Abstract

Augmented Reality (AR) technology has received great attention in the domain of assembly and maintenance tasks. This study investigates the development of a smart Augmented Reality (AR) task support system with an emphasis on point-of-completion detection and correctness classification. The objective of the study is to improve manual assembly operations by giving workers feedback about the current status of the task. Traditional AR experiences mostly depend on fixed instructions and lack the use of automatic feedback to its full strength. In this study, an advanced approach is developed to precisely detect the point of completion in complex manual assembly task by combining Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and a temporal voting scheme. In order to enhance the efficiency and accuracy of manual assembly tasks in a real-time AR environment, this research focuses on developing a robust system that integrates both point-of-completion detection and correctness classification. To accomplish this, dataset made up of video recordings of chair assembly jobs performed by users was gathered. The dataset was annotated carefully including the labeling of different substeps and their corresponding completion frames. The proposed models were trained and tested using this annotated dataset as the benchmark. The first phase of the research deals with the training of models for point-of-completion detection. A separate model was trained for each of the four assembly sub-steps to find the exact frame when each sub-step is completed. To achieve accurate detection results, a number of deep learning architectures, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), were trained and tested but, we got some excellent results by employing a temporal voting scheme. The models were validated using common performance metrics, such as accuracy, precision, recall, and F1 score, after being trained on the collected dataset. The training of correctness classification models was the main goal of the second phase. Once the point of completion was detected in a video stream, the target video clip was passed through a separate model to determine the correctness of the assembly task. The trained models showed excellent results, with accuracy, precision, recall, and F1 scores ranging from 0.80 to 0.91 across the different assembly steps. These results indicate the effectiveness of the proposed framework in accurately detecting the point of completion and evaluating the correctness of chair assembly tasks.