

In this paper we present a CNN based approach for a real time 3D-hand pose estimation from the depth sequence. Prior discriminative approaches have achieved remarkable success but are facing two main challenges: Firstly, the methods are fully supervised hence require large numbers of annotated training data to extract the dynamic information from a hand representation. Secondly, unreliable hand detectors based on strong assumptions or a weak detector which often fail in several situations like complex environment and multiple hands. In contrast to these methods, this paper presents an approach that can be considered as semi-supervised by performing predictive coding of image sequences of hand poses in order to capture latent features underlying a given image without supervision. The hand is modelled using a novel latent tree dependency model (LDTM) which transforms internal joint location to an explicit representation. Then the modeled hand topology is integrated with the pose estimator using data dependent method to jointly learn latent variables of the posterior pose appearance and the pose configuration respectively. Finally, an unsupervised error term which is a part of the recurrent architecture ensures smooth estimations of the final pose. Experiments on three challenging public datasets, ICVL, MSRA, and NYU demonstrate the significant performance of the proposed method which is comparable or better than state-of-the-art approaches.