3D Human Pose Estimation and Tracking from Silhouette Images

Leong Mei Chee (IMI)

Supervisor : Assoc Prof Lee Yong Tsui (MAE)
Co-Supervisor : Assoc Prof Lin Feng (SCSE)

14th Feb 2017

NANYANG TECHNOLOGICAL UNIVERSITY

Outline

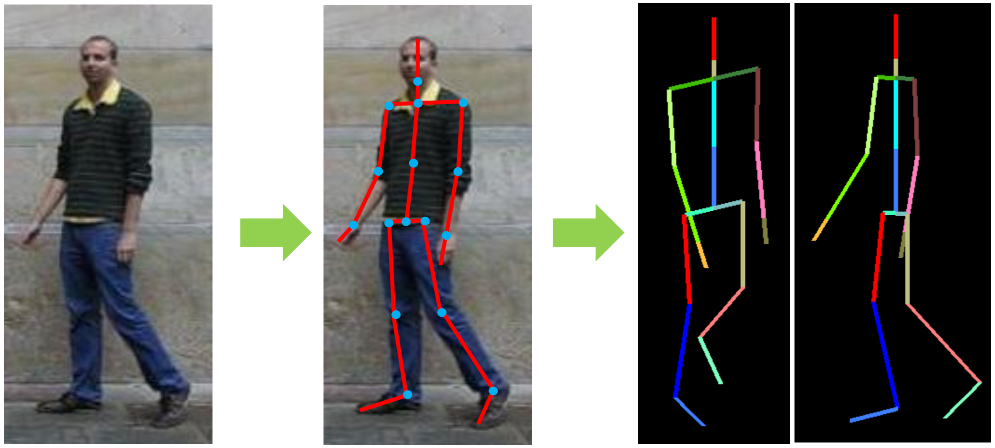
- Introduction and Background
- Literature Review
- Experimental Work
- Results
- Summary

3D Human Pose Estimation from A Single Image

NANYANG TECHNOLOGICAL UNIVERSITY

Introduction

- **3D human pose estimation**, the process of estimating the configuration of the human structure from 2D images.
- Applications in surveillance, entertainment, sports science and healthcare

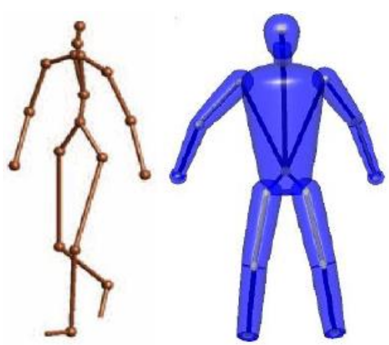


NANYANG TECHNOLOGICAL UNIVERSITY

Approaches for 3D Human Pose Recovery

Model-based method

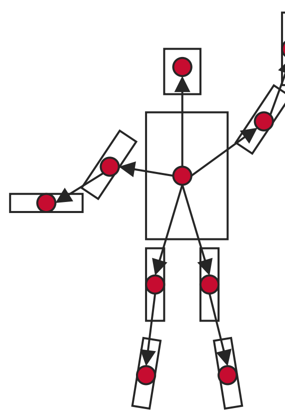
- Has an underlying human model
- Optimizes similarities between estimated pose projection and the given 2D image



(Hen and Paramesran 2009)

Deformable part model

- Assembly of body parts by detecting body parts location and impose spatial relation between parts



(Felzenszwalb and Huttenlocher 2005)

Model-free method

- Example-based and learning-based methods
- Learn mapping functions between 2D image and the 3D pose in a dataset

(Agarwal and Triggs 2004)

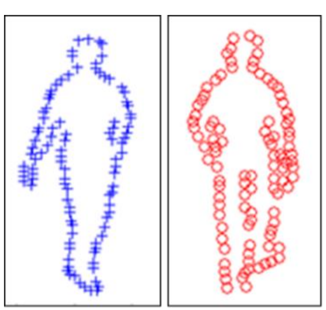
NANYANG TECHNOLOGICAL UNIVERSITY

Example-based Method

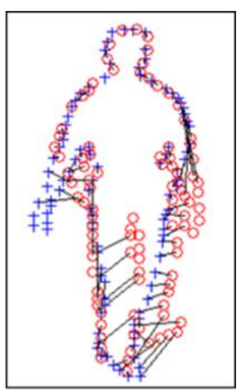
Feature Detection (Shape Context)

Feature matching to find nearest neighbors

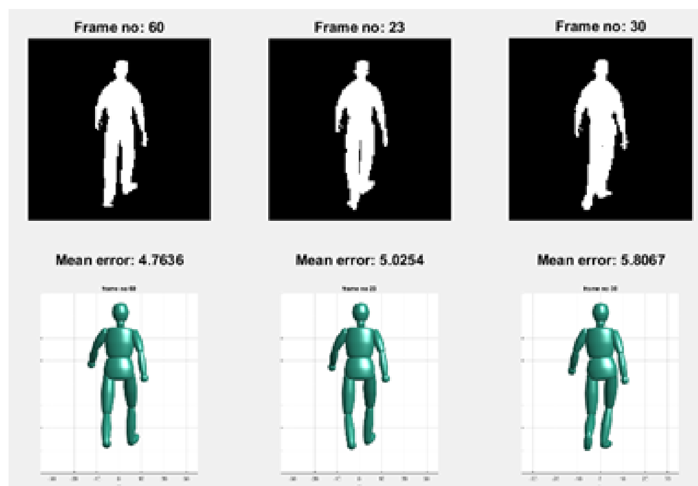
Averaging poses for nearest neighbors



Feature detection



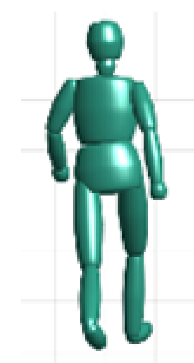
Feature matching



Nearest neighbors with corresponding 3D poses retrieved

Mean error: 4.51

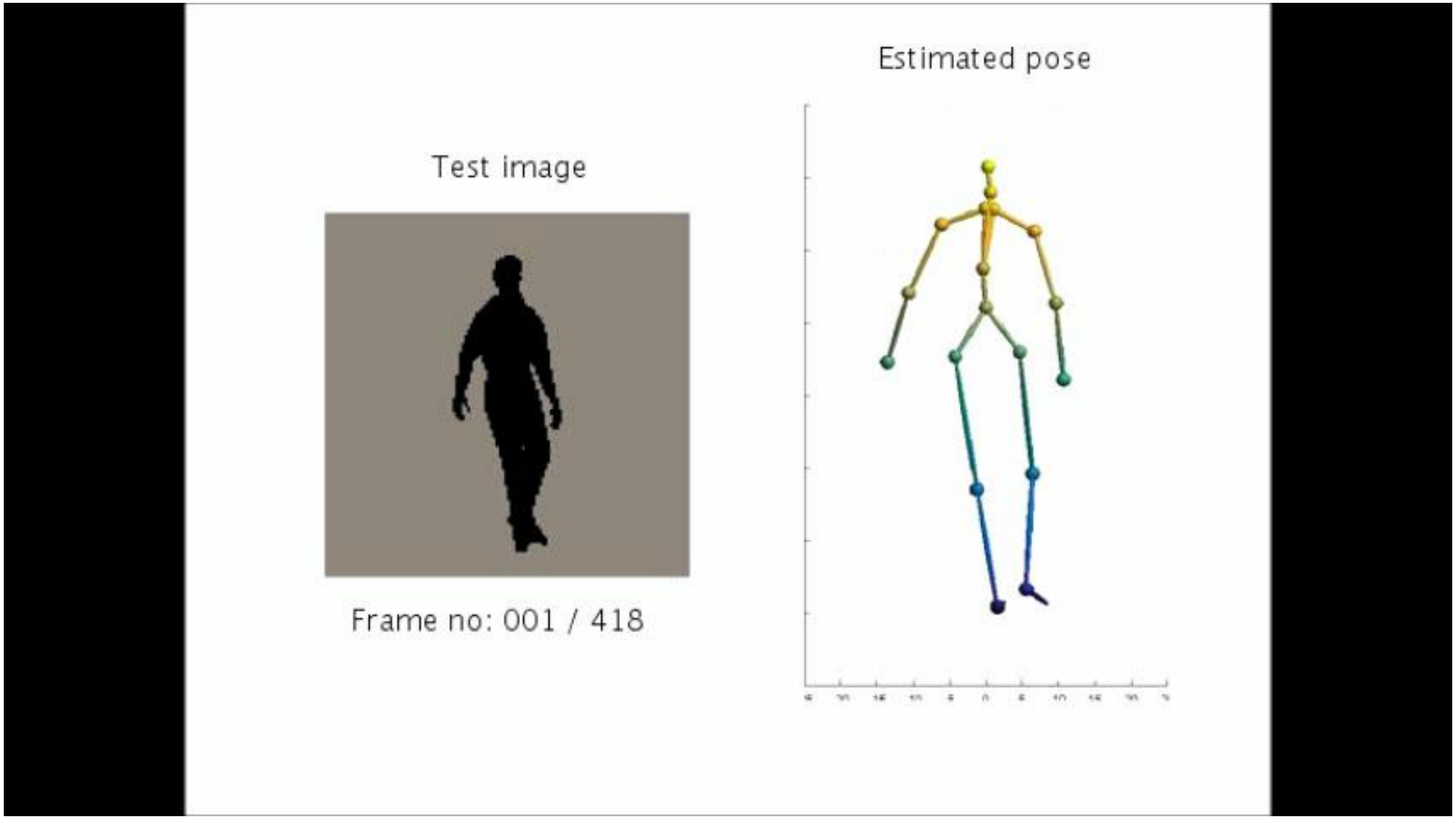
Estimated pose



Ground truth pose

NANYANG TECHNOLOGICAL UNIVERSITY

Result



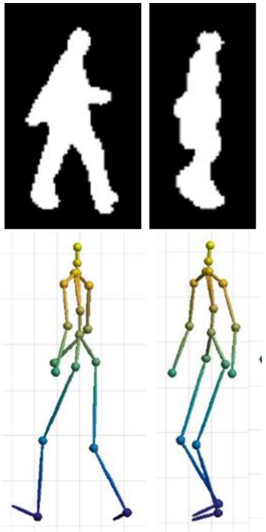
Sample result for 3D pose recovery from silhouette images

NANYANG TECHNOLOGICAL UNIVERSITY

Problems

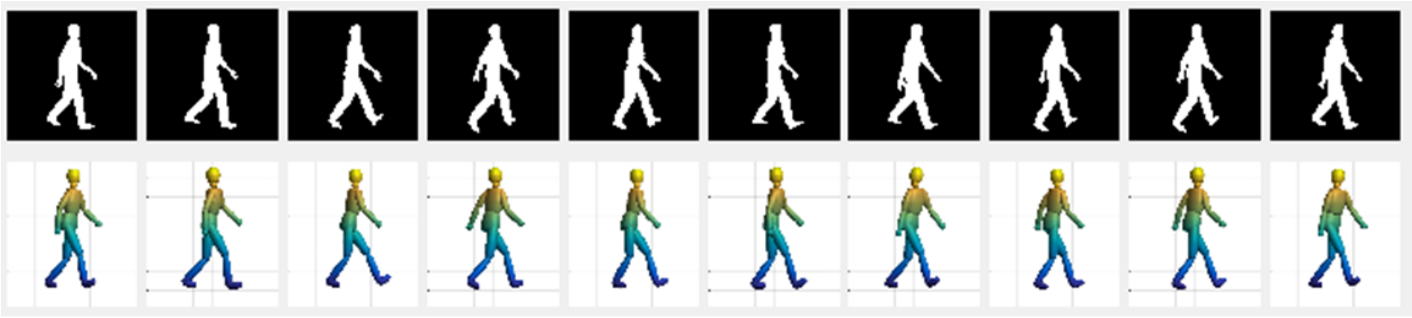
1. Projection ambiguity

- Ambiguity of a 2D projection image that can be mapped to more than one feasible 3D pose



2. Interpolation of poses

- Interpolation of ambiguous poses lead to counterbalance of the joint angles, results in wrongly estimated pose.



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Top 10 retrieved nearest neighbors with two variations of poses observed. One with right shoulder and left leg in front (c, d, e, h, i); while another with left shoulder and right leg in front (a, b, f, g, j).

3D Human Pose Tracking for A Sequence of Images

NANYANG TECHNOLOGICAL UNIVERSITY

Tracking Process

- Ensure smooth motion and recover the wrongly estimated pose
- Reconstructed pose from the current frame can be used for pose prediction in the next frame
- Temporal tracking can be used to resolve occlusion problems

Literature Review

Dynamic spatio-temporal modeling for example-based human silhouette recovery (X. Zhou and X. Li, 2015)

- Aims to effectively recover the original human silhouette signals from noisy corruption or partial occlusion by investigating in both spatial and temporal dimensions.
- A shape model is built to capture the spatial structural information on human silhouettes by sparseness constrained nonnegative matrix factorization (NMF)-based local feature learning.

$$\min_{W,H} \| V - WH \|_F$$

$$\text{s.t. } W \geq 0, H \geq 0$$

under optional constraints

$$sparseness(\mathrm{w}_i) = S_w, \quad \forall i$$

$$sparseness(\mathrm{h}_i) = S_h, \quad \forall i$$

Global basis images

Local basis images

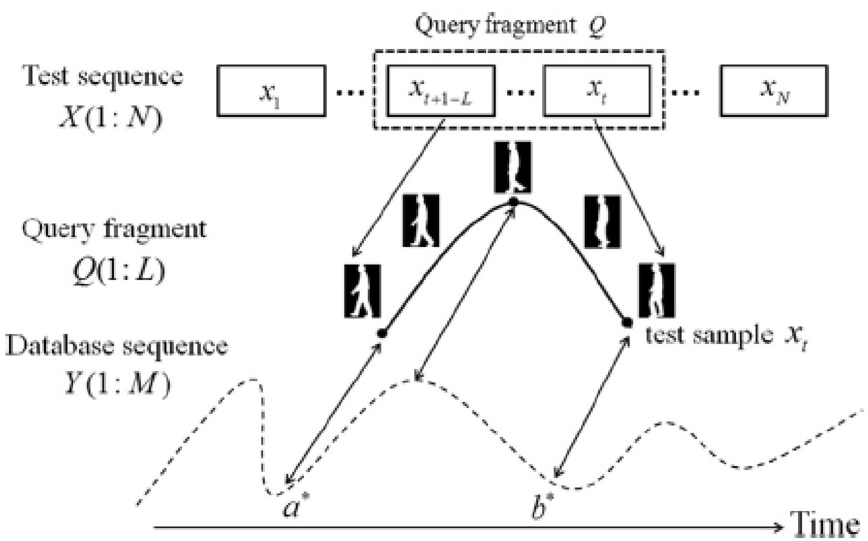
Recovered images for corrupted samples

NANYANG TECHNOLOGICAL UNIVERSITY

Literature Review

Dynamic spatio-temporal modeling for example-based human silhouette recovery (X. Zhou and X. Li, 2015)

- The underlying temporal correlations among adjacent silhouette frames are discovered by solving an adaptive time- series data alignment optimization problem using dynamic time warping (DTW).



x_t is reconstructed by a weighted combination of these exemplars

$$\hat{x}_t \approx \sum_{i=1}^{r} w_i y_{b_i^*}$$

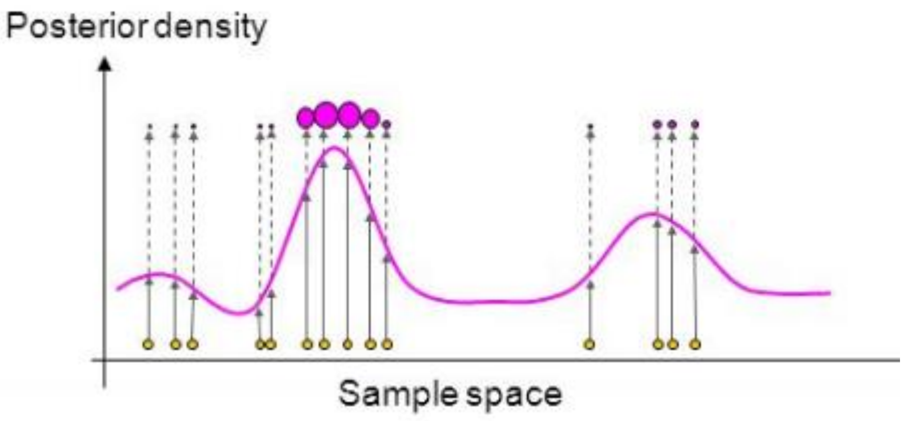
Optimal time alignment of query fragment Q with a subsequence within Y, from a* to b*. The samples associated with each ending point b* are used as the nearest exemplars for recovery.

NANYANG TECHNOLOGICAL UNIVERSITY

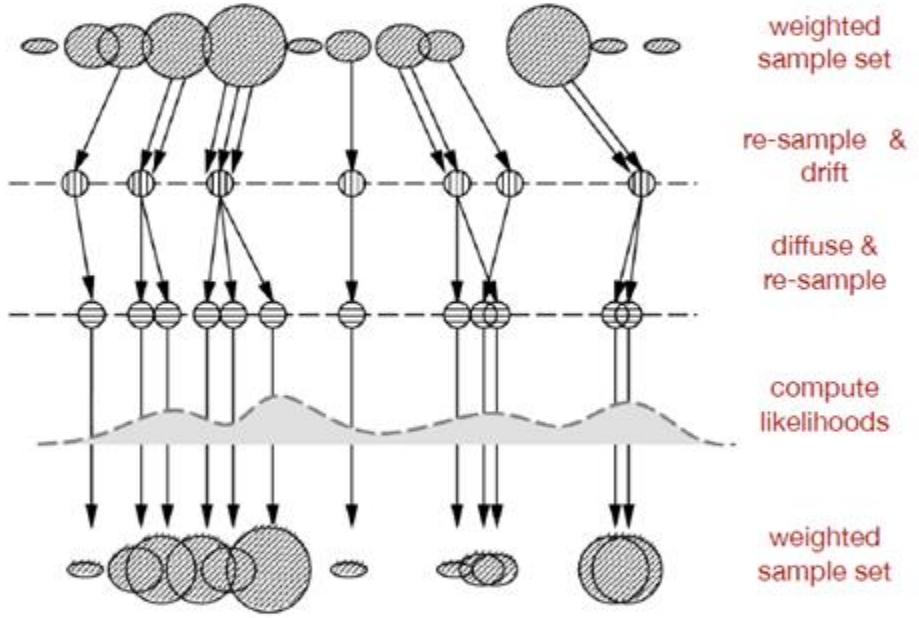
Literature Review

3D Human motion tracking by exemplar-based conditional particle filter (J. Liu, D. Liu, J. Dauwels, H.S. Seah, 2015)

- Particle filter – to represent the posterior density by a set of random particles with associated weights.



Particle filtering operation (Miodrag Bolic, 2004)



Prediction distributions (Isard and Blake, 1998)

NANYANG TECHNOLOGICAL UNIVERSITY

Literature Review

3D Human motion tracking by exemplar-based conditional particle filter (J. Liu, D. Liu, J. Dauwels, H.S. Seah, 2015)

- Introduce a conditional term with respect to exemplars and image data, to approximate a dynamic model and used to predict current states of particles in prediction phase.

Prediction Stage:

1. Compare current frame with dataset to find best examplar for 2D pose estimation. Then, reconstruct 3D pose.
2. Find the motion transition function:

$$\hat{\mathbf{x}}_{matched,t} = f_{exemplars,\mathbf{I}_{1:t}}(\hat{\mathbf{x}}_{t-1})$$

3. Prediction for each particle:

$$\hat{\mathbf{x}}_t^{(i)} = f_{exemplars,\mathbf{I}_{1:t}}(\hat{\mathbf{x}}_{t-1}^{(i)})$$

Literature Review

3D Human motion tracking by exemplar-based conditional particle filter (J. Liu, D. Liu, J. Dauwels, H.S. Seah, 2015)

- In update phase, weights of particles are refined by matching images with projected human model using a set of features.

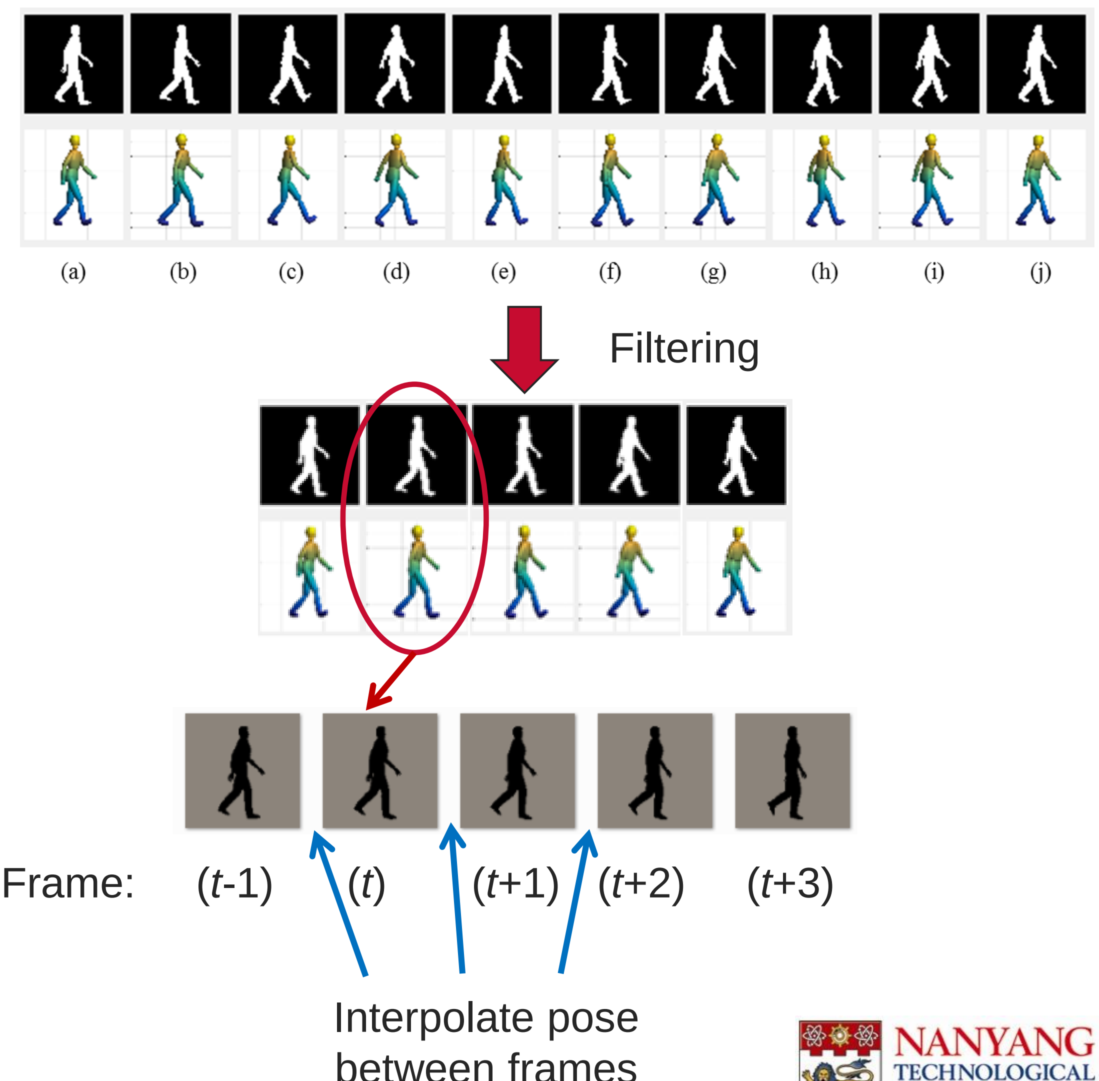
Update Stage:
1. Measure Sum of Squared Difference (SSD) for edge and silhouette features
2. Calculate weights of each particle according to likelihood measurement function

- This method retains advantages of classic particle filters while increases prediction accuracy which constrains evolved particles within an area closer to true state.

Experimental Work

Prediction Stage:

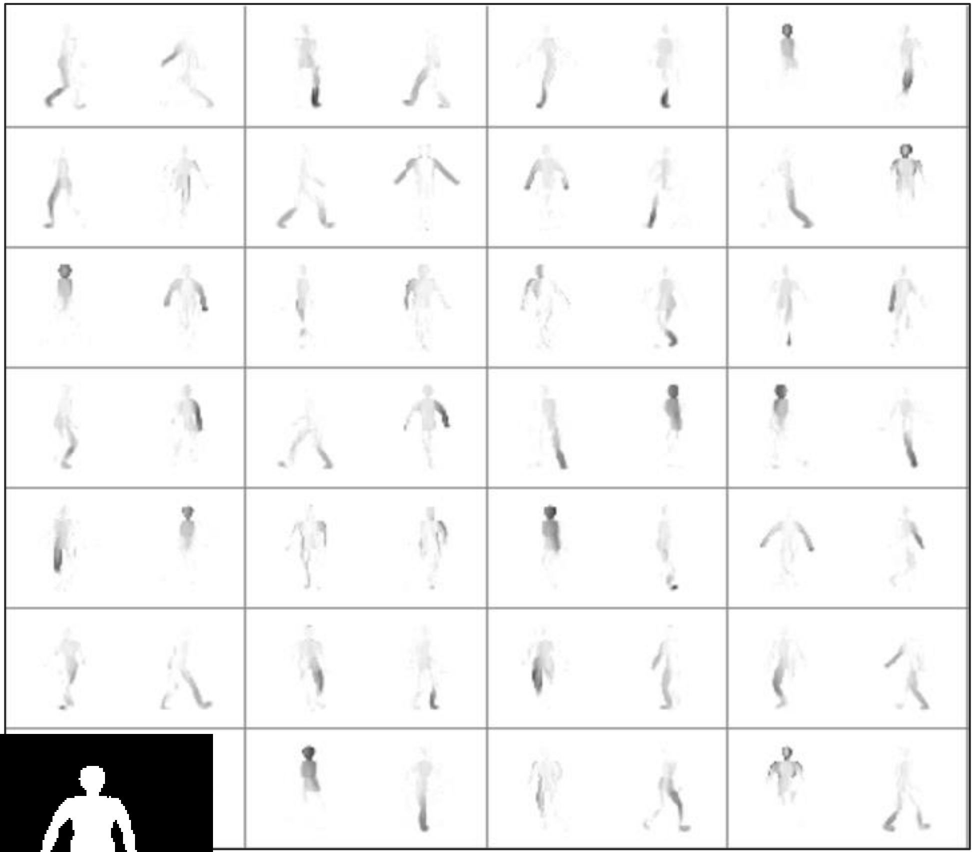
1. From current frame image, find nearest neighbors by comparing Shape Context.
2. Filter 3D poses where $x_i(t) - x_i(t\text{-}1) >$ angle constraint for joint i
3. For each pose, find frames from (t-1) to (t+3) and interpolate them as particle prediction.



NANYANG TECHNOLOGICAL UNIVERSITY

Experimental Work

Update Stage:
1. Find sparse representation for current image.
2. Compare sparse representation to 2D projections from particles to calculate weights.

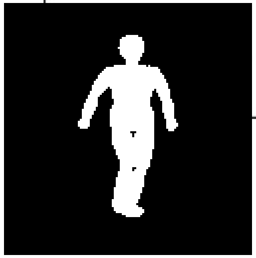


Sample basis images

Query image

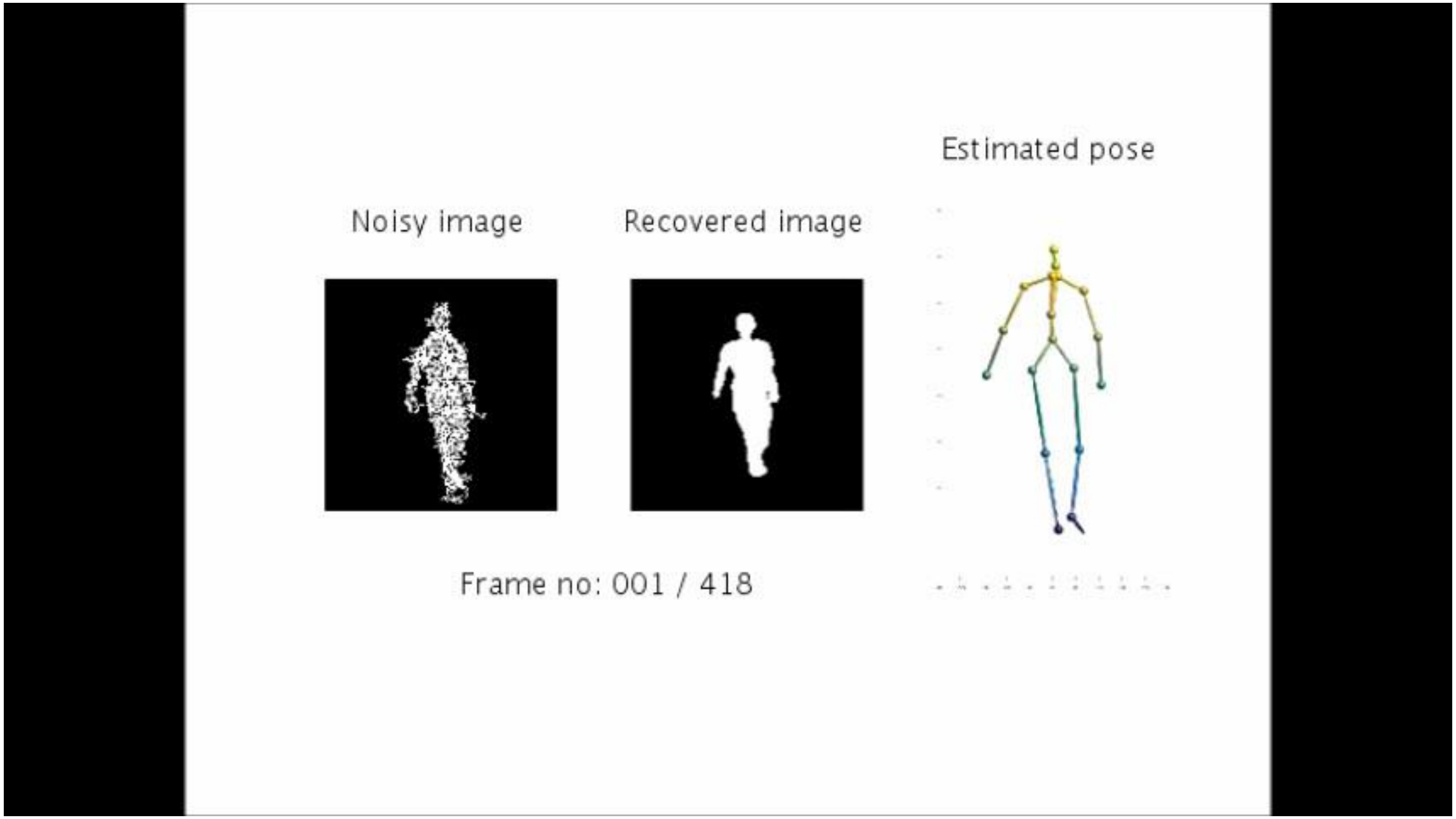
Sparse representation

compare



2D projection

Results



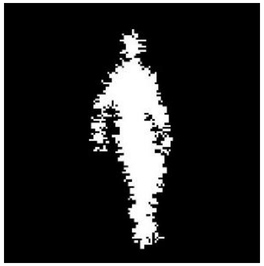
Sample result for 3D pose tracking from noisy images

Summary

1. 3D pose estimation with example-based approach. Filter ambiguous poses by comparing joints angle with previous frame.

2. Use frame sequence and interpolated poses as importance sampling for particles – to constraint particles within true space and obtain feasible poses.

3. Use sparse representation – to recover noisy image, can be applied to different body sizes

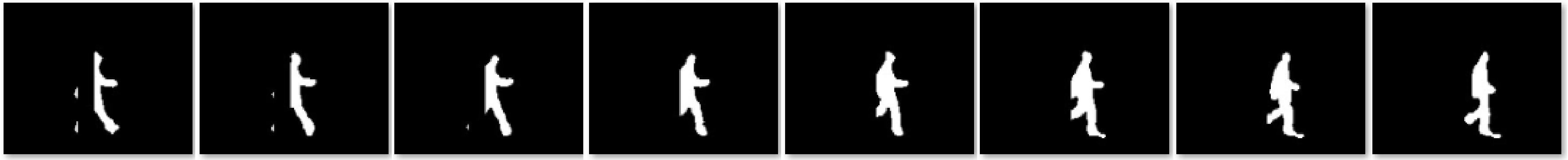


Noise image

Recovered image with sparse representation

Future Work

- Build sparse representation based on shape context feature (applicable to different image size)
- Work on different dataset (HumanEVA dataset) and subjects (CMU Motion Capture Database)
- Work on recovering occluded images



Sample images from Space Time Action dataset (M. Blank, et al, 2005)

THANK YOU

NANYANG
TECHNOLOGICAL
UNIVERSITY