

MDS-Based Multi-axial Dimensionality Reduction Model For Human Action Recognition



Introduction

- We want to develop a new unsupervised human recognition system which is able to recognize more than two classes with a low complexity.
- Recognize automatically low-level actions such as running, walking, hand clapping, etc. from an input video sequence.
- Based on the generation and fusion of a set of prototypes generated from different view-points of the data cube of the video sequence

Proposed Solution

- Background subtraction
- Non linear dimensionality reduction technique
- Classification and fusion

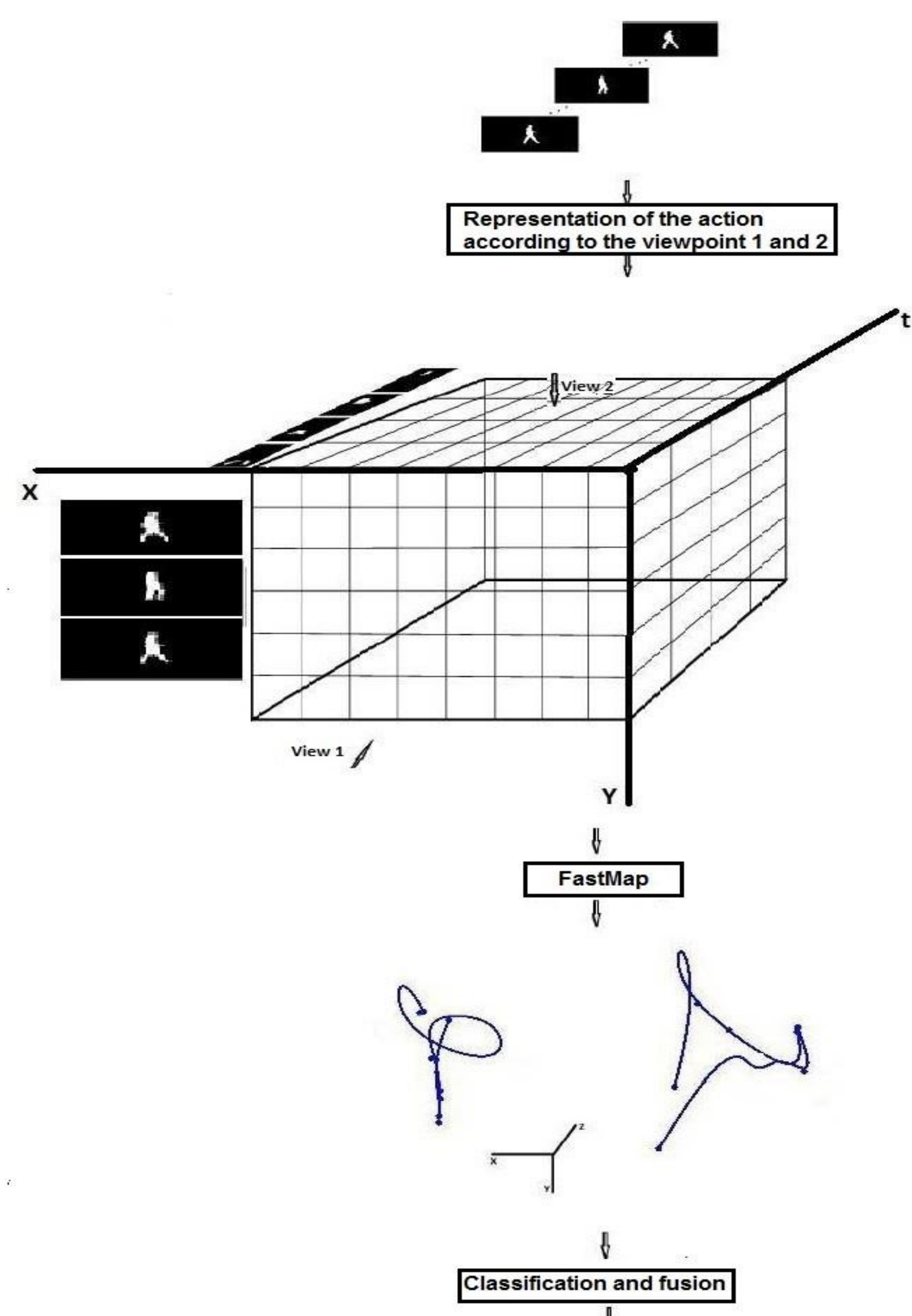


Fig. 1. Human action recognition system based on set of prototypes or 3D curves generated by a MDS-based dimensionality reduction according two viewpoints on sequence of binarized silhouettes

Preprocessing

- Consists in obtaining the binary video sequence of 2D silhouettes for each human action.
- Subtraction the median background from each image of the sequence and have then used a simple thresholding technique with a classical median filter
- Centering the gravity center of each binarized body silhouettes inside a rectangular fixed size bounding box

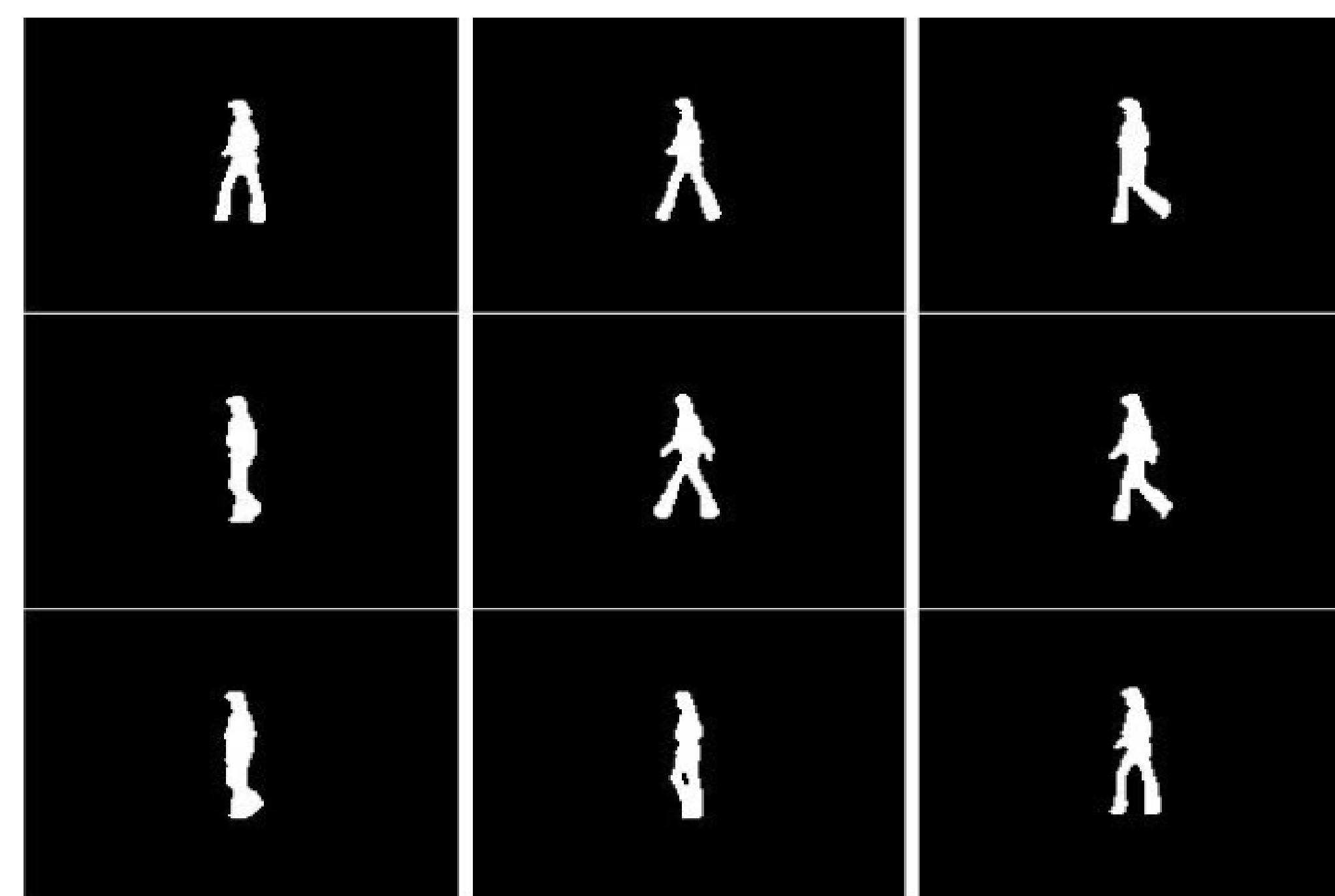


Fig. 2. Example of video sequence from the Weizmann data-set after the preprocessing step to extract and center the body silhouettes

Prototype Extraction

- Consists in building a set of prototypes or, more precisely, a set of reliable, compact and discriminative representations of the data cube containing the sequence of binarized silhouettes.
- Generated by an MDS based non linear dimensionality reduction technique from different view-points of the video cube, more precisely view-points 1 and 2.
- The first viewpoint aims at reducing the dimensionality of the image cube along the temporal axis of the video sequence by modeling every silhouette image frames in a low 3D dimensional space.
- The second viewpoint aims at reducing the dimensionality of the image cube through the spatial (line or column) axis of the set of 2D binarized silhouettes.

Classification and fusion

- For each view-point, classification prototype results by using the K -NN classifier generate a score vector both indicating the class and the sum of the number of nearest points of the prototype with the most nearest points
- Procedure fusion consists in adding the two score vectors generated by the first and second viewpoint and to classify a test prototype by the majority class

System validation

- Validated by using the *leave-one-out* procedure and the Weizmann human action data-set

Experimental Results

	walk	run	skip	jack	jump	pjump	side	wave1	wave2	bend
walk	0.89	0.11								
run	0.11	0.67	0.22							
skip	0.11	0.22	0.45	0.11	0.11					
jack				0.67						0.33
jumup					0.78			0.11	0.11	
pjump					0.11	0.78		0.11		
side							0.89	0.11		
wave1								0.89	0.11	
wave2								0.33	0.67	
bend									0.11	0.89

Confusion matrix associated to viewpoint 1

	walk	run	skip	jack	jump	pjump	side	wave1	wave2	bend
walk	0.78	0.11								
run		0.89	0.11							
skip		0.22	0.67							
jack		0.11		0.45	0.11			0.11	0.22	
jump				0.33	0.34	0.22	0.11			
pjump				0.11	0.11	0.78				
side		0.11					0.89			
wave1								0.66	0.22	
wave2								0.11	0.78	
bend									0.11	0.89

Confusion matrix associated to viewpoint 2

	walk	run	skip	jack	jump	pjump	side	wave1	wave2	bend
walk	1.0									
run		1.0								
skip			0.11	0.89						
jack				0.78						0.22
jump					0.78			0.11	0.11	
pjump				0.11		0.89				
side							1.0			
wave1								1.0		
wave2								0.11	0.89	
bend										1.0

Confusion matrix associated to the fusion of the two viewpoints

	View point 1	View point 2	Fusion
Number of K-NN	K=1	K=1	K=1
Recognition rate	75.8	71.3	92.3

Recognition rate of each view point and the fusion

Conclusion

- A simple human action recognition system based on two complementary prototypes.
- The sensitivity of the recognition system to each individual viewpoint is 75.8 and 71.3
- The efficiency of our multi-viewpoint based fusion is 92.3
- Give a better representation of the action in low dimension
- Does not exploit a spatio-temporal characteristics of action

References

- S. Ali and M. Shah, Human action recognition in videos using kinematic features and multiple instance learning, IEEE Trans. Pattern Anal. Mach. Intell., 32(2):288-303, 2010
- M. Blank and L. Gorelick and E. Shechtman and M. Irani and R. Basri, Actions as space-time shapes, Proceedings of the Tenth IEEE International Conference on Computer Vision, ICCV'05, pp. 1395--1402, 2005