Chapter 3
3D Technologies and Applications in Sign Language

Kiriakos Stefanidis
https://orcid.org/0000-0002-9922-1916
The Visual Computer Lab, Information Technologies Institute, Centre for Research and Technology, Hellas, Greece

Dimitrios Konstantinidis
https://orcid.org/0000-0002-7391-6875
The Visual Computing Lab, Information Technologies Institute, Centre for Research and Technology, Hellas, Greece

Athanasios Kalvourtzis
The Visual Computing Lab, Information Technologies Institute, Centre for Research and Technology, Hellas, Greece

Kosmas Dimitropoulos
The Visual Computing Lab, Information Technologies Institute, Centre for Research and Technology, Hellas, Greece

Petros Daras
https://orcid.org/0000-0003-3814-6710
The Visual Computing Lab, Information Technologies Institute, Centre for Research and Technology, Hellas, Greece

ABSTRACT

Millions of people suffering from partial or complete hearing loss use variants of sign language to communicate with each other or hearing people in their everyday life. Thus, it is imperative to develop systems to assist these people by removing the barriers that affect their social inclusion. These systems should aim towards capturing sign language in an accurate way, classifying sign language to natural words and representing sign language by having avatars or synthesized videos execute the exact same moves that convey a meaning in the sign language. This chapter reviews current state-of-the-art approaches that attempt to solve sign language recognition and representation and analyzes the challenges they face. Furthermore, this chapter presents a novel AI-based solution to the problem of robust sign language capturing and representation, as well as a solution to the unavailability of annotated sign language datasets before limitations and directions for future work are discussed.

DOI: 10.4018/978-1-5225-5294-9.ch003

INTRODUCTION

Sign language is the only way of communication among deaf or hearing-impaired people around the world. The importance of sign language can be illustrated by the fact that it is used not only for the communication among hearing-impaired people, but also for the interaction of hearing-impaired people with their environment and speaking people. Sign language enables people with hearing loss to remove the barriers to the use of mainstream products and services, such as TV and other media and improves their social inclusion by giving them equal opportunities with people having no hearing loss problems. Sign language is a visual-spatial language based on positional and visual components, such as the shape of fingers and hands, the location and orientation of the hands, arm and body movements and facial expressions. These components are combined together to form utterances that convey the meaning of words or sentences (Kendon, 2004).

Being able to capture and understand the relation between utterances and words is crucial for the deaf community in order to guide us to an era where the translation between utterances and words can be achieved automatically. To enable such translation, automatic systems should be developed that include methodologies for precise capturing of the hand and facial movements, robust recognition and classification of utterances to words and accurate representation of the utterances in a 2D environment using synthesized signed videos or in a 3D environment, using 3D avatars that represent humans performing utterances.

Unfortunately, current sign language capturing techniques either depend on too costly or complicated sensor setups or suffer from significant hand and finger occlusions. Moreover, sign language recognition methodologies face significant challenges as: a) each country has its own sign language, b) there is a lack of large publicly available annotated sign language recognition datasets, c) there are variations in the way people sign based on their individual signing style and d) each sign language consists of thousands of signs that can differ by subtle changes in hand, finger and facial movements. Finally, the 3D representation of utterances using avatars is challenging due to the mediocre accuracy of computed human skeletal joints, the lack of avatar realism and the problems in translation from skeletal points in the 3D space to quaternions. On the other hand, the synthesis of signed videos is still at its infancy due to lack of large datasets, although current deep learning approaches has shown promising results.

The need of deaf community for accurate and robust sign language recognition and representation makes the development of such systems imperative. However, the great many challenges these systems face put significant limitations and obstacles to current sign language recognition and representation systems. Although the sign language recognition and representation field has captured the attention of the research community and a lot of progress has been made in this field, there is still a lot of work that should be performed towards improving the accuracy and robustness of sign language recognition and representation systems.

In the next sections, we initially present a literature review of the state-of-the-art sign language capturing, recognition and representation approaches before we move on and present our proposed Artificial Intelligence (AI) – based approach to solve the sign language 3D animation task based on capturing and 3D representation. Finally, we also present our approach towards collecting annotated signed videos to form a large sign language repository based on the notion of crowdsourcing.