Estimating Human Pose from Occluded Images

Jia-Bin Huang and Ming-Hsuan Yang

{jbhuang,mhyang}@ieee.org

Electrical Engineering and Computer Science
University of California at Merced

ACCV, Xi' an, September , 2009

What this talk is about

- Estimating human poses from single 2D images as a direct nonlinear regression
- Recovering poses even when humans in the scenes are partially or heavily occluded via sparse approximation
- Achieving (implicitly) relevant feature selection

Outline

1. Introduction
2. Recovering Poses via Sparse Approximation
3. Experimental Setups
4. Conclusions

Introduction

Motivation - Applications

- **Motion Analysis**
  - Diagnostics of orthopedic patients
  - Analysis and optimization of an athletes' performances
  - Content-based retrieval and compression of video
  - Auto saftey: control of airbags, drowsiness detection, pedestrian detection, etc.
- **Surveillance**
  - People counting or crowd flux, flow, and congestion analysis
  - Analysis of actions, activities, and behaviors both for crowds and individuals
- **Human Computer Interaction**
  - Virtual Reality

Motivation - Applications

- **Motion Analysis**
  - Diagnostics of orthopedic patients
  - Analysis and optimization of an athletes' performances
  - Content-based retrieval and compression of video
  - Auto saftey: control of airbags, drowsiness detection, pedestrian detection, etc.
- **Surveillance**
  - People counting or crowd flux, flow, and congestion analysis
  - Analysis of actions, activities, and behaviors both for crowds and individuals
- **Human Computer Interaction**
  - Virtual Reality

Motivation - Applications

- **Motion Analysis**
  - Diagnostics of orthopedic patients
  - Analysis and optimization of an athletes' performances
  - Content-based retrieval and compression of video
  - Auto saftey: control of airbags, drowsiness detection, pedestrian detection, etc.
- **Surveillance**
  - People counting or crowd flux, flow, and congestion analysis
  - Analysis of actions, activities, and behaviors both for crowds and individuals
- **Human Computer Interaction**
  - Virtual Reality

Key Challenges

Why the problem is difficult?

- Ambiguity: Loss of depth information
- Variations of shape and appearance of human body
- Background clutters
- Occlusions

In this paper, we address the problems of occlusions and background clutters.

Related Works

Model-based (Generative)

- Employ a know model (e.g., tree structure) based on prior knowledge
- Include two parts: 1) Modeling, 2) Estimation

[Felzenszwalb, Huttenlocher '00], [Ioeffe et al. '01], [Ronfard et al. '02], [Ramnan, Forsyth '03], [Mori, Malik, '04]

Model-free (Discriminative)

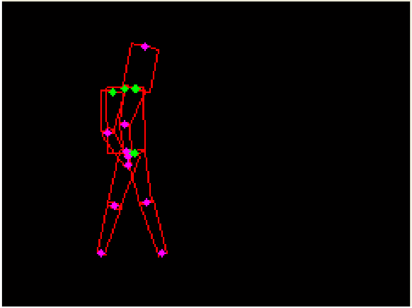
- Example-based
  [Shakhnarovich et al. '03], [Poppe '07]
- Learning-based
  [Rosales, Sclaroff, '01], [Agarwal, Triggs '06], [Sminchisescu et al. '07], [Bissacco et al. 07], [Okada, Soatto '08]

Recovering Poses via Sparse Approximation

Problem Formulation

- Given: N training samples $\{\mathbf{x}_1, \mathbf{y}_1\}, \{\mathbf{x}_2, \mathbf{y}_2\} \cdots, \{\mathbf{x}_N, \mathbf{y}_N\}$
- Input: test sample $\mathbf{b}$
- Output: pose descriptor vector $\mathbf{y}$

Input: single image



Output: pose

Test Image as a Sparse Linear Combination of Training Images

- Given N training samples $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N \in \mathbb{R}^m$, we represent a test sample b by

$$\mathbf{b} = \omega_1 \mathbf{x}_1 + \omega_2 \mathbf{x}_2 + \cdots + \omega_N \mathbf{x}_N$$

- Let $A = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N] \in \mathbb{R}^{m \times N}$,

$$\mathbf{b} = A\omega,$$

  where $\omega = [\omega_1, \omega_2, \ldots, \omega_N]^{\mathrm{T}}$ is the coefficient vector.
- We want to find the sparsest solution of the linear system.

Goal: Find the sparsest solution

Recover the sparsest solution via ℓ_1-norm minimization

- ℓ_0-norm solution–NP hard

$$\min_{\omega} \|\omega\|_0 \quad \text{subject to} \quad \mathbf{b} = A\omega$$

- ℓ_1-norm solution–convex optimization problem

$$\min_{\omega} \|\omega\|_1 \quad \text{subject to} \quad \mathbf{b} = A\omega$$

- Relaxed constraint on equality to allow small noise

$$\min_{\omega} \|\omega\|_1 \quad \text{subject to} \quad \|\mathbf{b} - A\omega\|_2 \leq \epsilon$$

Coping with Background Clutter and Occlusions

- When errors occur? 1) Misalignment, 2) Background clutter, 3) Occlusions.
- Introduce an error term $\mathbf{e}$

$$\mathbf{b} = A\omega + \mathbf{e} = [A \quad I] \begin{bmatrix} \omega \\ \mathbf{e} \end{bmatrix} = B\mathbf{v}$$

- Solve the extended linear system

$$\min \quad ||\mathbf{v}||_1 \quad \text{subject} \quad \text{to} \quad \|\mathbf{b} - B\mathbf{v}\|_2 \leq \epsilon$$

- Recovered test sample $\mathbf{b}_R$ can be represented as $A\omega$

Occlusion Recovery: An Example

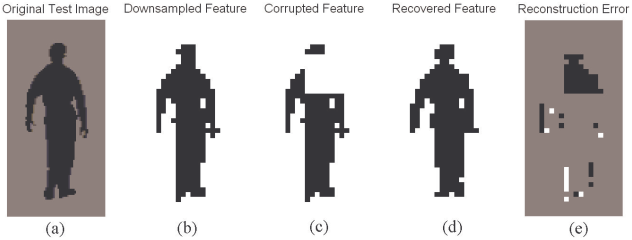


Figure: Occlusion recovery on a synthetic dataset. (a)(b) The original input image and its feature. (c) Corrupted feature via adding random block. (d) Recovered feature via find the sparsest solution. (e) Reconstruction error.

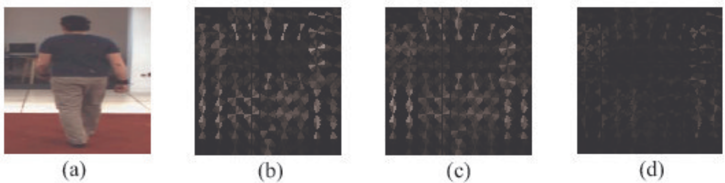


Figure: Feature selection example. (a) Original test image. (b) The HOG feature descriptor computed from (a). (c) Recovered feature vector by our algorithm. (d) The reconstruction error.

Experimental Results

Experimental Results

Settings

- 1.8 GHz PC with 2 GB RAM
- Matlab implementation

ℓ_1 solver

- ℓ_1 magic (second-order cone programming, SOCP)
- Sparse Lab
- Many other packages...

Regressor: Map image feature to pose vector

- Gaussian Process Regressor
- Relevant Vector Regressor [Agarwal, Triggs PAMI '06]
- Support Vector Regressor

Experimental Results

Settings

- 1.8 GHz PC with 2 GB RAM
- Matlab implementation

ℓ_1 solver

- ℓ_1 magic (second-order cone programming, SOCP)
- Sparse Lab
- Many other packages...

Regressor: Map image feature to pose vector

- Gaussian Process Regressor
- Relevant Vector Regressor [Agarwal, Triggs PAMI '06]
- Support Vector Regressor

Experimental Results

Settings

- 1.8 GHz PC with 2 GB RAM
- Matlab implementation

ℓ_1 solver

- ℓ_1 magic (second-order cone programming, SOCP)
- Sparse Lab
- Many other packages...

Regressor: Map image feature to pose vector

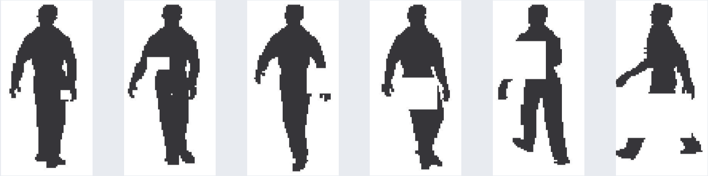
- Gaussian Process Regressor
- Relevant Vector Regressor [Agarwal, Triggs PAMI '06]
- Support Vector Regressor

Robustness to Occlusions

Synthetic data [Agarwal, Triggs PAMI '06]

- 1927 silhouette image for training and 418 images for testing
- Image descriptor: PCA and downsampled images
- Pose vector: 55-dimensional vector describe the joint angle (in degree)

Sample test images

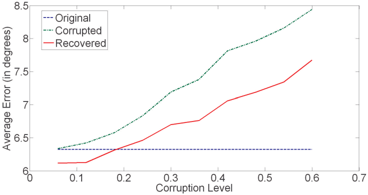


(a) 0.1 (b) 0.2 (c) 0.3 (d) 0.4 (e) 0.5 (f) 0.6

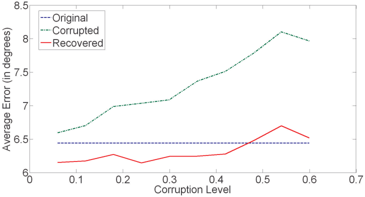
Robustness to Occlusions

Results on synthetic data set

- Estimating pose from original (blue), corrupted (green), and recovered (red) testing images



(a) PCA



(b) Downsample

Figure: Average error of pose estimation on synthetic data set using different features: (a) PCA with 20 coefficients. (b) downsampled (20×20) images.

Localized features are more suitable for error correction!

Robustness to Occlusions

Real Database: HumanEva I [Sigal, Black TR '06]

- Synchronized image and motion capture data
- Four views of four subjects
- Six predefined actions (walking, jogging, gesturing, throwing/catching, boxing, combo)
- We use the common motion walking sequences of three subjects from the first camera

Synthesized testing images

- For each image, we randomly generate two occluding blocks with various corruption level for synthesizing images with occlusions.

Robustness to Occlusions

Real Database: HumanEva I [Sigal, Black TR '06]

- Synchronized image and motion capture data
- Four views of four subjects
- Six predefined actions (walking, jogging, gesturing, throwing/catching, boxing, combo)
- We use the common motion walking sequences of three subjects from the first camera

Synthesized testing images

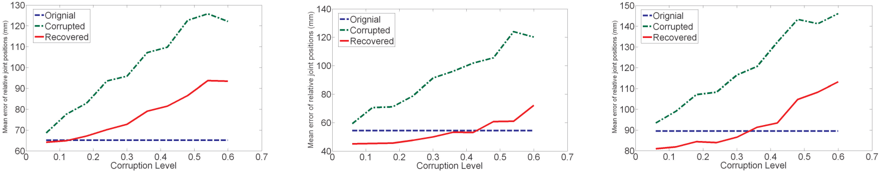
- For each image, we randomly generate two occluding blocks with various corruption level for synthesizing images with occlusions.

Robustness to Occlusions

HumanEva I: Sample test images

(a) 0.1 (b) 0.2 (c) 0.3 (d) 0.4 (e) 0.5 (f) 0.6

- Estimating pose from original (blue), corrupted (green), and recovered (red) testing images



(a) S1 (b) S2 (C) S3

Figure: Results of pose estimation on HumanEva data set I in walking sequences. (a) Subject 1. (b) Subject 2. (c) Subject 3.

Robustness to Background Clutters

Implicit relevant feature selection

- Compared with the original feature representation (HOG), the estimations induced from recovered feature are better.
- The improvements of mean position errors (mm) are 4.89, 10.84, and 7.87 for S1, S2, and S3, respectively

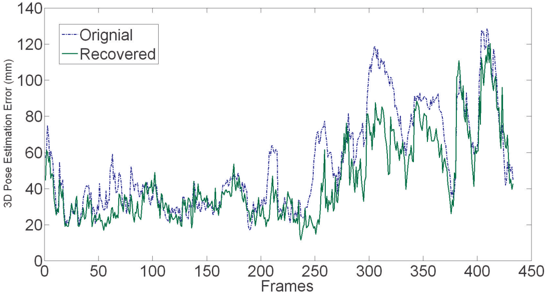


Figure: Mean 3D error plots for the walking sequences (S2).

Conclusions

Conclusions

- Spare approximation can be used for recovering errors resulted from occlusions when estimating humans from occluded images
- Without occlusions, the recovered features are still better than the original ones
- The recovery of feature representation is independent from the learning process, i.e., our method can be adopted as a preprocessing module of other model-free pose estimation approaches

Thank You! Questions?