ARTIFICIAL INTELLIGENCE(CS365)

3D ACTION RECOGNITION USING EIGEN-JOINTS

Kranthi Kumar, Prashant Kumar

Supervisor: Dr. Amitabha Mukerjee

Dept. of Computer Science and Engineering

PROBLEM STATEMENT

- To recognize human actions using 3D skeleton joints recovered from 3D depth data.
- 3D depth data is captured using RGB-D cameras such as Microsoft Kinect.

MOTIVATION

- Human activity recognition is one of the important problem in computer vision.
- It has uses in the fields of video surveillance, human-computer interaction, etc.
- Health Care.

MOTIVATION

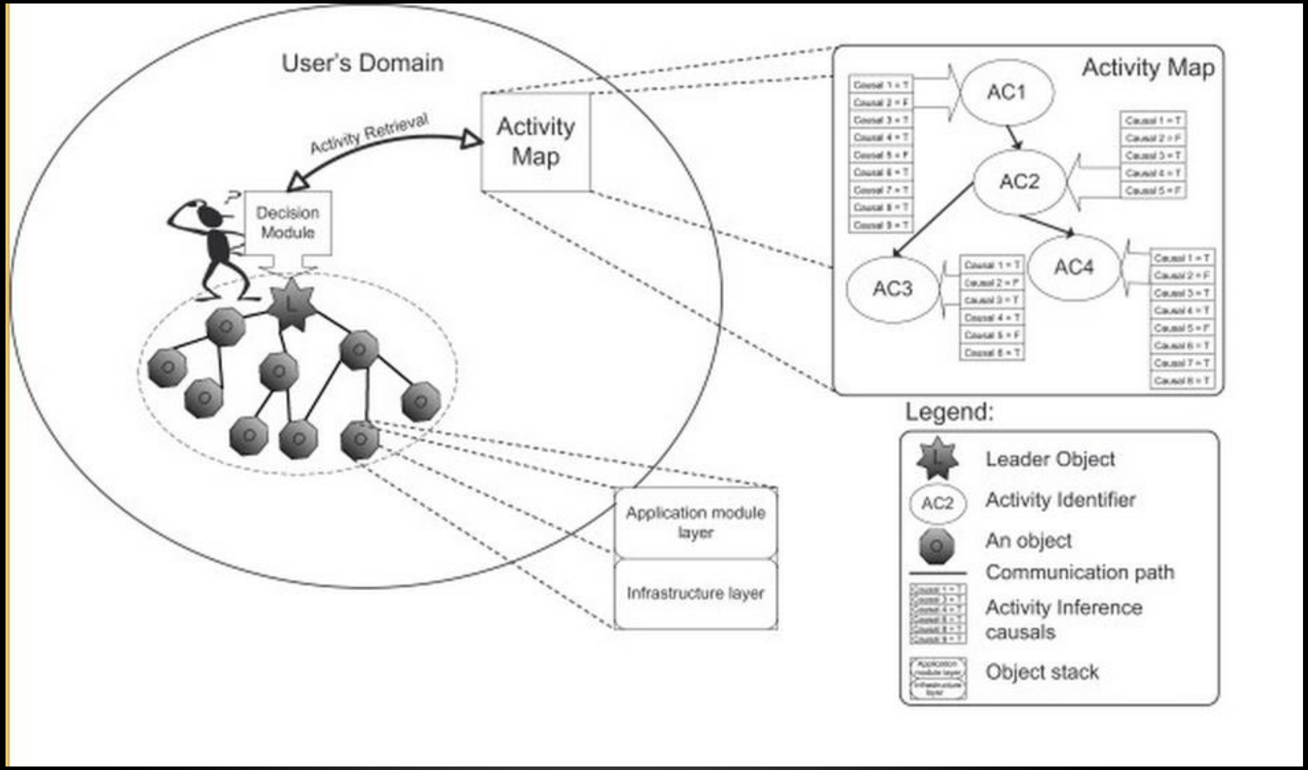
- Content-Based video search
- The video content is searched rather than metadata such as tag or keywords.
- It is difficult to manually annotate images with metadata in large databases and it may incorporate incorrect information.

MOTIVATION

- Xbox 360

MOTIVATION

- Health Care

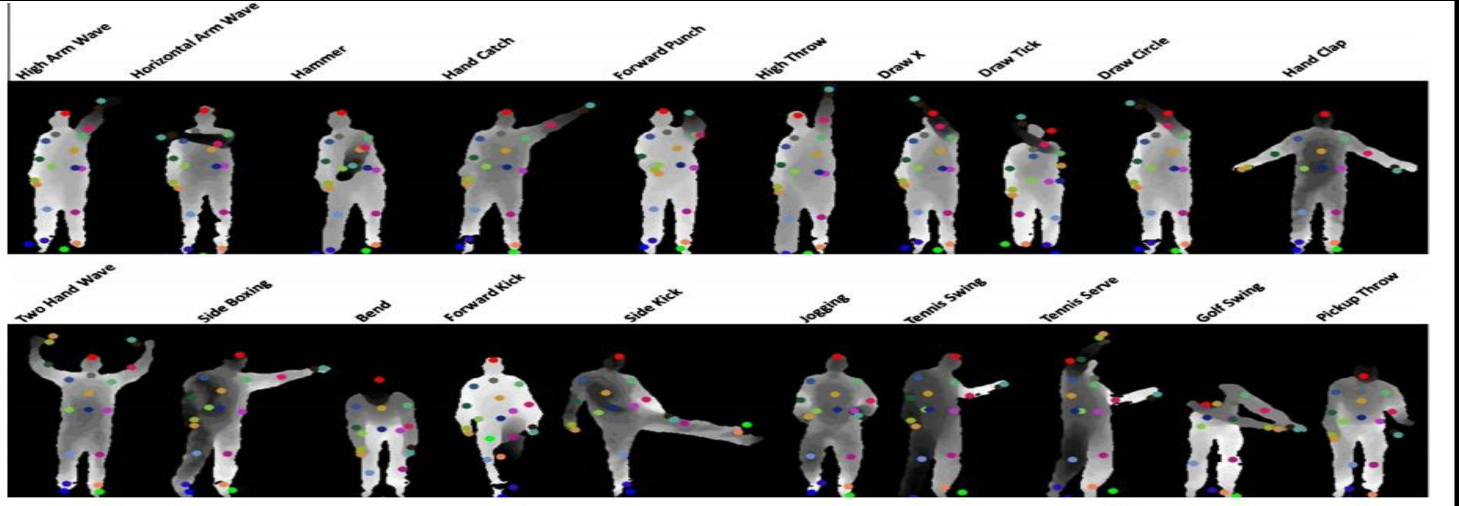


OVERVIEW

- Eigen-Joints Representation
- Naïve Bayes Nearest Neighbour Classification
- Informative Frame Selection

DATASET

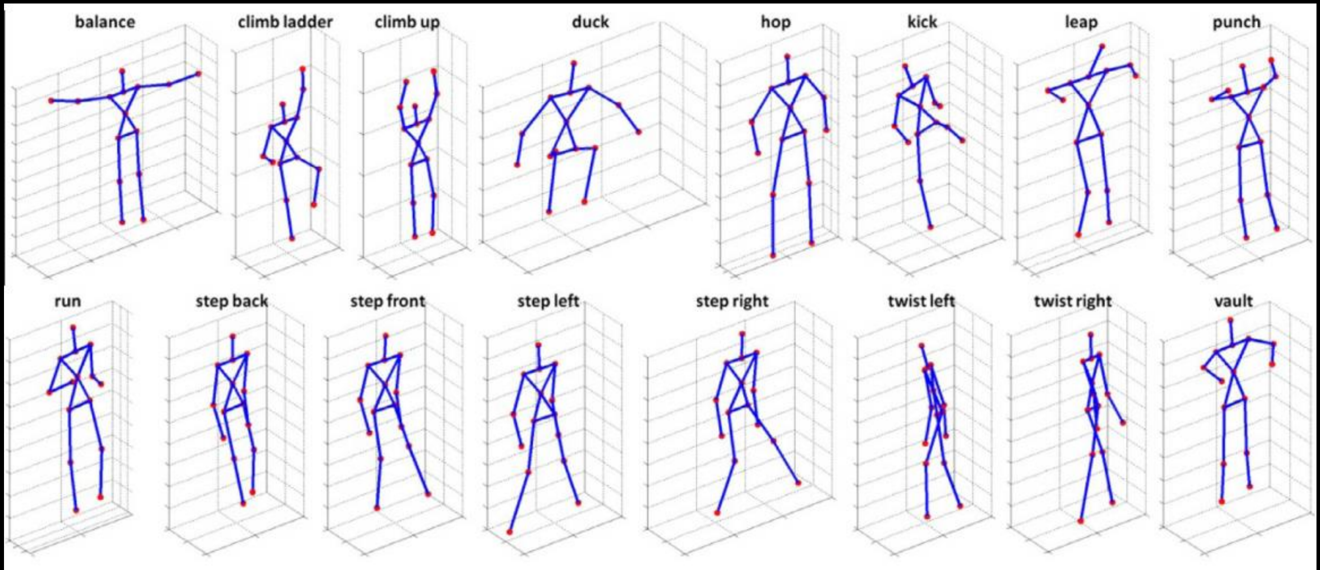
- MSR Action3D
  - 20 action types performed by 10 different subjects. Each subject performing an action 2 or 3 times.
  - Provides sequence of depth maps as well as skeleton joints.
  - Recorded with a depth sensor similar to the Kinect device..



Examples of depth maps and skeleton joints associated with each frame of twenty actions in the MSR Action3D dataset.

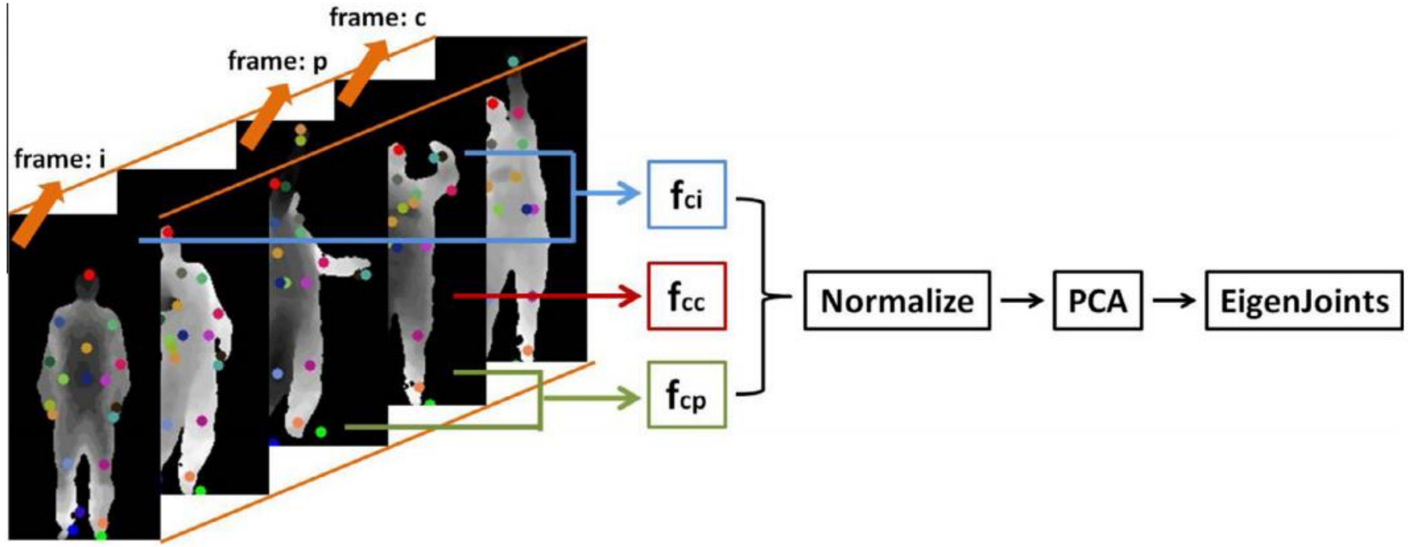
DATASET

- UCF Kinect
  - Each frame has 15 joints.
  - 16 actions performed by 16 different subjects
  - Depth maps are not provided



Sixteen actions and skeleton joints associated with each frame in the UCF Kinect dataset.

EIGEN-JOINTS REPRESENTATION



The framework of representing EigenJoints. In each frame, we compute three feature channels of f_{ci}, f_{cc}, and f_{cp} to capture the information of offset, posture, and motion. The normalization and PCA are then applied to obtain EigenJoints descriptor for each frame.

EIGEN-JOINTS REPRESENTATION

Static Posture Feature

$$f_{cc} = \{x_i - x_j | i, j = 1, 2, \ldots, N;\ i \neq j\}$$

Consecutive Motion Feature

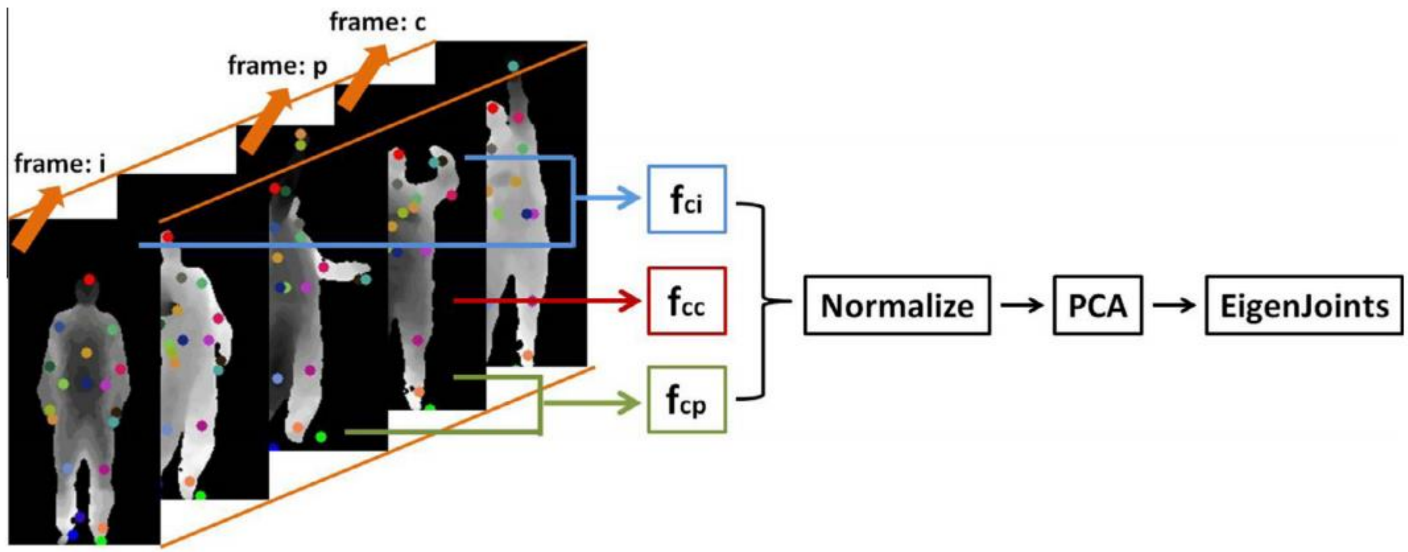
$$f_{cp} = \{x_i^c - x_j^p | x_i^c \in X_c; x_j^p \in X_p\}$$

Overall Dynamics Feature

$$f_{ci} = \{x_i^c - x_j^i | x_i^c \in X_c; x_j^i \in X_i\}$$

$$f_c = [f_{cc}, f_{cp}, f_{ci}]$$

EIGEN-JOINTS REPRESENTATION



The framework of representing EigenJoints. In each frame, we compute three feature channels of f_{ci}, f_{cc}, and f_{cp} to capture the information of offset, posture, and motion. The normalization and PCA are then applied to obtain EigenJoints descriptor for each frame.

NAÏVE BAYES NEAREST NEIGHBOUR(NBNN)

- Non parametric classifier for action classification
- No quantization of frame descriptors.
- Computation of Video-to-class distance, rather than conventional Video-to-Video distance.

$$C^* = \underset{c}{\operatorname{argmin}} \sum_{i=1}^{M} \|d_i - NN_c(d_i)\|^2$$

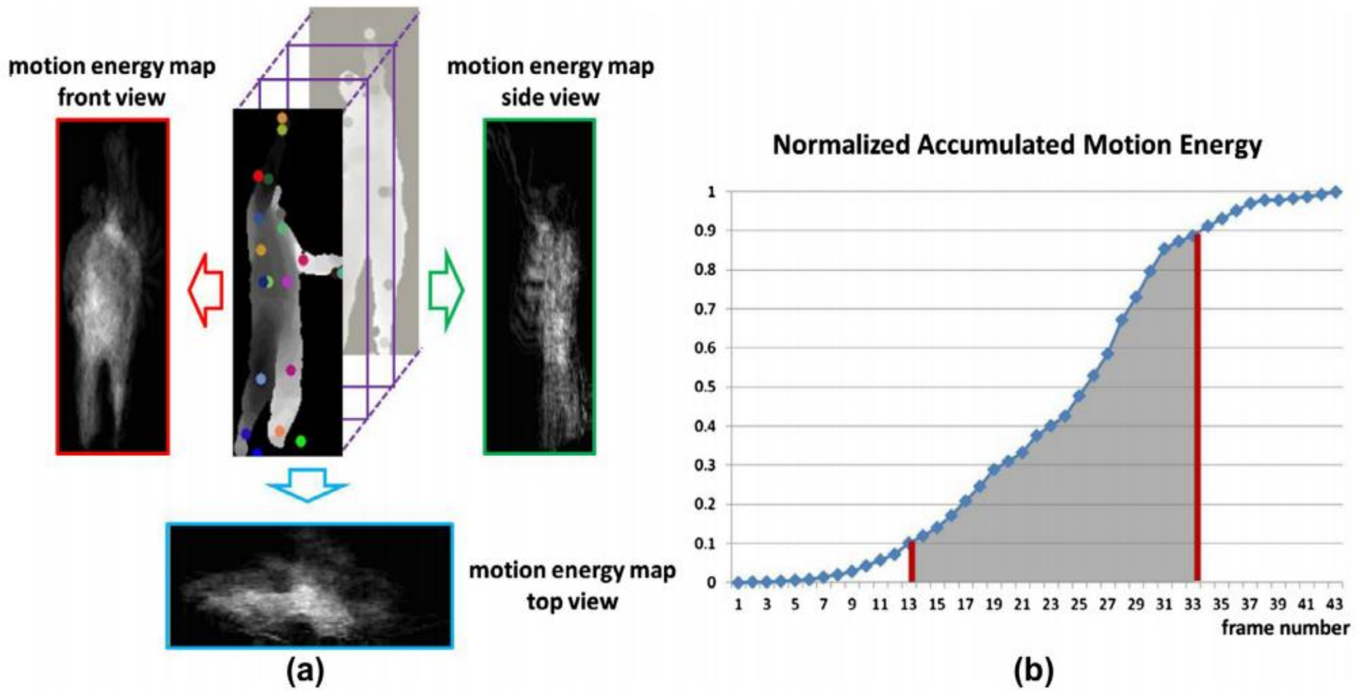
INFORMATIVE FRAME SELECTION

- All actions can be viewed as combination of four phases:-
  - Neutral
  - Onset
  - Apex
  - Offset
- Discriminative information between the frames is present mostly in the frames from onset and apex phases.
- So, extract frames from onset and apex phases and discard frames from neutral and offset phases.
- Reduces computational cost as the number of frames is reduced.

INFORMATIVE FRAME SELECTION

- 3D depth of each frame i is projected onto 3 orthogonal planes, which generate 3 projected frames $\boldsymbol{f_v}$, $v \in \{1,2,3\}$.

$$AME(i) = \sum_{v=1}^{3} \sum_{j=1}^{i} (|f_v^j - f_v^{j-1}| > \epsilon)$$



Computation of Accumulated Motion Energy (AME). (a) Motion energy maps associated with each projection view. (b) Normalized AME and selected informative frames.

REFERENCES

- X. Yang, Y. Tian, Effective 3D action recognition using EigenJoints, 2013.
- O. Boiman, E. Shechtman, M. Irani, In Defense of Nearest-Neighbor Based Image Classification, 2008.

THANK YOU!

ANY QUESTIONS?