Dance Movement Classification with Joint Detection

Stephanie Liu, Maia Ezratty – Advised by Andras Ferencz, Princeton University

MOTIVATION

- Human activity recognition (HAR): healthcare, security, video summary, highlight extraction
- Research has traditionally utilized a wide variety of feature representations
- Is activity recognition using detected joints effective?

GOAL AND APPROACH

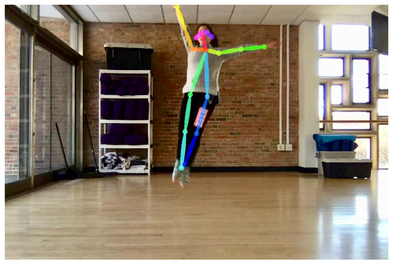
- Classify dance movement into one of the three categories performed—turn, jump, and fall
- Evaluate effectiveness of different feature representations **based on detected joints**
- Compare different KNN distance measures for each feature representation

RELATED WORK

- Joint detection
  - CMU OpenPose
  - Multi-person system to detect human body keypoints
- Feature representation
  - Actions as time-space models
  - Data from wearable accelerometers
- KNN distance measures and performance

DATA COLLECTION

- Self-collected data
- 60 training examples
  - 3 categories: turn, jump, and fall
  - 20 examples each — 10 per person
- 21 test cases: occlusion, orientation, variation
- 80 frames per video, passed through OpenPose



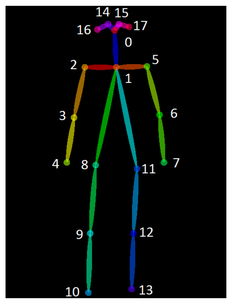
IMPLEMENTATION

Classification Algorithm

- K-Nearest Neighbors to classify the three movements: turn, jump, and fall
  - Simple and high performance
  - K = 1 to classify test cases
- Experimented with different distance measures
  - Manhattan (L-1)
  - Euclidean (L-2)
  - Chebyshev (L-infinity)

Feature Representation

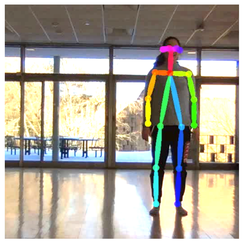
- Baseline (all features for all frames)



- Feature selection

Best features for L1	Best features for Euclidean	Best features for Chebyshev
right elbow y coord	right wrist y coord	right shoulder y coord
right wrist y coord	left elbow x coord	right wrist y coord
right hip y coord	right hip y coord	left elbow y coord
right ankle x coord		right ankle y coord
left hip y coord		left eye y coord

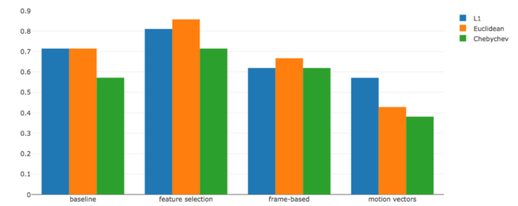
- Frame-based model
  - Each frame voted for the class that it was nearest to
  - Classification was based on number of votes



- Motion vector model
  - Calculated the change in distance between two frames for all joints
  - Normalized the distance by dividing by the distance between nose and neck

RESULTS

Distance Measure Performance



- Best: 85.7%
  - Feature selection representation with Euclidean distance measure
- Worst: 38.1%
  - Motion vector representation with Chebyshev distance measure

Test Case Performance

	Baseline (all features)	Feature Selection	Frame-based model	Motion vector model
Rise (unknown case)	x	x	x	x
Turn	✓	✓	✓	x
Turn	✓	✓	✓	x
Turn 45°	✓	✓	✓	✓
Turn 90°	✓	✓	✓	✓
Turn 180°	✓	✓	✓	✓
Turn close to camera	x	x	x	✓
Double turn	x	x	x	✓
Turn variation	x	✓	✓	x
Fall	✓	✓	✓	x
Fall 90°	✓	✓	✓	x
Fall 180°	✓	✓	✓	x
Fall close to camera	✓	✓	✓	x
Fall variation	✓	✓	✓	x
Jump	✓	✓	✓	✓
Jump 90°	x	✓	x	✓
Jump 180°	✓	✓	x	✓
Jump occluded	✓	✓	✓	✓
Jump close to camera	x	✓	x	✓
Jump variation	✓	✓	✓	✓
Jump variation	✓	✓	x	✓
ACCURACY	0.714	0.857	0.666	0.571

- Test cases covered unseen movement, different orientations, occlusion/missing joints, and variations of each type of movement
- Strengths
  - Orientation
  - Variation that did not resemble other classes
  - Did not correctly classify an unknown movement
- Weaknesses
  - Occlusion
  - Multi-person classification
  - Dependent on accurate joint detection

CHALLENGES

- The software used for joint detection, OpenPose, did not have feature tracking
  - As a result, multi-person movement classification was inaccurate
  - We eliminated these test cases from our accuracy calculations
- Joint detection sometimes fails due to occlusion or other confusion in a video
  - We replaced these missing joints with their position from the previous frame
  - This improved accuracy results

CONCLUSIONS

- Dance movement classification performs well with joint-based feature representations
- Given our training data, we found that using feature selection and the Euclidean distance gave the best classification accuracy
  - Potential reasons:
    - Dance movement is effectively represented by a few key joints
    - Euclidean distance is logical for body movement
- The motion vector model performed the worst of all the feature representations
  - Potential reasons:
    - Poor distance normalization
    - Does not utilize initial positions

FUTURE WORK

- Extension 1: Reject unseen actions
- Extension 2: Extend dance movement classification to dance recognition

ACKNOWLEDGEMENTS

Thank you to Andras Ferencz for your helpful advice, and Professor Russakovsky, Riley, and Kyle for your help this semester.