Víctor Ponce-López[1,2,3], Hugo Jair Escalante[4], Sergio Escalera[2,3], Xavier Baró[1,2,3]

[1]IN3, Open University of Catalonia, Rambla Poblenou 156, 08018 Barcelona
[2]Dept. MAiA, University of Barcelona, Gran Via de les Corts Catalanes 585, 08007 Barcelona
[3]Computer Vision Center, Building O, Campus UAB, 08193 Bellaterra, Barcelona
[4]Computational Sciences Dept. INAOE, Luis Enrique Erro 1, 72840 Tonantzintla, Puebla, Mexico
vponcel@uoc.edu, hugo.jair@gmail.com, sergio@maia.ub.es, xbaro@uoc.edu

Gesture and Action Recognition by Evolved Dynamic Subgestures

ABSTRACT

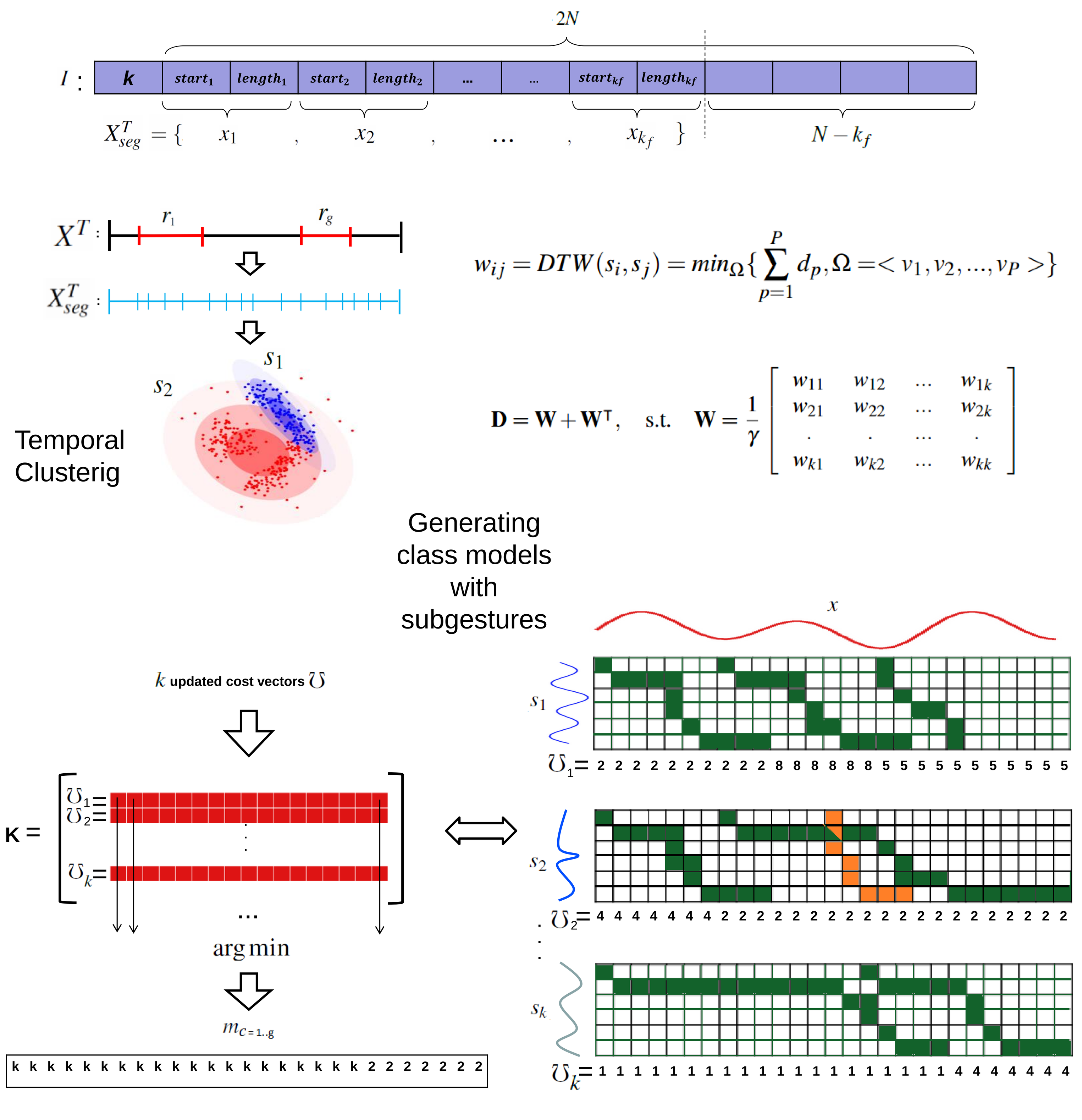
This paper introduces a framework for gesture and action recognition based on the evolution of temporal gesture primitives, or subgestures. Our work is inspired on the principle of producing genetic variations within a population of gesture subsequences, with the goal of obtaining a set of gesture units that enhance the generalization capability of standard gesture recognition approaches. In our context, gesture primitives are evolved over time using dynamic programming and generative models in order to recognize complex actions. In few generations, the proposed subgesture-based representation of actions and gestures outperforms the state of the art results on the MSRDaily3D and MSRAction3D datasets.

1. Motivation

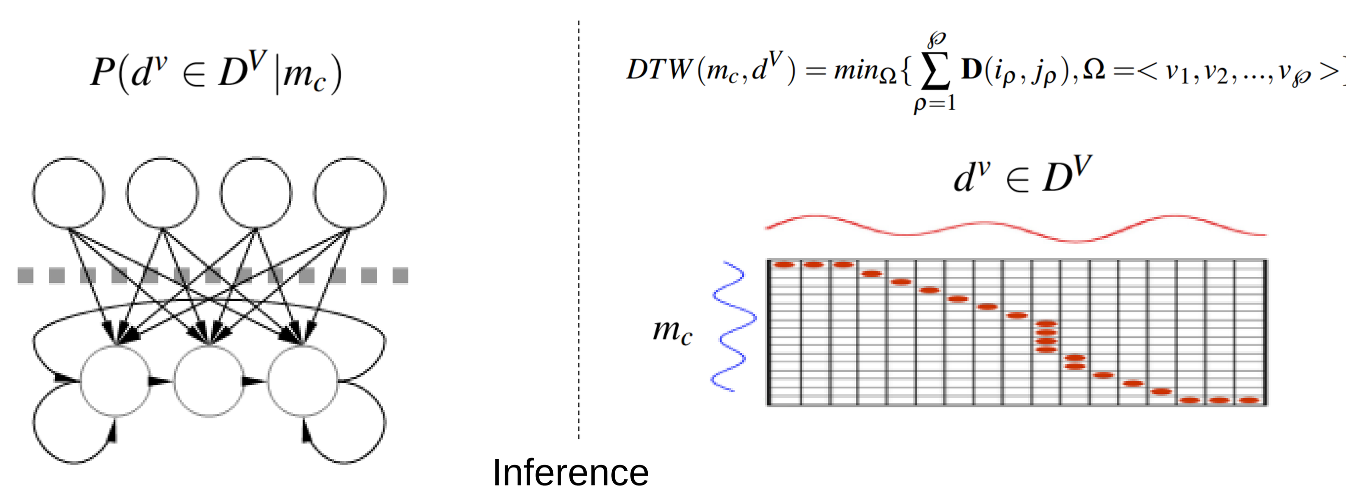
- Very recently, *evolutionary algorithms* have been also developed for key-frame *(Bag of Key Poses)* extraction [1-2].
- While these methods learn from a subset of frames, in *subgesture* modeling we aim at learning *spatio-temporal* units [3-6].
- *Class-specific* key poses/subgestures give a good performance. Nevertheless, we include the fact that there *are inter-class* subgestures that might be shared among different classes.
- We *evolve* such gesture *primitives* integrated into a *gesture recognition* framework coupled with either *DTW* or *HMM*.

2. Framework Overview

Selecting segments from the Population of the Genetic Algoritm



Obtain discrete vectors as subgestures indexes



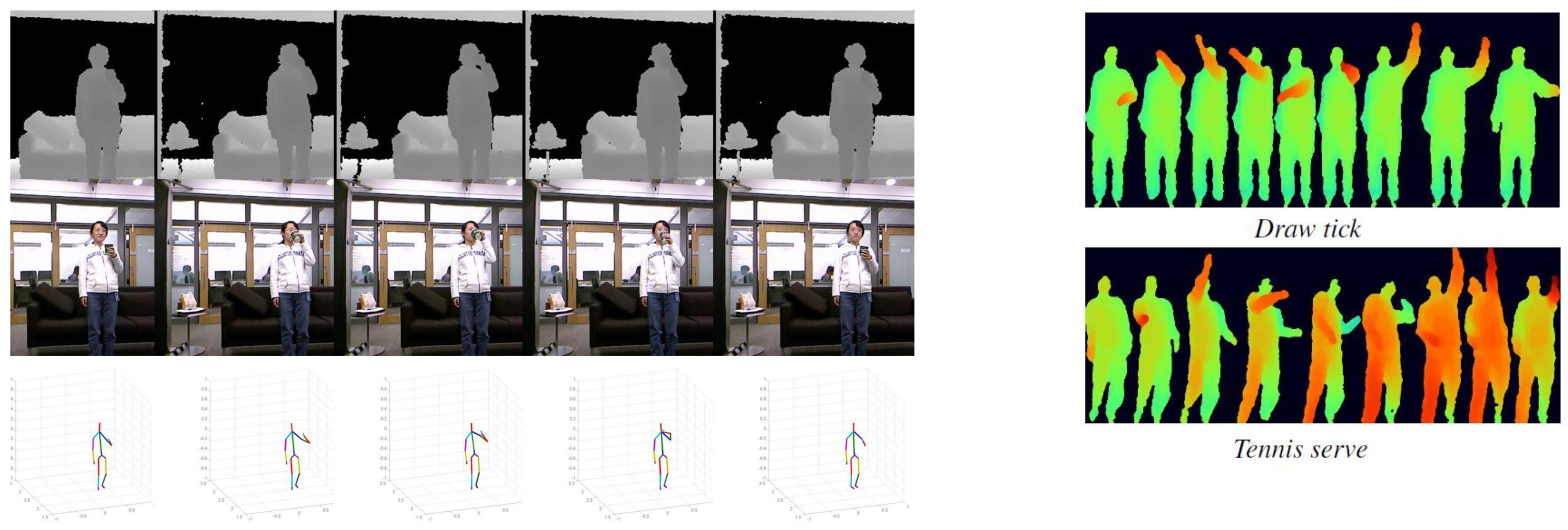
Hidden Markov Model

Inference

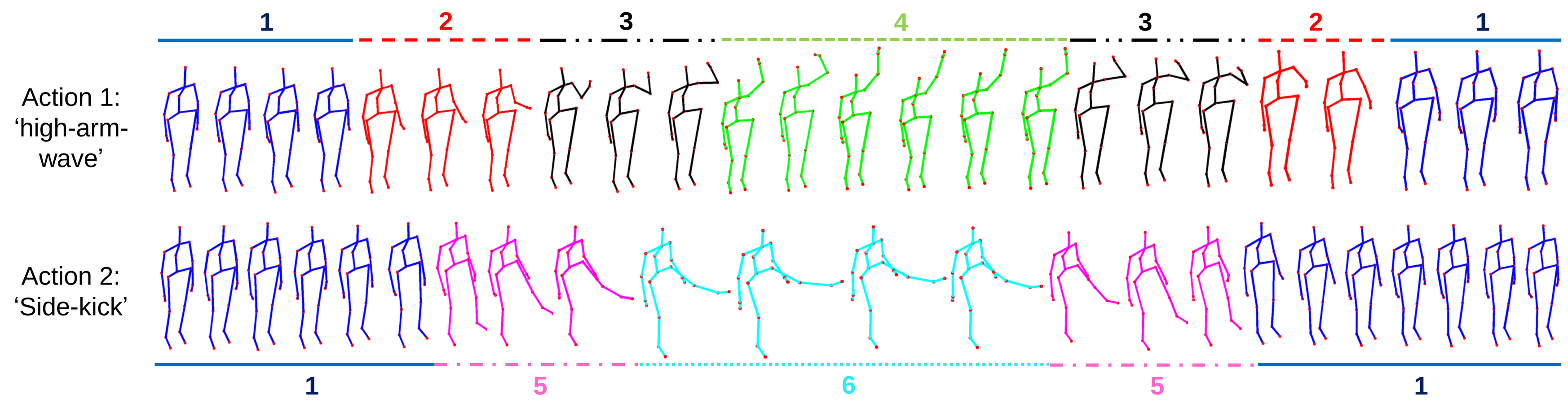
Dynamic Time Warping

3. Data and Results

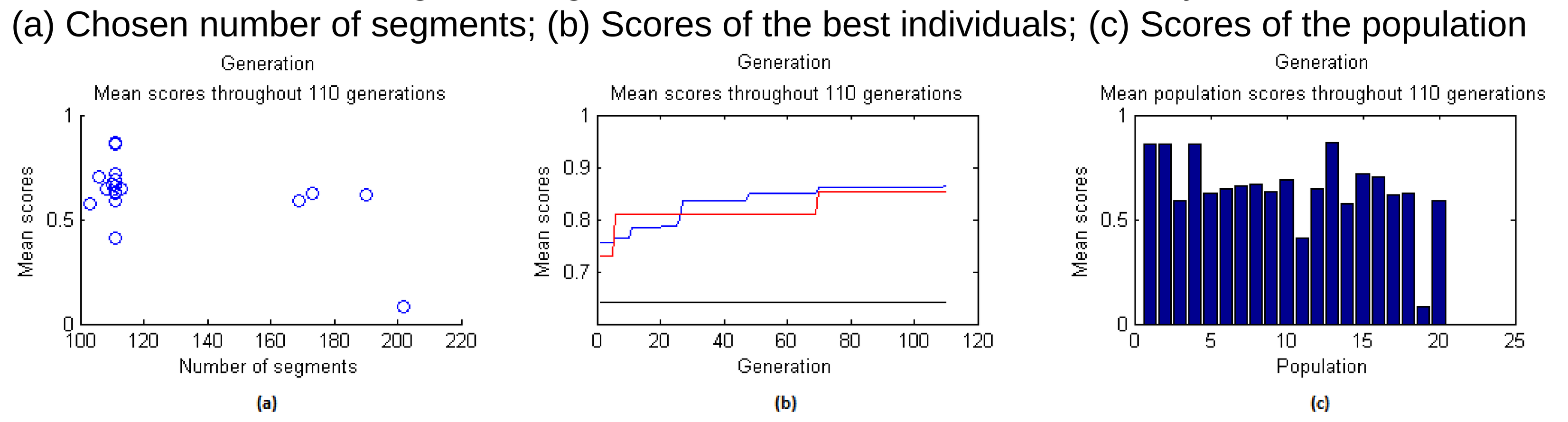
Examples of the MSR datasets



Recognition of an action test sequence given the subgesture models trained



Evolution of the genetic algorithm for the first fold of the MSRDaily3D dataset.
(a) Chosen number of segments; (b) Scores of the best individuals; (c) Scores of the population



Quantitative results of the recognition accuracy both on the Half-Split and on the 5-fold cross validation

MSRAction3D-HS		MSRDaily3D-CV		MSRDaily3D-HS	
Method	Accuracy	**Method**	Accuracy	**Method**	Accuracy
[32] (LOP+J.)	88.2%	[11] (SOSVM)	68.3%	[31] (LOP)	42.5%
[35] (DCSF)	89.3%	[12] (SMMED)	73.20%	[21] (DTW)	54%
[24] (HOPC)	91.64%	[35] (DCSF)	83.60%	[32] (MKL)	80.0%
[7] (PBR)	92.3%	[35] (DCSF+Skl.)	88.2	[16] (GP)	85.6%
[30] (MMTW)	92.7%	-	-	[32] (LOP+J.)	85.75%
Dynamic Time Warping					
Baseline	85.76%	Baseline	77.36%	Baseline	70.20%
Evolved	90.89%	Evolved	**89.51%**	Evolved	**88.16%**
Hidden Markov Model					
Baseline	70.85%	Baseline	74.62%	Baseline	69.29%
Evolved	**95%**	Evolved	**91.39%**	Evolved	**92.30%**

4. Conclusion

- We introduce a novel approach for learning dynamic gesture primitives for gesture and action recognition.
- An evolutionary computing framework is presented, which incorporates two most notable gesture recognition methodologies, namely DTW and HMMs.
- Experimental results show the competitiveness of our methods, outperforming state of the art results in benchmark data sets after few generations
- Proposed subgesture learning methodology enhances the recognition performance of traditional techniques.
- Future work includes extending the framework for related tasks (e.g. gesture spotting, event detection) and an extensive evaluation under different parameter settings.

References

[1] A.A. Chaaraoui and F. Florez-Revuelta. Adaptive human action recognition with an evolving bag of key poses. IEEE Transactions on Autonomous Mental Development, 6(2):139–152, 2014.
[2] H.J. Escalante, J. Martinez, S. Escalera, V. Ponce-López, and X. Baró. Improving the bag of visual words with genetic programming. In IJCNN, 2015.
[3] K. Li, J. Hu, and Y. Fu. Modeling complex temporal composition of actionlets for activity prediction. In ECCV, volume 7572, pages 286–299, 2012
[4] M. R. Malgireddy, I. Nwogu, S. Ghosh, and V. Govindaraju. A shared parameter model for gesture and sub-gesture analysis. In Combinatorial Image Analysis, volume 6636, pages 483–493, 2011.
[5] V. Ponce, M. Gorga, X. Baro, and S. Escalera. Human behavior analysis from video data using bag-of-gestures. In IJCAI, 2011.
[6] L. Wang, Y. Qiao, , and X. Tang. Video action detection with relational dynamic-poselets. In ECCV, 2014.