State-of-the-art approaches on hand pose estimation from depth images have reported promising results under quite controlled considerations. In this paper we propose a two-step pipeline for recovering the hand pose from a sequence of depth images. The pipeline has been designed to deal with images taken from any viewpoint and exhibiting a high degree of finger occlusion. In a first step we initialize the hand pose using a part-based model, fitting a set of hand components in the depth images. In a second step we consider temporal data and estimate the parameters of a trained bilinear model consisting of shape and trajectory bases. Results on a synthetic, highly-occluded dataset demonstrate that the proposed method outperforms most recent pose recovering approaches, including those based on CNNs.