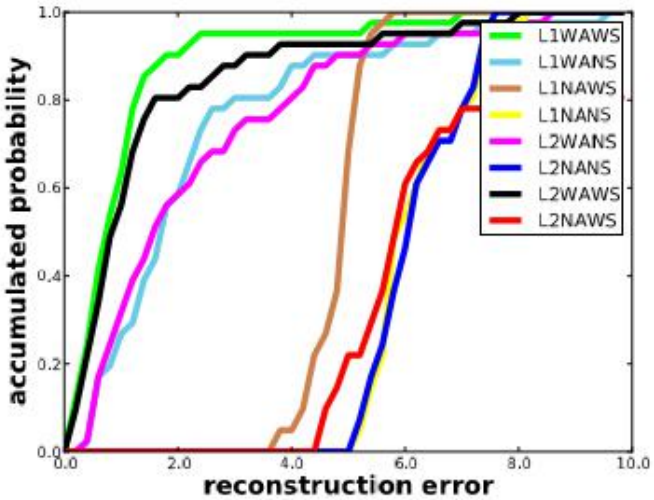
Basis Learning

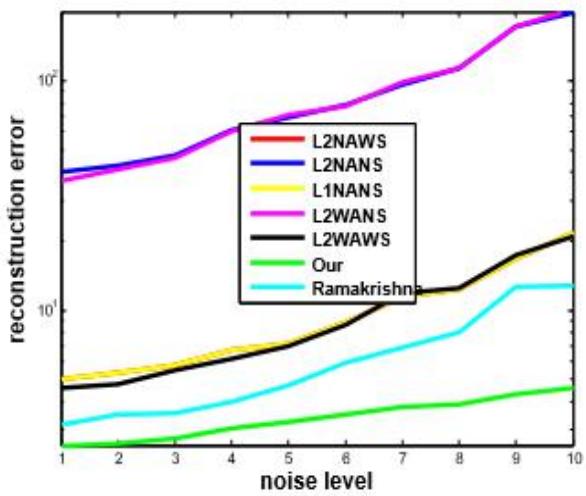


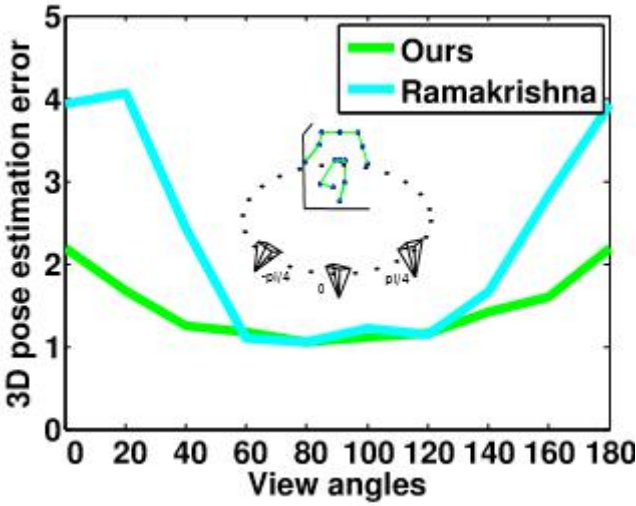
3D pose $B = \{b_1, \cdots, b_k\}$ $y = \sum_{i=1}^{k} \alpha_i \cdot b_i + \mu$

Controlled Experiments

Premise: 2D joint locations are known







Of Three Factors
norm/anthropo
morphic/sparsity

Of Inaccurate 2D

Of Human-
Camera

Real Experiments

Detecting the 2D joint locations by a detector firstly

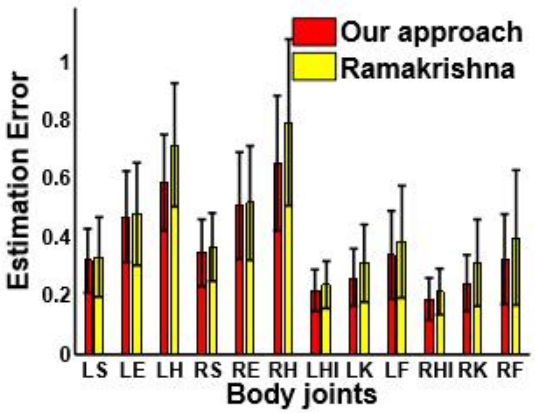


Figure 7. **Real experiment on the UvA dataset:** comparison with a state-of-the-art [12]. Both average estimation errors and standard deviations are shown for each joint (*i.e.* left shoulder, left elbow, left hand, right shoulder, right elbow, right hand, left hip, left knee, left foot, right hip, right knee and right foot). See Section 4.4.1.

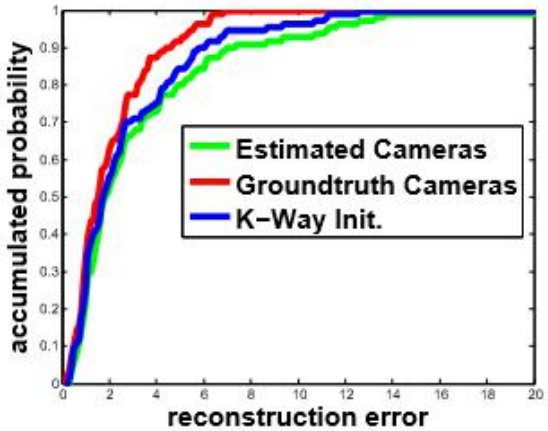


Figure 8. **Real experiment on the CMU dataset:** cumulative distribution of 3D pose estimation errors when camera parameters are (1) assigned by groundtruth, estimated by initializing the 3D pose with (2) mean pose, or (3) 30 cluster centers. The y-axis is the percentage of the cases whose estimation error is less than or equal to the corresponding x-axis value on the curves.

My Experiments

CMU Graphics Lab Motion Capture Database

Subject #1 (climb, swing, hang on playground equipment)		
Image	Trial #	Motion Description
	1	playground - forward jumps, turn around
	2	playground - climb
	3	playground - climb, hang, swing
	4	playground - climb
	5	playground - climb, go under
	6	playground - climb, sit, dangle legs, descend
	7	playground - climb, sit, dangle legs, jump down
	8	playground - climb, sit, dangle legs, rock back, lower self to ground
	9	playground - climb, hang, hold self up with arms straight, swing, drop,
	10	playground - climb, swing, lean back, drop
	11	playground - climb, hang, lean over, jump down
	12	playground - climb, pull up, dangle, sit, lower self to ground
	13	playground - climb, go under, jump down
	14	playground - climb, jump down, dangle, legs push off against

My Experiments

CVX RESEARCH

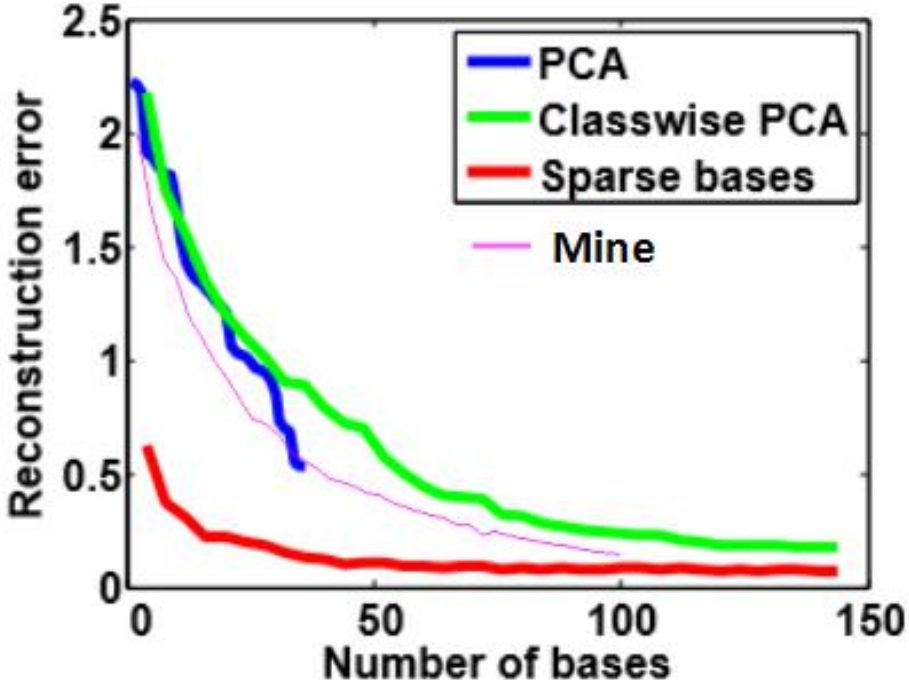
SPArse Modeling Software

Stanford University

Inria INVENTEURS DU MONDE NUMÉRIQUE

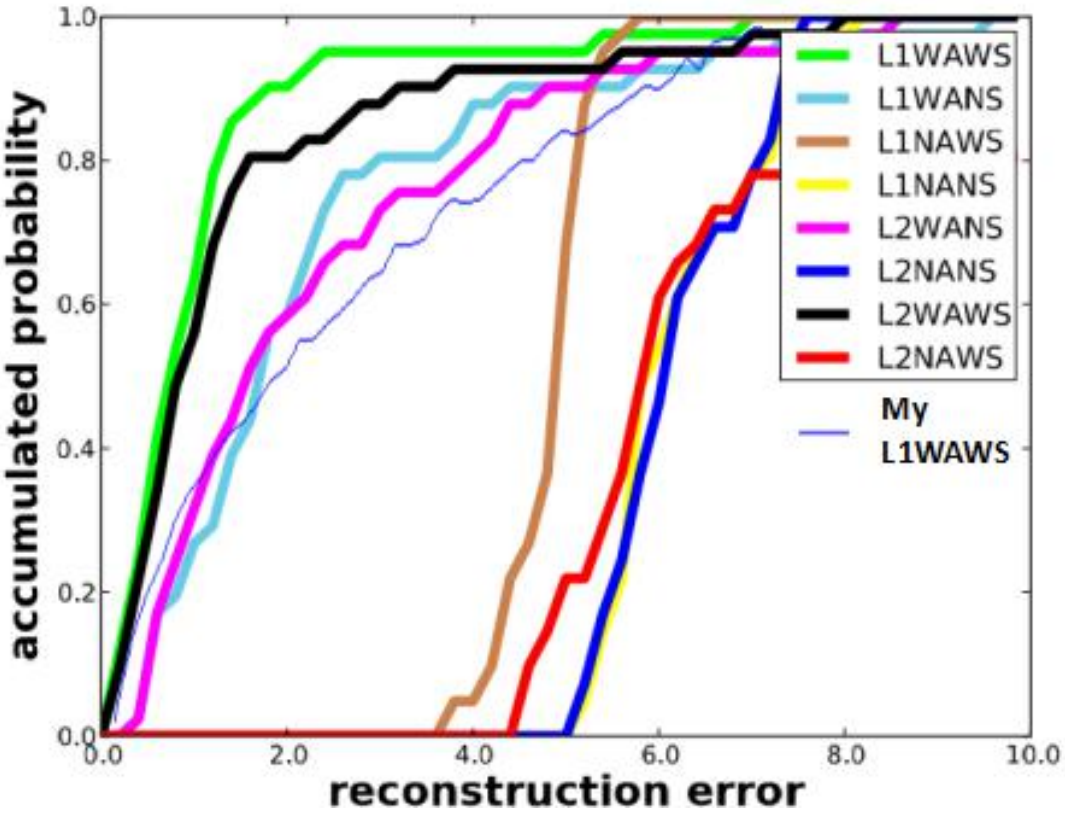
$$
\begin{array}{ll}
\text{minimize} & \|Ax - b\|_2 \\
\text{subject to} & Cx = d \\
& \|x\|_\infty \leq e
\end{array}
$$

My Experiments -Basis Learning



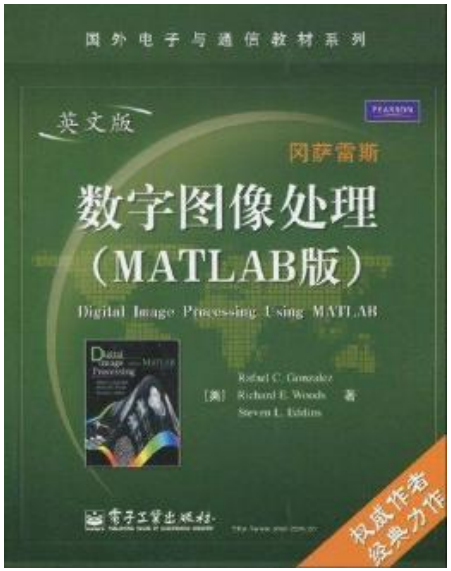
3D pose $B = \{b_1, \cdots, b_k\}$ $y = \sum_{i=1}^{k} \alpha_i \cdot b_i + \mu$

My Experiments -Controlled Experiment



MORE

Model



Vs

Machine Learning

Thanks !