Action and Activity Recognition

Exploring Motion History Images for Human Action Classification

By: Khalid Mahmoud, Ethan Grosvenor, Seth Ojo, Devstutya Pandey, Mack Bourne

What is Action and Activity Recognition?

- The main idea of action recognition is to be able to identify human movements from video data.

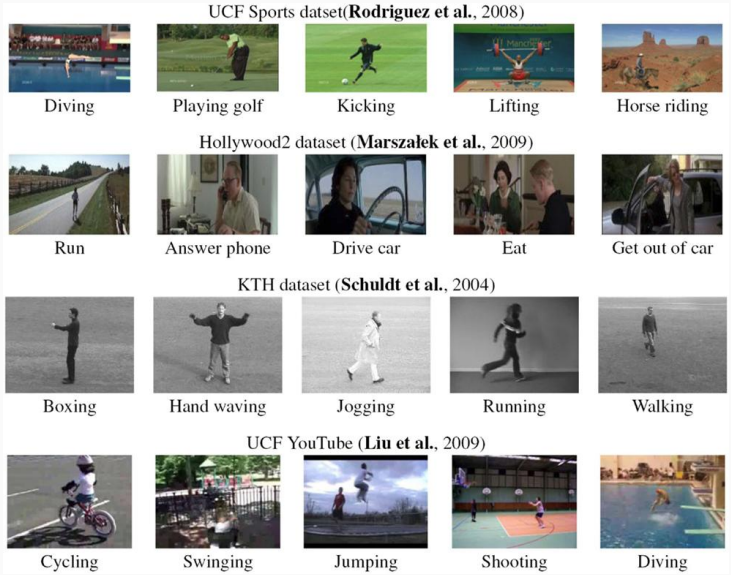
- Has a wide range of real-world applications:

- Sports analytics for tracking player performance or techniques used.

- Surveillance and security to improve threat detection

- Healthcare & Rehabilitation – Monitoring patients and assisting therapy.

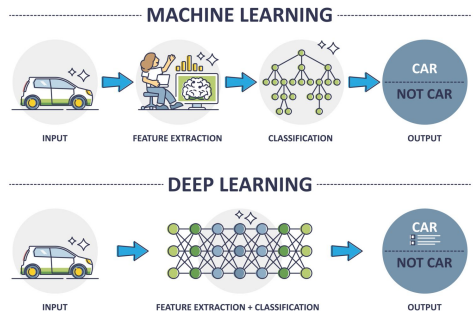
- Human-Computer Interaction – Enabling gesture control in gaming and VR.



Background

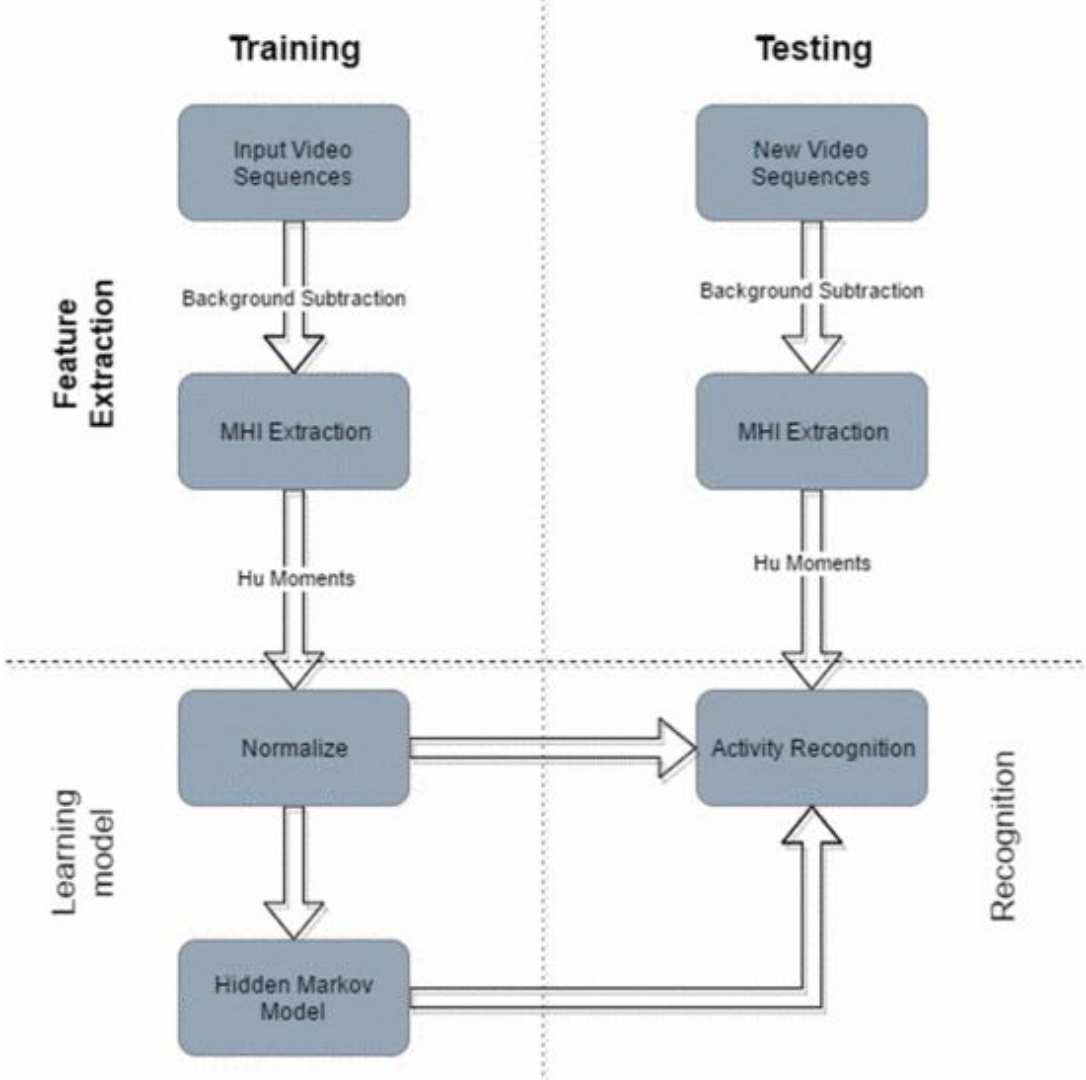
Key Studies on Action Recognition Methods

Reviewed Papers:

	E. C. Alp and H. Y. Keles, "Action recognition using MHI based Hu moments with HMMs," IEEE EUROCON 2017 -17th International Conference on Smart Technologies, Ohrid, Macedonia, 2017
	CNNs, RNNs and Transformers in Human Action Recognition: A Survey and a Hybrid Model" by Khaled Alomar, Halil Ibrahim Aysel, and Xiaohao Cai from the University of Southampton, UK.

Baseline Technique:

Action Recognition Using MHI Based Hu Moments With HMMs


Training
Testing
Input Video Sequences
New Video Sequences
Feature Extraction
Background Subtraction
Background Subtraction
MHI Extraction
MHI Extraction
Hu Moments
Hu Moments
Normalize
Activity Recognition
Learning model
Recognition
Hidden Markov Model

Preprocessing - Silhouette Masks (Background Subtraction)

Silhouette Masks:

- Binary images that isolate the subject from the background.

How They Help:

- Focus on **human motion**, removing background noise.
  Improve **motion analysis** for better action recognition.

Advantages:

- **Noise Reduction** – Eliminates unnecessary background details.
  Consistency – Maintains a uniform shape representation.

Impact on Performance:

- Clean, pre-processed data improves **action classification** accuracy.

Feature Extraction: Motion History Image (MHI):

What is MHI?

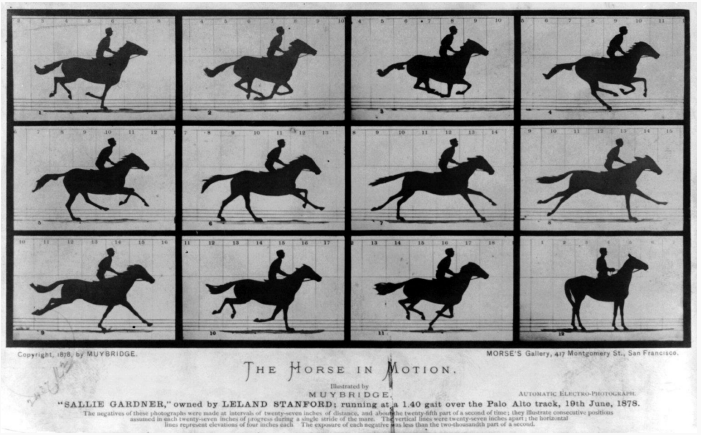
- A grayscale image capturing motion history.
  Brighter pixels = recent movement; darker pixels fade over time.

How It Works

- Updates pixel values based on motion in consecutive frames.
  Fades older motion while emphasizing recent activity.

Why Use MHI?

- **Encodes motion timing** (e.g., walking vs. running).
- **Captures motion direction** (tracks movement flow).
- **Compact representation** (summarizes motion in one image).



Feature Extraction - Hu Moments

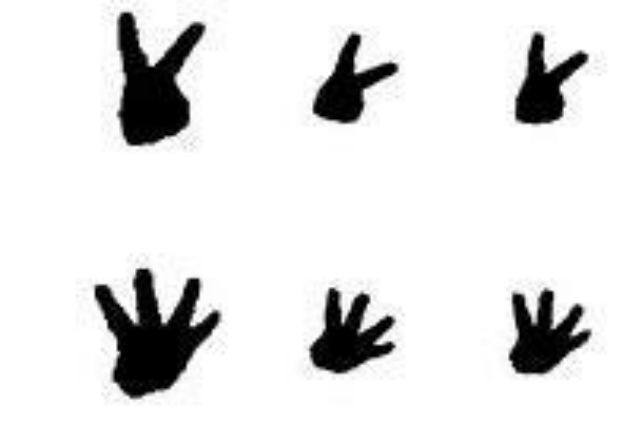
Hu Moments

- Mathematical descriptors capturing overall shape and structure of motion.

- Summarize human movement patterns (e.g., walking, running).

Benefits:

- Robust for summarizing global motion features.
- Invariant to scale, rotation, and translation.
- Ideal for motion recognition regardless of variations in viewing angles or sizes.



Hidden Markov Models (HMMs) & Validation

HMMs for Temporal Modeling

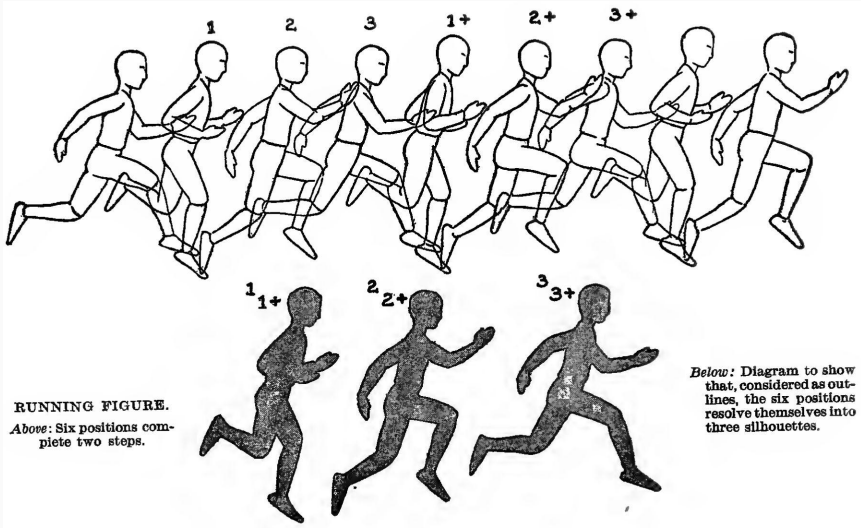
- Model how actions unfold over time.
- Capture sequential patterns in motion data.
- Ideal for recognizing temporal patterns.

Cross-Validation Approach

- Leave-One-Out Cross-Validation.
- One sequence held out per action for training/testing.

Generalization

- Ensures the system generalizes well to new individuals.
- Improves robustness and performance on unseen data.



Challenges in Baseline Technique

Why Classic MHI Falls Short

- No Motion Intensity: Treats all movements equally (improved with optical flow).

- **Loses Temporal Order**: Captures only final shape, not sequence.

- **No Directional Awareness**: Can't distinguish movement direction (partially improved with optical flow).

- **Limited to Static Backgrounds**: Noise from moving backgrounds requires silhouette masks.

Our Approach

Basis for Our Project

- Elit Cenk Alp and Hacer Yalim Keles (IEEE) presents a method for action recognition using MHI and Hu moments, combined with Hidden Markov Models (HMMs).
    - Achieves 96% classification accuracy on the Weizmann dataset

Areas for Improvement

- **Complex Backgrounds**: Better handle complex backgrounds with heavy motion.
- **Performance in Outdoor Settings**: It achieves its lowest accuracy in outdoor activities (walking, running and skipping).

How We Build On It

1. Improved Feature Extraction
2. Dynamic Action Representation
3. Enhanced Temporal Understanding

Dataset

1. Walking
2. Running
3. Jumping jacks
4. Galloping sideways
5. Bending over
6. One hand wave
7. Two hand wave
8. Jumping in place
9. Jumping
10. Skipping

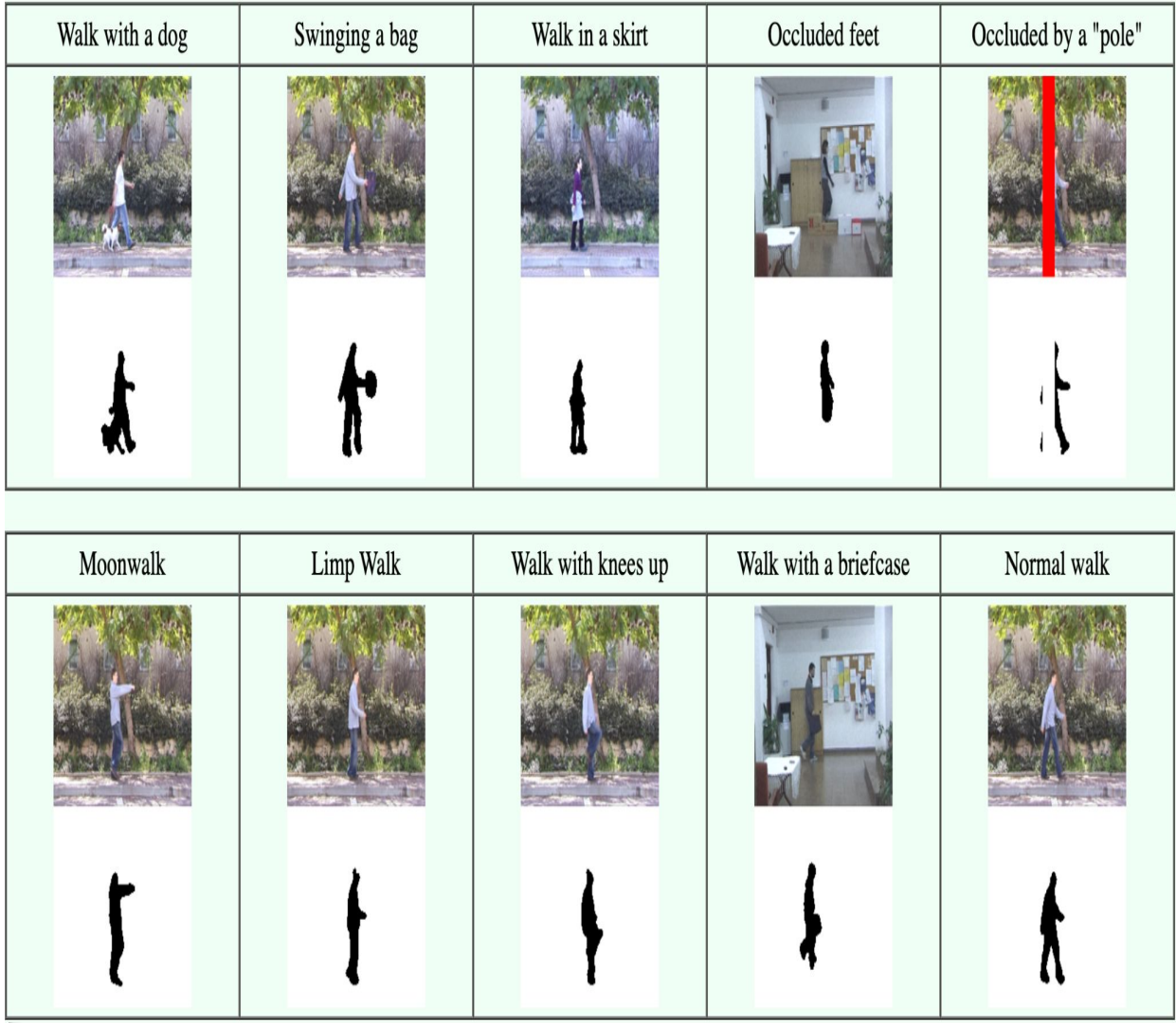
- **Videos of 9 subjects**, **10 different actions**
- **= 90 sequences** total

The dataset includes:

- RGB video clips
- **Binary silhouette masks** extracted (`original_masks.mat`)

Frame size: 144 × 180 pixels

Frame count per video: Varies (~30–90 frames)
Format: `.avi` for videos, `.mat` for extracted masks



Making MHI Better: Our Innovations

1. Optical Flow-Weighted MHI

- Uses Optical Flow (Farneback)
- Improves Motion Intensity Representation
- Enhances Temporal Dynamics

2. Temporal Pyramid MHI Representation

- Divides Video into Time Segments
- Computes MHI for Each Segment.
- Captures Motion Evolution Over Time

3. Directional Motion Energy Images (DMEI)

- Separate Motion Masks for Each Direction.
- Extracts HOG & LBP Features
- Improves Classification Accuracy

How Our System Works

Pipeline: Unchanged Components

Input Format

- **Data Type:** Pre-extracted binary **silhouette masks** from original_masks.mat
- **Shape:** Each sample is a video volume (3D array: height × width × frames

Preprocessing & Cropping

- Every frame is **cropped** to remove padding
- **Thresholded** to binary (foreground = 255, background = 0)

Motion Encoding (MHI Base)

- **Motion History Image (MHI):** Encodes temporal motion by accumulating silhouette differences with a decay factor

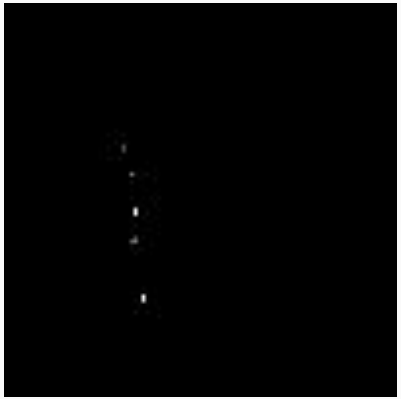
Feature Extraction

- Extract **Hu Moments** from final MHI

Temporal Modeling

- Hidden Markov Model (HMM)
- Leave-One-Out Cross Validation (LOOCV)
- Tested using a **sliding window** strategy for robustness

Motion Visualization from Dataset



Optical Flow-Weighted MHI

What Changed:

- Added **optical flow magnitude weighting** into MHI construction
- Optical flow computed between grayscale frames using `cv2.calcOpticalFlowFarneback`
- Silhouette is weighted by `(1 + normalized_magnitude)` before being added to the MHI
- Improves sensitivity to motion **intensity** and speed (e.g., distinguishing walk vs run)

Why It Matters:

- Helps highlight **subtle or fast movements** more clearly
- Reduces confusion between visually similar actions that differ temporally

Temporal Pyramid MHI

What Changed:

- Introduced **Temporal Pyramid Segmentation**
- Each video is split into n equal **temporal segments** (n = 3)
- Separate MHI is constructed per segment
- Features from each segment are extracted and **concatenated**

Why It Matters:

- Captures **phase-specific motion** (e.g., jump's buildup vs landing)
- Aims to enhance temporal discriminability of actions

Note: MHI construction and feature extraction per segment remain the same.

DMHI (Directional MHI)

What Changed:

- Attempted to implement **Directional Motion History Image** by computing directional gradients
- Intended to capture **motion direction** in addition to intensity

Note: May require cleaner directional encoding (e.g., histogram of flow angles)

Challenges:

- Implementation instability — resulted in significant **misclassification**
- Example: all `wave2` actions predicted as `bend`, suggesting corrupted or noisy features

Combined Optical Flow-Weighted + DMHI

What Changed:

- Combines **directional motion encoding** with **optical flow magnitude weighting**
- Silhouettes weighted by optical flow magnitude
- Directional variation captured

Observations:

- Conceptually strong (combining direction + speed)
- Performance may depend heavily on **angle binning** and gradient representation quality

Combined Optical Flow-Weighted + Temporal Pyramid MHI

What Changed:

- Combined enhancements
- Each temporal segment's MHI is **optical flow weighted**
- Features from each weighted segment are concatenated for final descriptor

Why It Matters:

- Combines **motion magnitude sensitivity** with **temporal phase-awareness**
- Best of both worlds: recognizes subtle variations **and** phase evolution

Results

Baseline

Classification report:

	precision	recall	f1-score	support
bend	0.96	1.00	0.98	360
jack	1.00	1.00	1.00	450
jump	1.00	0.99	1.00	179
pjump	1.00	0.95	0.97	259
run	0.92	0.96	0.94	98
side	0.83	0.98	0.90	165
skip	0.97	0.98	0.97	176
walk	0.98	1.00	0.99	401
wave1	1.00	0.98	0.99	374
wave2	1.00	0.90	0.94	345
accuracy			0.97	2807
macro avg	0.97	0.97	0.97	2807
weighted avg	0.98	0.97	0.97	2807

1st Method:
Temporal Pyramid Motion Histogram Imaging

Classification report:

	precision	recall	f1-score	support
bend	0.96	1.00	0.98	360
jack	1.00	1.00	1.00	450
jump	1.00	0.99	1.00	179
pjump	1.00	0.95	0.97	259
run	0.92	0.96	0.94	98
side	0.83	0.98	0.90	165
skip	0.97	0.98	0.97	176
walk	0.98	1.00	0.99	401
wave1	1.00	0.98	0.99	374
wave2	1.00	0.90	0.94	345
accuracy			0.97	2807
macro avg	0.97	0.97	0.97	2807
weighted avg	0.98	0.97	0.97	2807

Classification report:

	precision	recall	f1-score	support
bend	0.96	1.00	0.98	360
jack	1.00	1.00	1.00	450
jump	1.00	0.99	1.00	179
pjump	1.00	0.95	0.97	259
run	0.92	0.96	0.94	98
side	0.83	0.98	0.90	165
skip	0.97	0.98	0.97	176
walk	0.98	1.00	0.99	401
wave1	1.00	0.98	0.99	374
wave2	1.00	0.90	0.94	345
accuracy			0.97	2807
macro avg	0.97	0.97	0.97	2807
weighted avg	0.98	0.97	0.97	2807

Baseline vs. Temporal Pyramid

Baseline (Original MHI)
Accuracy: 97%
Key Observations:

- Strong performance across most action classes
- Some misclassification of similar actions (side gallop mistaken as jumping jacks)
- Struggles with dynamic actions, particularly for subtle differences.

T**emporal Pyramid MHI**
Accuracy: 97%
Key Observations:

- No improvement from baseline.
- Model insensitivity

2nd Method:
Directional Motion History Image (DMHI)

Classification report:

	precision	recall	f1-score	support
bend	0.96	1.00	0.98	360
jack	1.00	1.00	1.00	450
jump	1.00	0.99	1.00	179
pjump	1.00	0.95	0.97	259
run	0.92	0.96	0.94	98
side	0.83	0.98	0.90	165
skip	0.97	0.98	0.97	176
walk	0.98	1.00	0.99	401
wave1	1.00	0.98	0.99	374
wave2	1.00	0.90	0.94	345
accuracy			0.97	2807
macro avg	0.97	0.97	0.97	2807
weighted avg	0.98	0.97	0.97	2807

Classification report:

	precision	recall	f1-score	support
bend	0.66	0.80	0.72	369
jack	0.84	1.00	0.91	459
jump	0.95	0.74	0.83	188
pjump	0.85	0.81	0.83	268
run	1.00	0.96	0.98	107
side	1.00	0.87	0.93	174
skip	0.59	1.00	0.74	186
walk	0.99	1.00	1.00	411
wave1	0.74	0.53	0.62	383
wave2	0.73	0.51	0.60	354
accuracy			0.81	2899
macro avg	0.84	0.82	0.82	2899
weighted avg	0.82	0.81	0.80	2899

Baseline vs DMHI

Baseline (Original MHI)
Accuracy: 97%
Key Observations:

- Strong performance across most action classes
- Some misclassification of similar actions (side gallop mistaken as jumping jacks)
- Struggles with dynamic actions, particularly for subtle differences.

DMHI
Accuracy: 77%

Key Observations:
- Significant decrease in classification, particularly with skip and bend
- Better for actions with directional movement
- Possibly increase classifications using optical flow

3rd Method:
Optical Flow + DMHI

Classification report:

	precision	recall	f1-score	support
bend	0.96	1.00	0.98	360
jack	1.00	1.00	1.00	450
jump	1.00	0.99	1.00	179
pjump	1.00	0.95	0.97	259
run	0.92	0.96	0.94	98
side	0.83	0.98	0.90	165
skip	0.97	0.98	0.97	176
walk	0.98	1.00	0.99	401
wave1	1.00	0.98	0.99	374
wave2	1.00	0.90	0.94	345
accuracy			0.97	2807
macro avg	0.97	0.97	0.97	2807
weighted avg	0.98	0.97	0.97	2807

Classification report:

	precision	recall	f1-score	support
bend	0.66	0.80	0.72	369
jack	0.84	1.00	0.91	459
jump	0.95	0.74	0.83	188
pjump	0.85	0.81	0.83	268
run	1.00	0.96	0.98	107
side	1.00	0.87	0.93	174
skip	0.59	1.00	0.74	186
walk	0.99	1.00	1.00	411
wave1	0.74	0.53	0.62	383
wave2	0.73	0.51	0.60	354
accuracy			0.81	2899
macro avg	0.84	0.82	0.82	2899
weighted avg	0.82	0.81	0.80	2899

Baseline vs DMHI + Optical Flow

Baseline:
Accuracy: 97%
Key Observations:

- Strong performance across most action classes
- Some misclassification of similar actions (side gallop mistaken as jumping jacks)
- Struggles with dynamic actions, particularly for subtle differences.

DMHI + Optical Flow:
Accuracy: 77%
Key Observations:

- No improvements from base DMHI implementation
- Optical flow works best in detecting excess information
- Motion template was most likely being overwritten by poor encoding

Final Improvement:
Optical Flow + Temporal Pyramid

Classification report:

	precision	recall	f1-score	support
bend	0.96	1.00	0.98	360
jack	1.00	1.00	1.00	450
jump	1.00	0.99	1.00	179
pjump	1.00	0.95	0.97	259
run	0.92	0.96	0.94	98
side	0.83	0.98	0.90	165
skip	0.97	0.98	0.97	176
walk	0.98	1.00	0.99	401
wave1	1.00	0.98	0.99	374
wave2	1.00	0.90	0.94	345
accuracy			0.97	2807
macro avg	0.97	0.97	0.97	2807
weighted avg	0.98	0.97	0.97	2807

Classification report:

	precision	recall	f1-score	support
bend	0.99	0.99	0.99	360
jack	1.00	0.99	1.00	450
jump	0.92	0.98	0.95	179
pjump	1.00	0.96	0.98	259
run	0.99	0.97	0.98	98
side	0.96	0.98	0.97	165
skip	0.98	0.98	0.98	176
walk	0.96	1.00	0.98	401
wave1	1.00	0.98	0.99	374
wave2	1.00	0.98	0.99	345
accuracy			0.98	2807
macro avg	0.98	0.98	0.98	2807
weighted avg	0.98	0.98	0.98	2807

Baseline vs Optical Flow + Temporal Pyramid

Baseline (Original MHI):
Accuracy: 97%
Key Observations:

- Strong performance across most action classes
- Some misclassification of similar actions (side gallop mistaken as jumping jacks)
- Struggles with dynamic actions, particularly for subtle differences.

Optical Flow + Temporal Pyramid:

Accuracy: 98%

Key Observations:

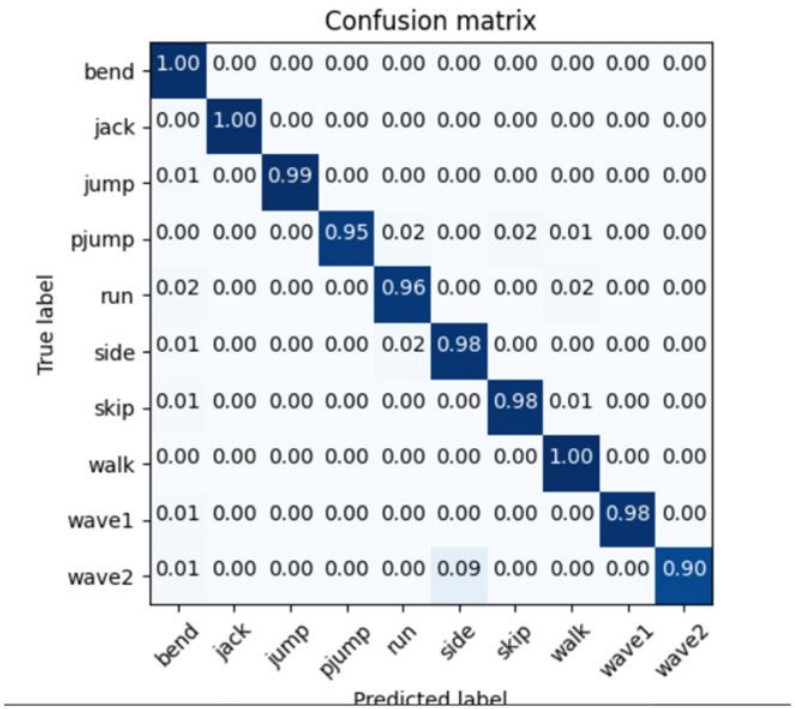
- Improvements to dynamic movements
- Significant decrease in jumping accuracy
- Noisy calculations

Baseline Confusion Matrix

Key Characteristics: Uses the traditional Motion History Image (MHI) or simple feature extraction + HMM training

Performance:

Overall accuracy remains at a basic level: some classes show relatively high accuracy, while others exhibit notable confusion.

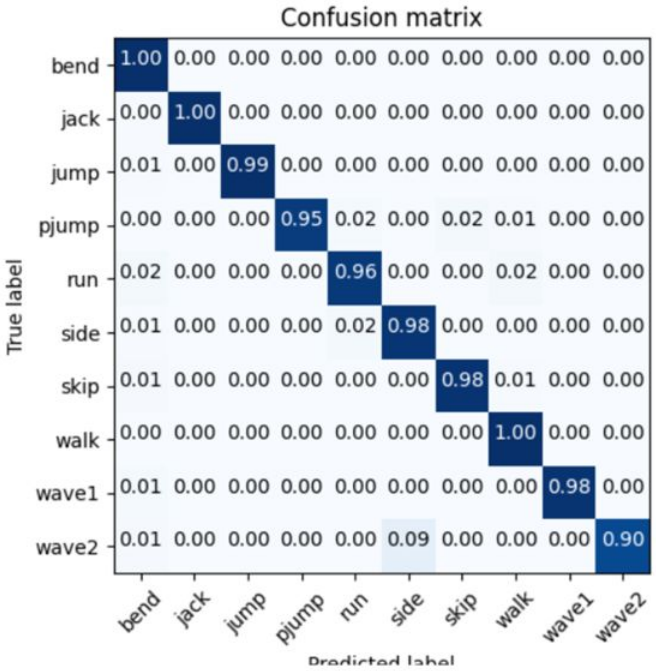


Temporal Pyramid Confusion Matrix

Key Idea: Divides video into segments, computes MHI per segment, and concatenates to capture action stages.

Benefits:

Accuracy is similar to baseline, with no major improvement.

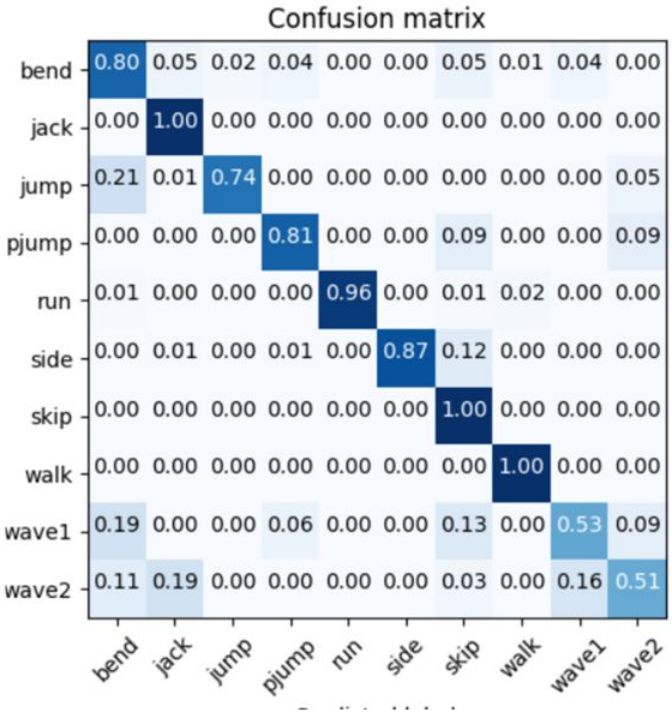


DMHI Confusion Matrix

Key Idea: Captures motion direction (up/down/left/right) to emphasize directional differences in actions.

Drawbacks:

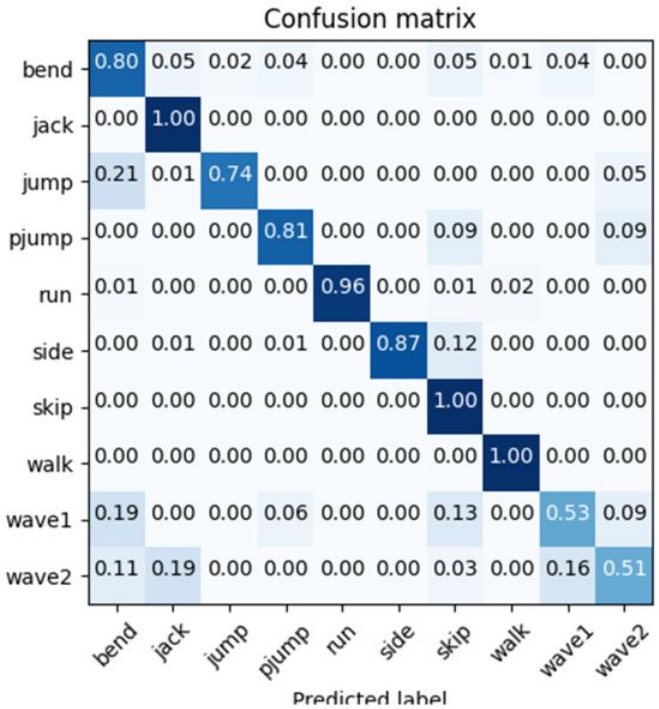
1. Improves accuracy for multi-directional actions (e.g., skip).
2. Struggles with actions differing by speed or subtle posture—misses amplitude/shape cues.
3. Overall accuracy drops, not ideal.



Combined Optical flow and DMHI

Key Idea:
Captures motion direction (up/down/left/right) and flow-weighted MHIs to better distinguish multi-directional actions.

Drawbacks:
1. Overemphasizes direction, missing posture and motion amplitude—hurts accuracy for actions like *bend*, *pjump*.
2. Higher feature dimensionality leads to overfitting and weaker generalization under LOOCV.
3. Overall Accuracy decreased

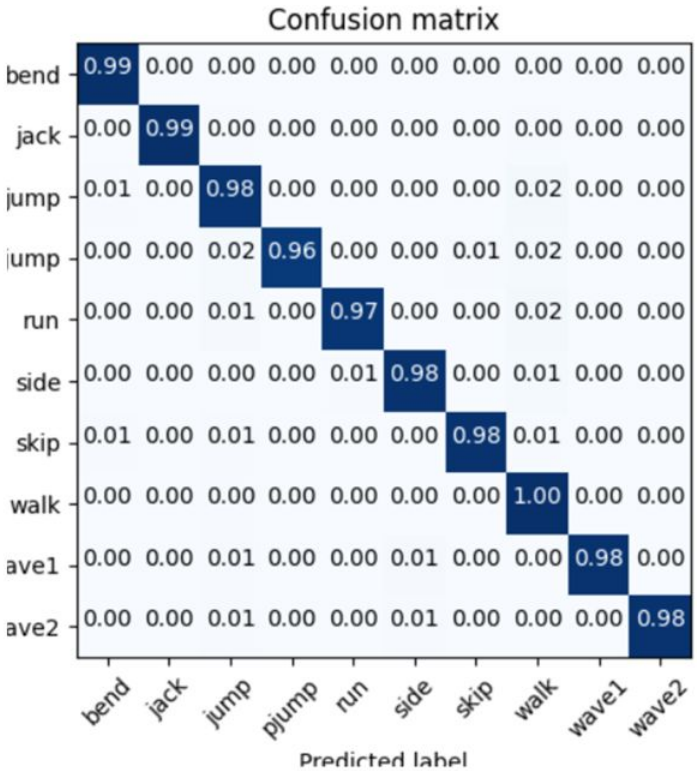


Combined Optical flow and Temporal

Key Idea: Combines optical flow-weighted MHI with temporal pyramid to emphasize strong motions and action phases.

Benefits:

1. Better detects large, fast actions (e.g., run, pjump).
2. Clarifies phase-based differences (e.g., wave1 vs. wave2).
3. Lowest confusion and highest overall accuracy—best-performing method.



Live Demo!

Limitations

- Performance variations across different Actions
- Sensitivity to Background Complexity
- **Expanding Datasets**: Testing on more extensive and diverse datasets for better generalization and real-world performance.
- DMHI require cleaner directional encoding

Conclusion

- Enhanced Action Recognition System:
  - MHI, silhouette masks & motion features.
- Key Findings:
  - **Optical Flow + Temporal Pyramid** yielded the highest accuracy.
  - Integration of motion across time scales enhances precision.
- Real World Impact:
  - Applications in **sports analytics, surveillance, and human-computer interaction**.
  - Provides a **lightweight, real-time alternative** to deep learning models.
- Future Work:
  - **Improving background robustness** for complex, dynamic environments.
  - **Expanding datasets** for better generalization.

Thank You!