The paper introduces a method using CFG to recognize human action by stochastic parsing. The idea is to consider action as one low level primitives and high level structure. By the idea of CFG, high level structure provides a set of grammar that only generates specific strings. In syntactic recognition, only specific strings can pass through the state machine followed by the CFG rules.

Given a set of states, the sequence of terminals are decided by producing the maximum probability of Viterbi algorithm. A skip rule is added to the model to handle some noisy event to make the model more robust. The method only requires the consistence of internal states. The purpose is to keep the action recognition clean so that it reduces the probability of detecting a mixture of multiple actions. To keep the concurrency, a distance function will be applied on consecutive inner states or neighbor states around a particular states. But it also rises the problems such as partial occlusion. The high structure is dependent on low primitives, while the primitives should be clear to get good recognition accuracy.

However, the model is much more complicated than other methods. Generating the set of states could be expensive. Also the size of state set could be huge to achieve good accuracy.