Information Sciences 281 (2014) 762–780

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

An approach to generate and embed sign language video tracks into multimedia contents

Tiago Maritan U. de Araújo *, Felipe L.S. Ferreira, Danilo A.N.S. Silva, Leonardo D. Oliveira, Eduardo L. Falcão, Leonardo A. Domingues, Vandhuy F. Martins, Igor A.C. Portela, Yurika S. Nóbrega, Hozana R.G. Lima, Guido L. Souza Filho, Tatiana A. Tavares, Alexandre N. Duarte

Digital Video Applications Lab (LAViD), Federal University of Paraíba, Paraíba, Brazil

ARTICLE INFO

Article history:
Received 1 February 2013
Received in revised form 6 March 2014
Accepted 8 April 2014
Available online 19 April 2014

Keywords:
Accessible multimedia content
Brazilian sign language
Machine translation
Accessible technologies for the deaf
Sign synthesis

ABSTRACT

Deaf people have serious problems to access information due to their inherent difficulties to deal with spoken and written languages. This work tries to address this problem by proposing a solution for automatic generation and insertion of sign language video tracks into captioned digital multimedia content. Our solution can process a subtitle stream and generate the sign language track in real-time. Furthermore, it has a set of mechanisms that exploit human computation to generate and maintain their linguistic constructions. The solution was instantiated for the Digital TV, Web and Digital Cinema platforms and evaluated through a set of experiments with deaf users.

1. Introduction

Deaf people naturally communicate using sign languages. As a result, many of them have difficulties in understanding and communicating through texts in written languages. Since these languages are based on sounds, most of them spends several years in school and fail to learn to read and write the written language of their own country [36]. Reading comprehension tests performed by Wauters [40] with deaf children aged 7–20 in the Netherlands showed that only 25% of them read at or above the level of a nine-year-old hearing child. In Brazil, about 97% of the deaf people do not finish the high school [21].

In addition, the Information and Communication Technologies (ICT) rarely address the specific requirements and needs of deaf people [16]. The support for sign language, for example, is rarely addressed in the design of these technologies. On TV, for example, sign languages support is generally limited to a window with an interpreter presented along with the original video program (wipe). This solution has high operational costs for generation and production (cameras, studio, staff, etc.) and requires full time human interpreters, which reduces their presence to a small portion of the TV programming. Furthermore, this traditional approach is not feasible for platforms with dynamic contents such as the Web. These difficulties result in major barriers to communicate, to access information and to acquire knowledge.

* Corresponding author. Tel.: +55 8332167093.

E-mail addresses: maritan@lavid.ufpb.br (Tiago Maritan U. de Araújo), lacet@lavid.ufpb.br (F.L.S. Ferreira), danilo@lavid.ufpb.br (D.A.N.S. Silva), leodantas@lavid.ufpb.br (L.D. Oliveira), eduardolf@lavid.ufpb.br (E.L. Falcão), leonardo.araujo@lavid.ufpb.br (L.A. Domingues), vandhuy@lavid.ufpb.br (V.F. Martins), igor.portela@lavid.ufpb.br (I.A.C. Portela), yurika@lavid.ufpb.br (Y.S. Nóbrega), hozana@lavid.ufpb.br (H.R.G. Lima), guido@lavid.ufpb.br (G.L. Souza Filho), tatiana@lavid.ufpb.br (T.A. Tavares), alexandre@lavid.ufpb.br (A.N. Duarte), alexandre@lavid.ufpb.br (A.N. Duarte).

http://dx.doi.org/10.1016/j.ins.2014.04.008
0020-0255/

The scientific literature includes some works addressing the communication needs of the deaf [17,18,26,27,35]. These works offer technological solutions for daily activities enabling deaf people to watch and understand television, to interact with other people, to write a letter, among others. The use of dynamic [17,18] and emotive captioning [26] in movies and television programs and the development of games for training deaf children [27] are examples of this type of solution.

Other works deal with machine translation for sign languages [4,13,19,20,30–33,38,42]. Veale et al. [38], for example, proposed a multilingual translation system for translating English texts into Japanese Sign Language (JSL), American Sign Language (ASL) and Irish Sign Language (ISL). The work explores and extends some Artificial Intelligence (AI) concepts to sign languages (SL), such as, knowledge representation, metaphorical reasoning, among others [30], but there is no testing or experimentation to evaluate the solution. Then, it is not possible to draw conclusions about its feasibility and translation speed and quality.

Zhao et al. [42] developed an interlanguage-based approach for translating English text into American Sign Language (ASL). It analyses the input data to generate an intermediate representation (IR) from their syntactic and morphological information. Then, a sign synthesizer uses the IR information to generate the signs. However, as well as in Veale et al.'s work [38], the solution lacks experimental evaluation. Morrissey [30] proposed an example-based machine translation (EBMT) system for translating text into ISL. However, the data set was developed from a set of "children's stories", which restricts the translation for that particular domain.

Fotinea et al. [13] developed a system for translating Greek texts into Greek Sign Language (GSL). This work uses a transfer-based approach for generate the sentences in GSL, but it main focus is the strategy of animation that explores the parallel structures of sign languages (e.g., the ability to present a hand movement with a facial expression simultaneously). To perform this task, a 3D avatar was developed to explore the parallel structure of sign languages. However, no testing or experimentation was conducted to evaluate it translation speed and quality.

Huenerfauth et al. [19,20] proposed modeling classifiers predicate[1] in a English to American Sign Language (ASL) translation system. Some tests performed with deaf users showed that contents exploring the use of classifier predicates (generated by the Huenerfauth solution) were significantly more natural, grammatically correct and understandable than the contents based on direct translation. The translation speed, however, was not evaluated by author.

Anuja et al. [4] proposed a system for translating English speech into Indian Sign Language (ISL) focused on helping deaf people to interact in public places, such as banks and railroads. The system also uses a transfer-based approach for translating speech entries into ISL animations. This solution is restricted to a specific domain and according to authors it takes a long (and unacceptable) time to generate the translation (the time values, however, were not described in the work).

San-Segundo et al. [31–33] proposed an architecture for translating speech into Spanish Sign Language (LSE) focused on helping deaf people when they want to renew their identity card or driver's license. This translation system consists of three modules: a speech recognizer, a natural language translator and an animation module. However, as well as in Anuja, Suryapriya and Idicula work, this solution is also restricted to a particular (or specific) domain and the time needed for translating speech into LSE (speech recognition, translation and signing) is around 8 s per sentence, which makes the solution unfeasible for real time domains (e.g., television).

These works can be separated in two classes: one class of works that translate speech in the source spoken language to the target sign language (i.e., they use speech recognition) [4,31–33], and other class that translate written texts to the target sign language (i.e., they do not use speech recognition) [13,19,20,30,38,42]. However, all these works have some limitations. The class of works that use speech recognition [4,31–33], for example, are just applied to specific domains and are not efficient considering signing and translation speed. Other works do not have an assessment of the feasibility and quality of the solution [13,38,42] or are also applied to specific domains [19,20,30]. These limitations reduce their applicability to real-time and open-domain scenarios, such as TV.

Another difficulty is that the development of their linguistic constructions (translation rules, signs dictionary, etc.) is in general a non-trivial task and requires much manual work. Moreover, as sign languages are natural and living languages, new signs and new grammatical constructions can arise spontaneously over time. This implies that these new signs and constructions must also be included in the solution, otherwise the quality of content generated by it tend to deteriorate over time, making it outdated.

To reduce these problems, in this paper, we propose a solution to generate and embed sign language video tracks in multimedia contents. Our current implementation targets the Brazilian Sign Language (LIBRAS), but we believe that the general solution can be extended for other target sign languages. The LIBRAS video tracks are generated from the translation of subtitle tracks in Brazilian Portuguese and are embedded in the multimedia content as an extra layer of accessible content.

A 3D avatar reproduces the signs and the solution also explores human computation strategies to allow human collaborators to generate and maintain their linguistic constructions (translation rules and signs). The implementation utilizes also a set of optimization strategies, such as a textual machine translation strategy for Brazilian Portuguese to gloss (a LIBRAS textual representation), which consumes little computational time, and LIBRAS dictionaries to avoid rendering the signs in real time, reducing the computational resources required to generate the LIBRAS video.

[1] Classifiers are linguistic phenomena used by sign language interpreters to make the signs more natural and easier to understand. They make use of the space around the signer in a topologically meaningful way. The interpreter's hands represent an imaginary entity in space in front of them, and they position, move, trace our re-orient this imaginary object to indicate location, movement, shape, among others [19].