Gesture and Action Recognition by Evolved Dynamic Subgestures

Víctor Ponce-López [1,2,3]
http://victorponce.org

Hugo Jair Escalante [4]
http://hugojair.org/

Sergio Escalera [2,3]
http://www.maia.ub.es/~sergio/

Xavier Baró [1,2,3]
https://xbaro.wordpress.com/

[1] IN3, Open University of Catalonia
Rambla Poblenou, 156, 08018 Barcelona, Spain

[2] Dept. MAiA, University of Barcelona
Gran Via de Les Corts Catalanes, 585, 08007, Barcelona, Spain

[3] Computer Vision Center
Building O, Campus UAB, 08193, Bellaterra, Spain

[4] Computational Sciences Dept. INAOE
Luis Enrique Erro, 1, 72840, Tonantzintla, Puebla, Mexico

Gesture and human *action* recognition are two widely studied topics in computer vision and machine learning. In this work we perform gesture-action recognition base on the evolution of *temporal* gesture *primitives*, or *subgestures*. It 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. The underlying assumption is that whole gestures are composed by primitives (that can be shared or not among gestures from different categories), and the hypothesis is that learning with primitives leads to better recognition performance.

Very recently, evolutionary algorithms have been also developed for keyframe extraction [1, 2]. In these works, a bag-of-key-poses representation is adopted and an evolutionary algorithm was used to select the number of key-poses for the vocabulary (using $k-$means for clustering), the training set, features and parameters of the model (using DTW for recognition). These methods look for a subset of frames, whereas in subgesture modeling we aim at learning spatio-temporal units (subgestures) [3, 4, 5, 6]. On the other hand, many works of the literature assume and demonstrate that class-specific key poses/subgestures give a good performance. Nevertheless, we include the fact that some classes may contain or share similar subgestures [5]. Under this additional assumption, our method also reaches the state of the art performance and provides considerable improvements in gesture and action recognition domains.

In this work, a genetic algorithm is used to evolve gesture primitives integrated into a gesture recognition framework coupled with either DTW or HMMs. Different from most of the reviewed work, our approach obtains dynamic subgestures (*i.e.*, sequences of frames of different lengths) and simultaneously learns the parameters of the recognition model (either DTW or HMM). To consider an appropriate measure as DTW so as to treat temporal deformations and obtain subgestures, we implement a temporal clustering algorithm based on an extension of k-means that incorporates the temporal dimension both to represent and to cluster samples on the space and time. Then, we design each class model by means of representing each class sequence-samples in terms of subgestures, as illustrated in Figure 1. For the evaluation, we compute the mean score of classifying each sequence given the learned model parameters either for the DTW or HMM approaches.

In few generations, the proposed subgesture-based representation of actions and gestures achieves the state of the art performance on the MSR-Daily3D and MSRAction3D data sets, and outperforms previous methods for the different validation settings.

[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.

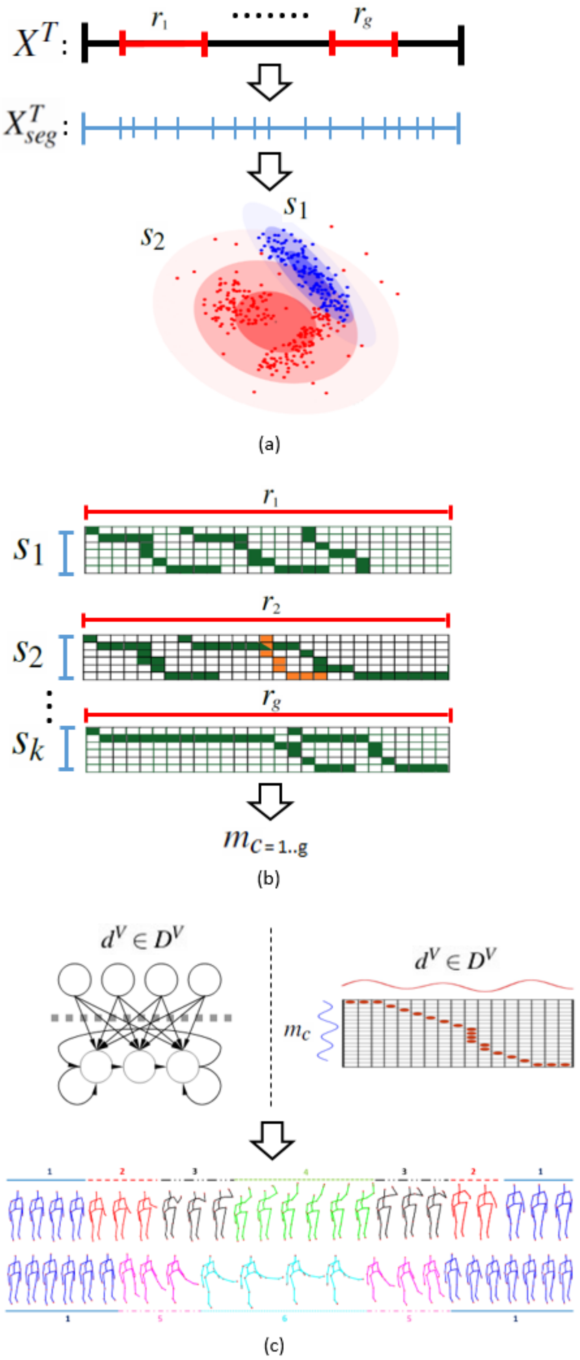

Figure 1: General diagram of the subgesture framework within the evolutionary procedure. In image (a), first we obtain the representants for each class from the whole training data X^T, and besides X^T is split in random segments to obtain subgestures by means of the aligned temporal clustering method. Image (b) shows the backward loop algorithm used both to design each class model $mc \in M$ for the DTW version, where M is the set of g class models $C = \{c_1, c_2, ..., c_g\}$, and to discretize the input sequences for the HMM version. Image (c) shows the recognition of an action/gesture test sequence given the models trained either with HMM or DTW, so that it can be represented as a set of subgesture models (best seen in color).