Recognizing Human Actions by Multi-Layer Growing Grid Architecture

Authors : Z. Gharaee

Abstract : Recognizing actions performed by others is important in our daily lives since it is necessary for communicating with others in a proper way. We perceive an action by observing the kinematics of motions involved in the performance. We use our experience and concepts to make a correct recognition of the actions. Although building the action concepts is a life-long process, which is repeated throughout life, we are very efficient in applying our learned concepts in analyzing motions and recognizing actions. Experiments on the subjects observing the actions performed by an actor show that an action is recognized after only about two hundred milliseconds of observation. In this study, hierarchical action recognition architecture is proposed by using growing grid layers. The first-layer growing grid receives the pre-processed data of consecutive 3D postures of joint positions and applies some heuristics during the growth phase to allocate areas of the map by inserting new neurons. As a result of training the first-layer growing grid, action pattern vectors are generated by connecting the elicited activations of the learned map. The ordered vector representation layer receives action pattern vectors to create time-invariant vectors of key elicited activations. Time-invariant vectors are sent to second-layer growing grid for categorization. This grid creates the clusters representing the actions. Finally, one-layer neural network developed by a delta rule labels the action categories in the last layer. System performance has been evaluated in an experiment with the publicly available MSR-Action3D dataset. There are actions performed by using different parts of human body: Hand Clap, Two Hands Wave, Side Boxing, Bend, Forward Kick, Side Kick, Jogging, Tennis Serve, Golf Swing, Pick Up and Throw. The growing grid architecture was trained by applying several random selections of generalization test data fed to the system during on average 100 epochs for each training of the first-layer growing grid and around 75 epochs for each training of the second-layer growing grid. The average generalization test accuracy is 92.6%. A comparison analysis between the performance of growing grid architecture and self-organizing map (SOM) architecture in terms of accuracy and learning speed show that the growing grid architecture is superior to the SOM architecture in action recognition task. The SOM architecture completes learning the same dataset of actions in around 150 epochs for each training of the first-layer SOM while it takes 1200 epochs for each training of the second-layer SOM and it achieves the average recognition accuracy of 90% for generalization test data. In summary, using the growing grid network preserves the fundamental features of SOMs, such as topographic organization of neurons, lateral interactions, the abilities of unsupervised learning and representing high dimensional input space in the lower dimensional maps. The architecture also benefits from an automatic size setting mechanism resulting in higher flexibility and robustness. Moreover, by utilizing growing grids the system automatically obtains a prior knowledge of input space during the growth phase and applies this information to expand the map by inserting new neurons wherever there is high representational demand.

Keywords : action recognition, growing grid, hierarchical architecture, neural networks, system performance