3D Human Pose Search using Oriented Cylinders

Selen Pehlivan
Pınar Duygulu

Department of Computer Engineering
Bilkent University

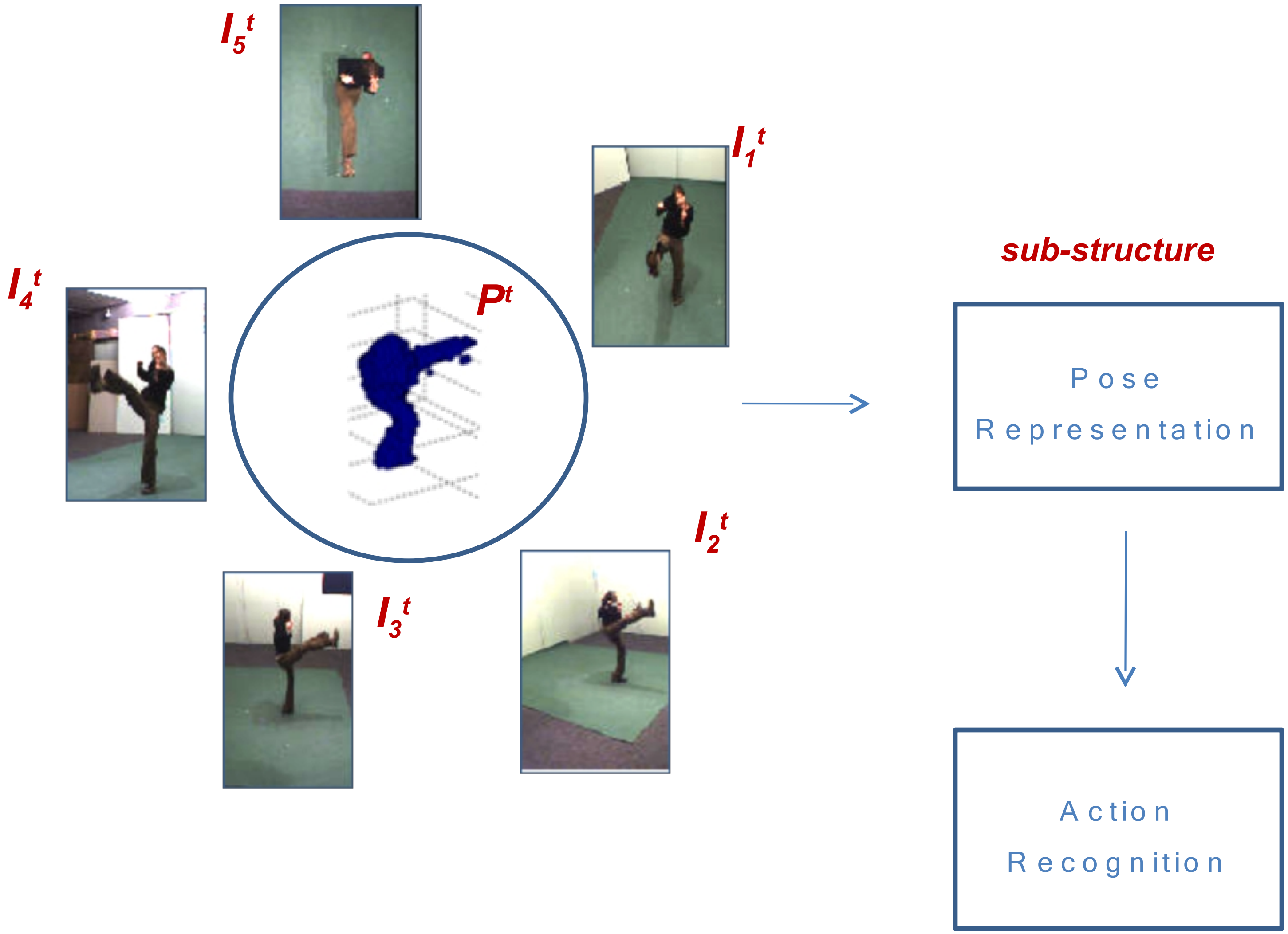
Problem

- Human action recognition
  - Applications
    - Surveillance, monitoring, HCI, games

- Multi-camera systems
  - Poses as volumes [*Weinland et al. 06, Chen et al. 03, Cohen et al. 03, Huang et al. 05, Pierobon et al. 05, Lv et al. 07*]

Human Poses

- Different than rigid body objects [*Ankerst et al. 99, Kazhdan et al. 03, Johnson et al. 03*]
- Articulated structure
- High number of potential configuration
- A compact representation


I_5^t
I_1^t
I_4^t
P^t
I_2^t
I_3^t
sub-structure
Pose Representation
Action Recognition

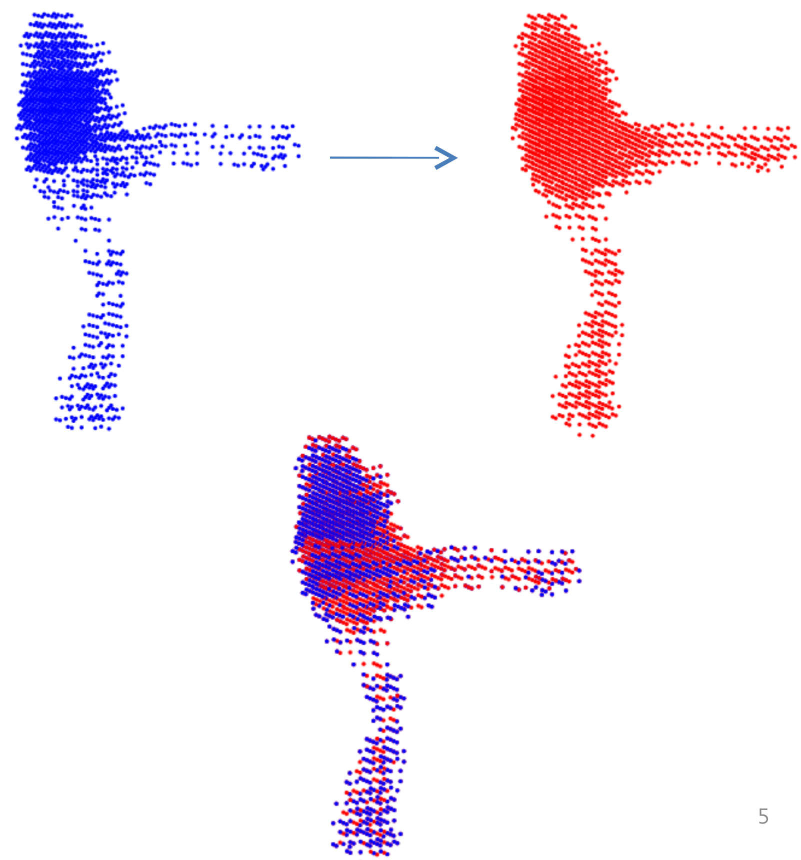
3D Poses

Reconstruction

- Obtain a visual hull
- In the form of voxel grid
- Shape from silhouette technique
- EPVH [*Franco and Boyer 03*]

Enhancement

- Perform morphological closing
- Sphere



3D Poses → Representation

- body parts look like cylinders [*Bilford 71* , *Marr* and *Nishihara 78*]

- varying
  - size
  - orientation

Approach

Step 1 Form Cylindrical Filters

Step 2 Search over 3D Pose

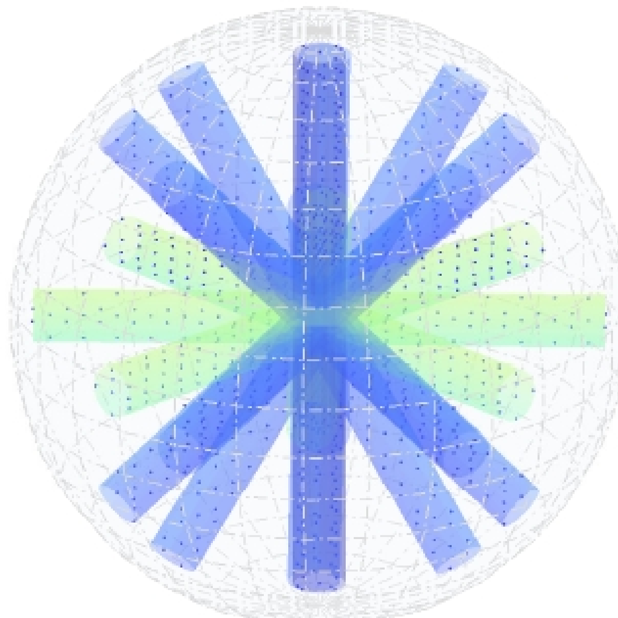
Step 3 Select High Response Regions

Step 4 Form Histograms

Pose Representation
Step 1 Form Cylindrical Filters

- 3 D 0-1 Filters
- Filter set K per cylinder size
  - Cylinder [r x h]
  - Rotate α° apart around local axis
  - Let K be the filter set , the number of filters in K with α° apart:

$$|K| = 1 + (\frac{90}{\alpha} - 1)(\frac{360}{\alpha}) + (\frac{180}{\alpha})$$

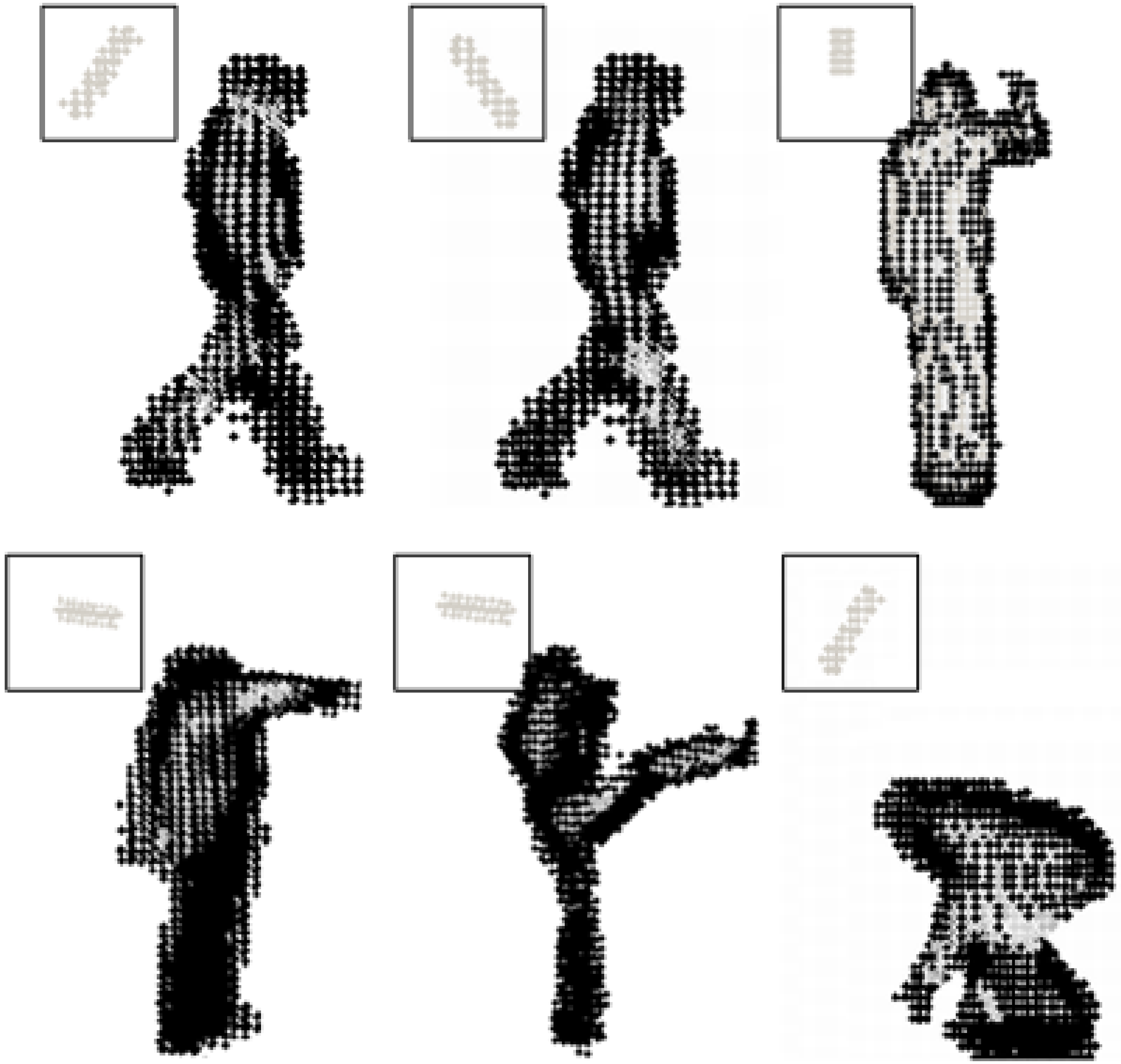


- 3D filters are the grid located inside cylinders

Pose Representation
Step 2 Search over 3D Pose

- Convolve with 3D pose

Search Results



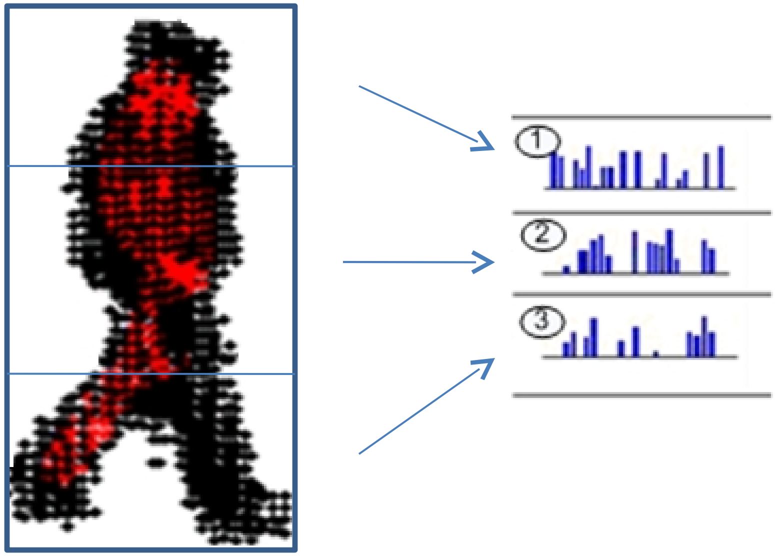
Pose Representation
Step 3 Select High Response Regions

- High response regions
  - body parts with the same orientation

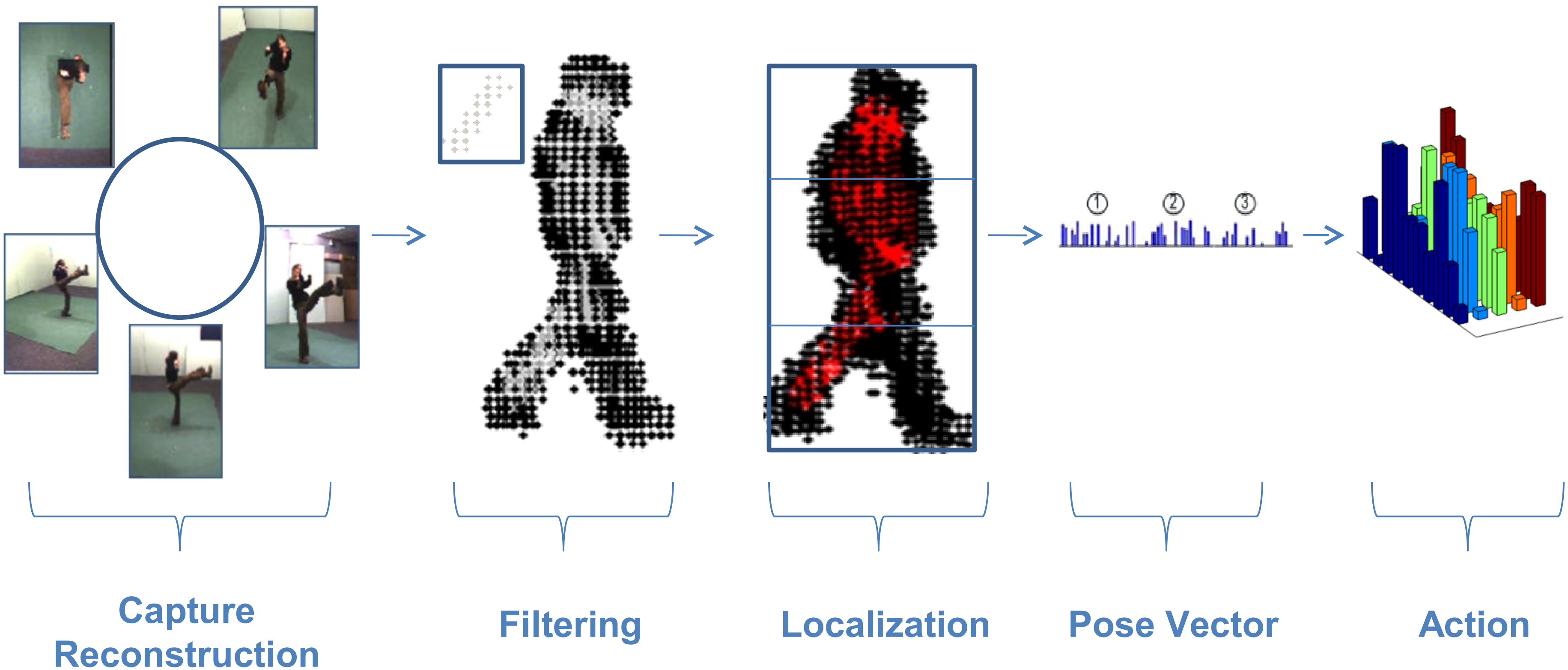
- Scale responses to the range of [0, 1]
- Select voxels with a score greater than a threshold

Pose Representation
Step 4 Form Histograms

- Divide into N sub-volumes
- Form histograms of cylinders with a given size and orientation
- Then combine N histograms
- Normalize the feature vector

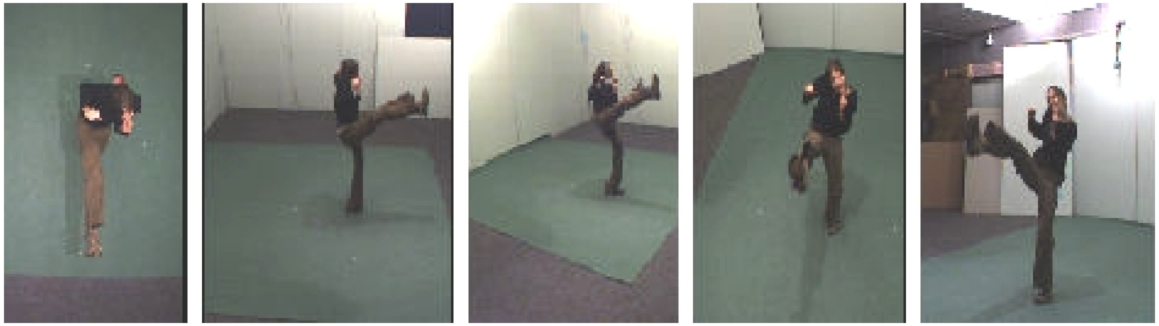


Pose Representation
Distribution of Oriented Cylinders



Dataset

- IXMAS Dataset with 5 cameras [*Weinland et al. 06*]
- 5 actions
  - Walk
  - Wave
  - Kick
  - Punch
  - Pick up
- [64x64x64]



Pose Retrieval

Method 1 NN-based Classification

[r x h]
[1x5]
[1x10]
[1x20]

$\alpha°$
30

T
0.8

N
3

Poses	Accuracy All	Accuracy [1 x5]	Accuracy [1x10]	Accuracy [1x20]
Walk	96.97	93.94	96.97	96.97
Wave	90.91	90.91	93.94	81.82
Punch	63.64	48.48	69.70	72.73
Kick	87.88	84.85	84.85	75.76
Pick	96.97	90.91	96.97	93.94

- Euclidean distance
- Leave one out

Pose Retrieval
Method 2 SVM-based Classification

[r x h]
[1x5]
[1x10]
[1x20]

α°
30

T
0.8

N
3

Poses	Accuracy All	Accuracy [1 x5]	Accuracy [1x10]	Accuracy [1x20]
Walk	100	87.50	91.67	95.83
Wave	91.67	54.17	95.83	79.17
Punch	66.67	66.67	66.67	45.83
Kick	100	100	100	95.83
Pick	95.83	95.83	95.83	95.83

- Multi-class SVM
- RBF kernel
- Training set: 3actors

Action Recognition

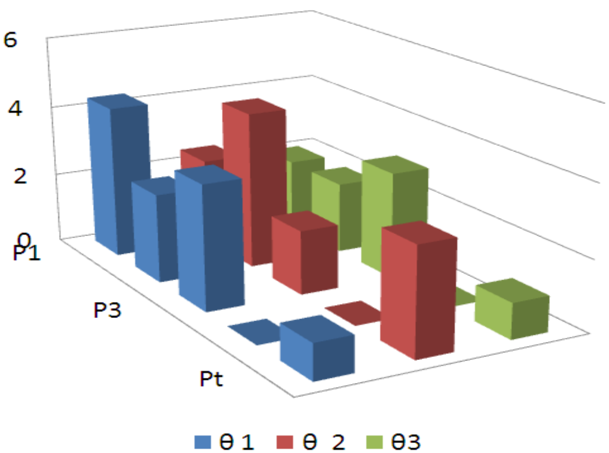
Method 1 Dynamic Time Warping[Rabiner et al. 93]

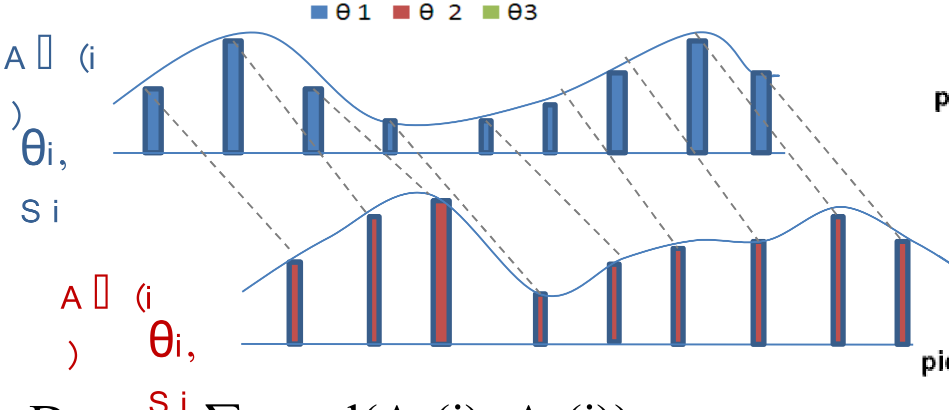
Method 2 Hidden Markov Model[Rabiner et al. 89]

Action Recognition
Method 1 Dynamic Time Warping

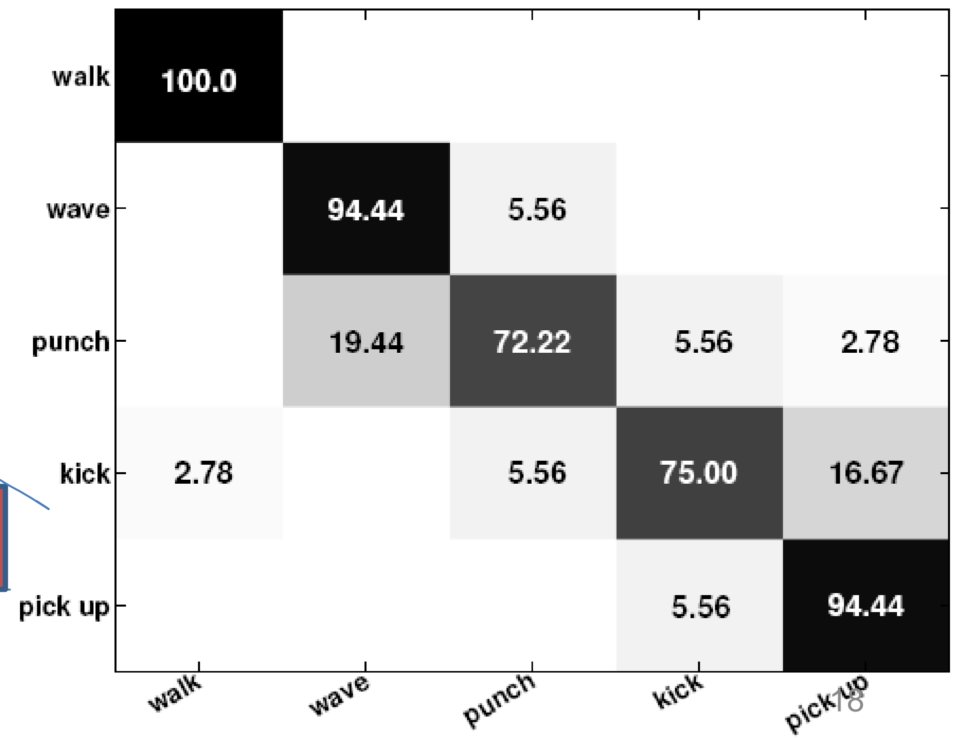
- Use the approach of [Duygulu and Ikizler 07]

[r x h]	α°	T	N
[1x5]	30	0.8	3
[1x10]			
[1x20]			





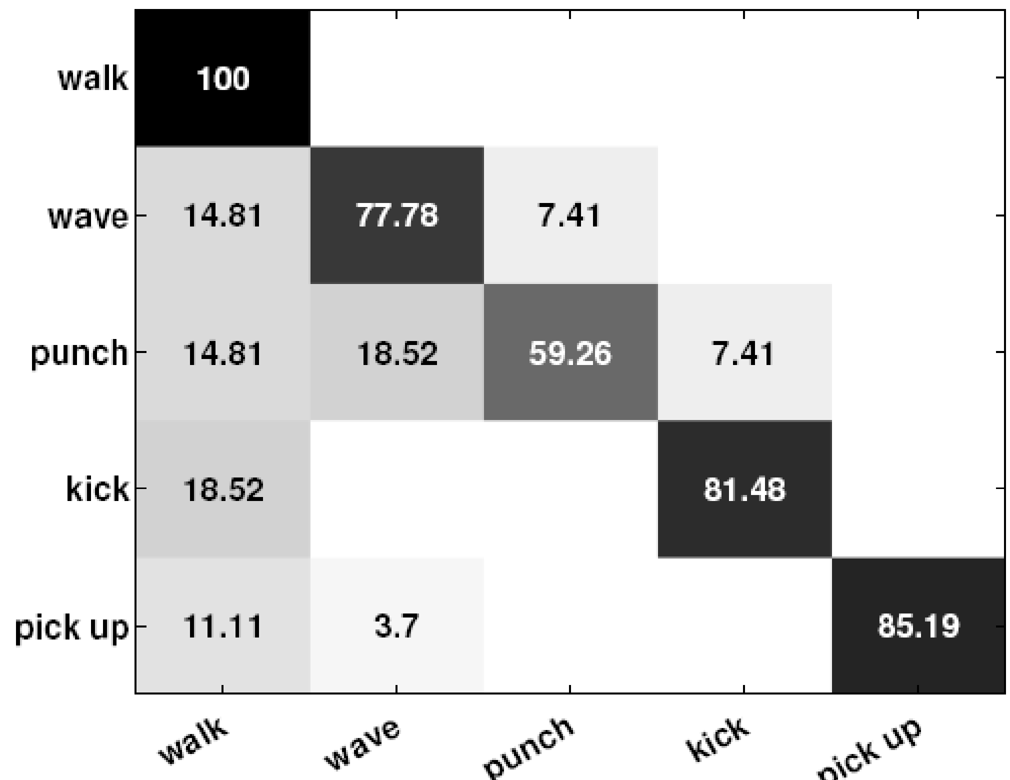
$$D_{global} = \sum_{i=1..n} d(A_1(i), A_2(i))$$



Action Recognition
Method 2 Hidden Markov Model

- Cluster into 80 pose-words
- 3 states

[r x h]	α°	T	N
[1x5]	30	0.8	3
[1x10]			



DTW vs. HMM

Poses	DTW All	DTW [1 x5][1x10]	HMM All	HMM [1x5][1x10]
Walk	97.22	100	100	100
Wave	94.44	94.44	77.78	62.96
Punch	69.44	72.22	59.26	70.37
Kick	66.67	75	81.48	62.96
Pick	91.67	94.44	85.19	74.07

Other Works

Weinland et al.	93.33%
Liu et al.	82.8%
Weinland et al.	81.27%
Lv et al.	80.6%
Yan et al.	78.0%

Conclusion & Discussion

- Samples are enough to handle viewpoint variations
- We can obtain better results with denser data
- We observe that cylinders with different lengths more robust than with different radiuses

Future work:

Experiments on

- more action categories
- different resolutions of volumes

Questions

?