This paper presents an effective method for continuous gesture recognition. The method consists of two modules: segmentation and recognition. In the segmentation module, a continuous gesture sequence is segmented into isolated gesture sequences by classifying the frames into gesture frames and transitional frames using two stream CNNs. In the recognition module, our method exploits the spatiotemporal information embedded in RGB and depth sequences. For the depth modality, our method converts a sequence into Dynamic Images and Motion Dynamic Images through rank pooling and input them to Convolutional Neural Networks respectively. For the RGB modality, our method adopts Convolutional LSTM Networks to learn long-term spatiotemporal features from short-term spatiotemporal features obtained by a 3D convolutional neural network. Our method has been evaluated on ChaLearn LAP Large-scale Continuous Gesture Dataset and achieved the state-of-the-art performance.