Human Pose Estimation

A primary survey

By: Nima Pourdamghani
Advisor: Dr. Hamid Reza Rabiee

Outline

- Human Pose Estimation: The Problem
  - Applications
  - Different Approaches
  - Single Camera: Input Features
  - Silhouettes
  - Output Models
  - Performance Measures
- Challenges
  - Depth Ambiguity
  - Other Challenges
- Previous Works
  - Model Based Methods
  - Learning Based Methods
  - Methods Which Reduce Dimensionality
  - Main Surveys
  - Main Labs
- Data Bases
  - Existing Data Bases
  - Problem of ground truth
- My Approach

Human Pose Estimation

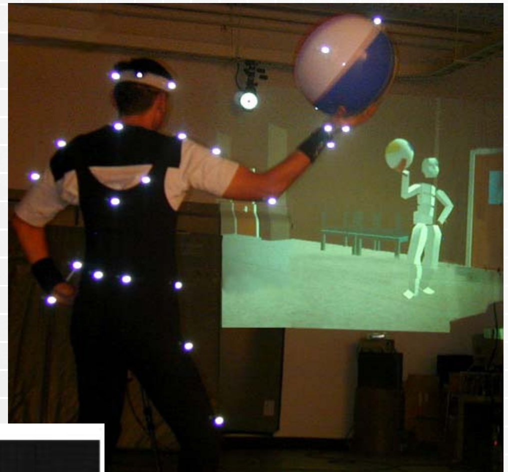
The Problem

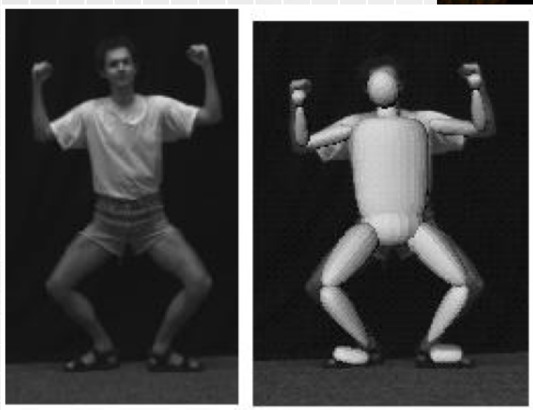
Applications

- It's the problem of inferring a model of human body's state from input data
- Computer Games Industry
  - Realistic animations
  - Human Computer Interaction
- Sports Science
  - Visually analysis the movements
- Physiotherapy
  - Gait analysis
- Video surveillance
  - Motion tracking

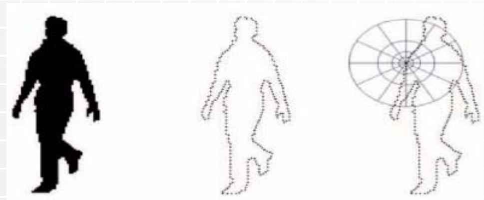
Different Approaches

- De-facto Approaches
  - Easier to infer the pose
  - Expensive
  - Not available in real problems
- Vision Based
  - Single Camera
  - Multi Camera
  - Widely used





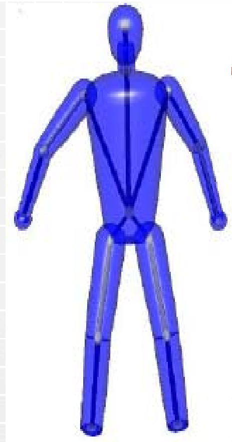
Single Camera: Input Features

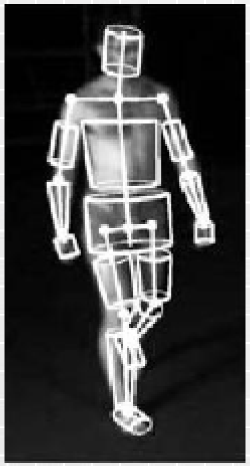


- Most important: silhouette
  - Simple and reliable extraction
    - Assuming robust background
  - Insensitive to irrelevant surface attributes
  - Preserves most of the information
  - Edges
  - Shape contexts
    - **Local** histogram of edges
- Edges
  - Combination of edge and silhouette
- Motions
  - Direct use: HMM, Particle Filters
    - Needs initialization
  - Use into learning, combination with other features
- Colors
- Morphological skeleton
- Center of limbs
- …

Output Models

- ❖ THERE EXISTS NO STANDARD ☹
  - Different applications
- ❖ Kinematic tree model (mostly used)
  - How many joints?
  - How many degree of freedom per joint? (max 3)
  - Dimensionality changes.
- ❖ Volumetric models
  - Represent variations in body size
- ❖ Each model has its own variations





Performance Measures

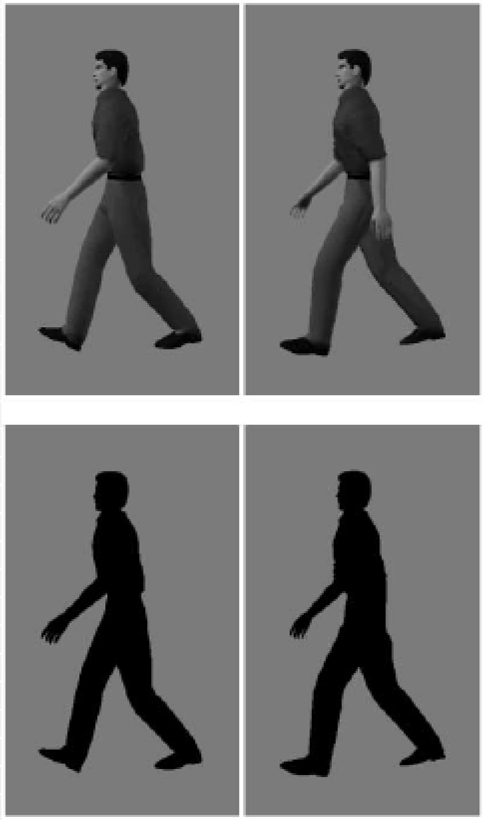
- mean (over all angels) Root Mean Square (RMS) absolute difference errors
  - $D(x,x') = \frac{1}{m}\sum_{i=1}^{m}\left|\left((x_i - x_i' + 180) \bmod 360\right) - 180\right|$
- We don't always have the ground truth
- different applications need different output models
  - No standard
  - Most Comparisons are subjective
  - Most papers didn't compare with any other method

Human Pose Estimation

Challenges

Depth Ambiguity

- The problem is ill posed!
  - 3D world is projected into 2D
- Depth ambiguity is the main challenge
- Use multiple Cameras
- Combine other features
  - motion
  - Limbs' edges
  - Color



Other Challenges

- Self occlusion
- High input dimension
- High output dimension
- General image processing challenges
  - Motion blurs
  - Unconstrained lightening
  - These are usually ignored

Human Pose Estimation

Previous Works

Model Based Methods

- Not Learning Based!
  - Deals with individual frame/video
- Estimate directly from each frame
  - Has many possible solutions
  - Needs to know positions of each body part
- Solve an optimization problem
  - Forward rendering to predict the images
  - Solve over pose variables
  - Expensive
  - Need good initialization
  - Have many local minima
- [7] uses skin color to segment the body into limbs and combine color, shape and contour features for pose estimation (PAMI 2009)
- [8] enforces constant appearance and integrate appearance information besides edges to reliably find body segments. (PAMI 2007)

Learning Based Methods (Supervised)

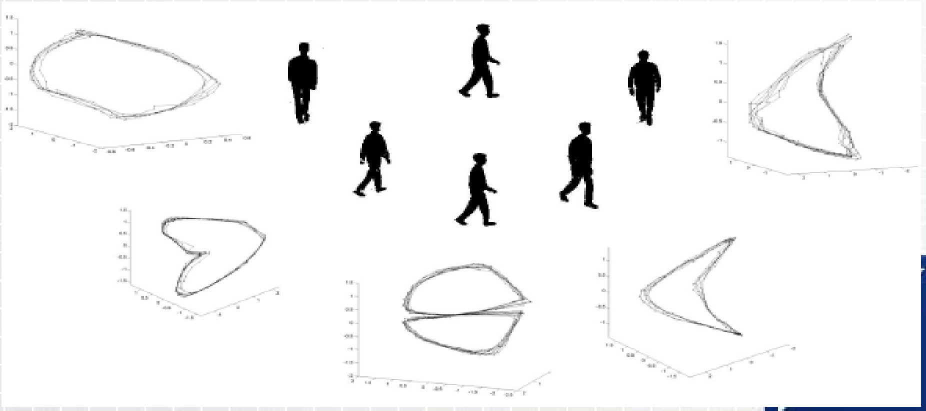
- Little work exists
- Supervised methods
  - [4] learn a perceptron between image and model (ICCV 2000)
  - [5] uses KNN (ICCV 2003)
  - [6] uses RVM (Relevance Vector Machine an extension to SVM) and includes prev. frames to reduce ambiguity (PAMI 2006)
  - [9] combines silhouettes and internal edges and uses K-means (PAMI 2007)

Learning Based Methods (Semi-Supervised)

- There are so few semi-supervised works
  - [2] (ICPR 2008) and [3] (ICML 2007) use Gaussian Process
  - [12,13] presented a new learning method, they learn GMM using semi-supervised techniques and tested the method on human pose estimation
  - [14] uses manifold regularization

Methods Which Reduce Dimensionality

- Reduce data dimensionality prior to estimation
- [10] (CVPR 2004) Used LLE to Learn a nonlinear projection from input space to a low dimensional (3D) manifold space and another projection from manifold space to model space
  - Discuss more: eigenvectors of the Laplacian matrix are good basis for a nonlinear transformation, so graph based methods might implicitly do this work.
- [11] (ICCV 2007) extends [10] for other activities than gait



Main Surveys

- Most (almost all) surveys are on human motion tracking
- Here are some good ones:
  - Hen YW, Paramesran R. **"Single camera 3D human pose estimation: A Review of current techniques."** In: *2009 International Conference for Technical Postgraduates (TECHPOS)*. IEEE; **2009**:1-8.
  - Sminchisescu C. **"3D Human Motion Analysis in Monocular Video Techniques and Challenges."** *COMPUTATIONAL IMAGING AND VISION*. **2008**;36:185.
    - Cited by 9
  - Poppe R. **"Vision-based human motion analysis: An overview."** *Computer Vision and Image Understanding*. **2007**;108(1-2):4-18.
    - Cited by 115
  - Moeslund TB, Hilton A, Krüger V. **"A survey of advances in vision-based human motion capture and analysis."** *Computer Vision and Image Understanding*. **2006**;104(2-3):90-126.
    - Cited bye 505
  - Moeslund T. **"A Survey of Computer Vision-Based Human Motion Capture."** *Computer Vision and Image Understanding*. **2001**;81(3):231-268.
    - Cited bye 871

Main Labs

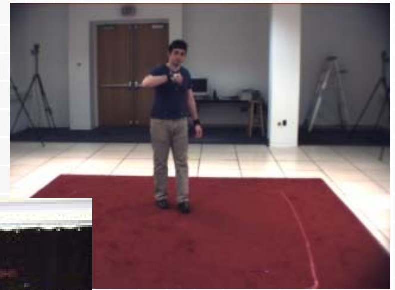
- Most Labs work on human motion tracking or activity recognition
- CMU Graphics Lab (mocap)
  - CMU motion capture database
- CMU, Entertainment Technology Center (ETC) Project
  - The Master Motion project is using an optical motion capture system and wireless VR to explore how virtual reality can be used to learn physical movement.
- University of Washington, Department of Computer Science & Engineering, motion capture lab
  - Its primary purpose is to advance current cutting-edge research in computer animation tools and techniques.
- Vision lab in Brown university
  - HumanEva

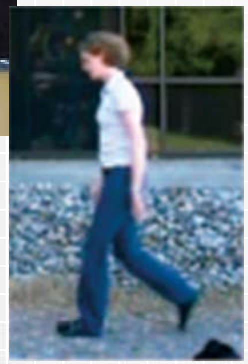
Human Pose Estimation

Databases

Existing Databases

- HumanEva (I & II)
  - From vision lab of Brown university
  - 7 calibrated video sequences
  - 4 subjects performing a 6 common actions
  - Contains **training**, validation and testing sets
- CMU Graphics Lab Motion Capture Database
  - For activity recognition
  - De-facto
  - Contains many activities
- HEDVIG KJELLSTRÖM Database
  - Walking straight, Walking in a circle
- Georgia Tech, GVU Center/College of Computing
  - Gait (walking) data
- There are others but…
  - HumanEva is the best
  - Hedvig ‘s is commonly used





Problem of Ground Truth

- Finding the ground truth for pose estimation is hard
- Most databases don’t have training set (except for humaneva)
- A trick: Use Poser
  - A software for creating realistic animations.
  - This way we will have clean silhouettes and precise training data
  - Some other works including [6] (PAMI 2006) have used poser.

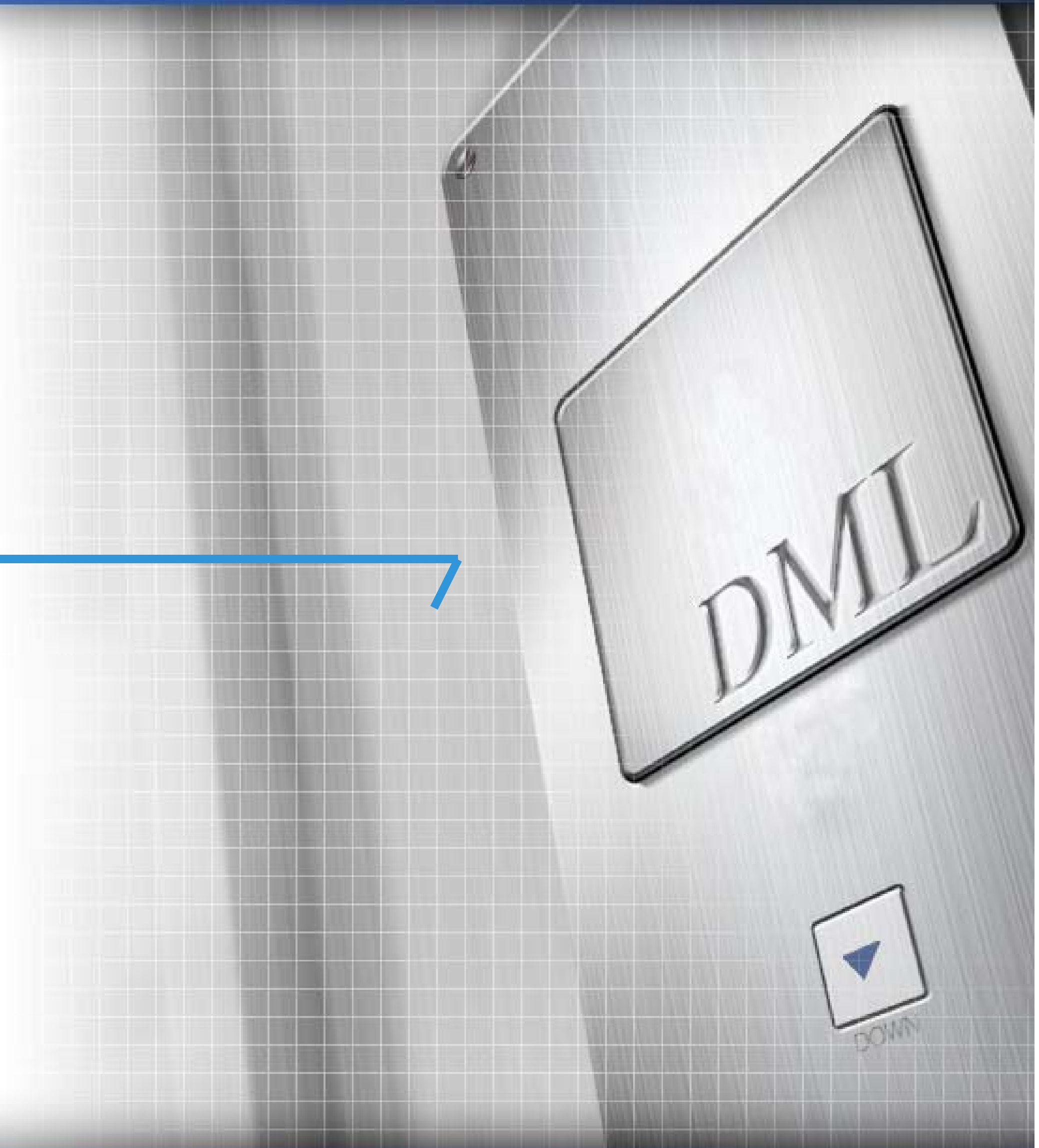
Human Pose Estimation

My Approach

My Approach

- Use silhouette plus some other feature to reduce ambiguity as input data
  - Options: prev. frames, internal edges, center of limbs
- Use a Kinematic tree model as output
- Use POSER or HumanEva as training data
- Use HumanEva or/and Some other dataset as test data
- Learn a regression function (probably using graph based methods) over training data
  - The work done by [10,11] proves the existence of a well-formed manifold over the data
- Compare to some method like [6]

Q&A



Referrences

- [1] Hen YW, Paramesran R. Single camera 3D human pose estimation: A Review of current techniques. In: *2009 International Conference for Technical Postgraduates (TECHPOS)*. Ieee; 2009:1-8.
- [2] Zhao X, Ning H, Liu Y, Huang T. Discriminative estimation of 3D human pose using gaussian processes. In: *19th International Conference on Pattern Recognition, 2008. ICPR 2008*. Ieee; 2008:1–4.
- [3] Ek C, Torr P, Lawrence N. Gaussian process latent variable models for human pose estimation. In: *Proceedings of the 4th international conference on Machine learning for multimodal interaction*. Springer-Verlag; 2007:132–143.
- [4] Rosales R, Sclaroff S. Inferring body pose without tracking body parts. In: *IEEE conference on Computer Vision and Pattern Recognition (cvpr)*.; 2000.
- [5] Shakhnarovich G, Viola P, Darrell T. Fast pose estimation with parameter-sensitive hashing. In: *Proceedings Ninth IEEE International Conference on Computer Vision*. Ieee; 2003:750-757 vol.2.
- [6] Agarwal A, Triggs B, Rhone-Alpes I, Montbonnot F. Recovering 3D human pose from monocular images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*. 2006;28(1):44–58.
- [7] Lee MW, Nevatia R. Human pose tracking in monocular sequence using multilevel structured models. *IEEE Transactions on Pattern Analysis and Pattern Recognition(PAMI)*. 2009;31(1):27-38.
- [8] 1. Ramanan D, Forsyth D, Zisserman A. Tracking people by learning their appearance. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*. 2007;29:pp65-81.
- [9] 1. Sminchisescu C, Kanaujia A, Metaxas DN. BMA^{3}E : Discriminative Density Propagation for Visual Tracking. *IEEE Transactions on Pattern Analysis and Pattern Recognition(PAMI)*. 2007;29(11):2030-2044.
- [10] Elgammal a. Inferring 3D body pose from silhouettes using activity manifold learning. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* 2004:681-688.
- [11] Lee C, Elgammal A. Modeling View and Posture Manifolds for Tracking. *2007 IEEE 11th International Conference on Computer Vision*. 2007:1-8.
- [12] 1. Navaratnam R, Fitzgibbon A, Cipolla R. Semisupervised learning of joint density models for human pose estimation. In: *Proc. British Machine Vision Conference BMVC.*; 2006:679-688.
- [13] 1. Navaratnam R, Fitzgibbon AW, Cipolla R. The Joint Manifold Model for Semi-supervised Multi-valued Regression. In: *2007 IEEE 11th International Conference on Computer Vision*. Ieee; 2007:1-8.
- [14] 1. Li Y, Jia K, Zhang G. Semi-Supervised Human Pose Estimation Piloted by Manifold Structure. In: *2009 International Conference on Information Engineering and Computer Science*. Ieee; 2009:1-4.