Web-Based Sign Language Synthesis and Animation for On-line Assistive Technologies

Zdeněk Krňoul
University of West Bohemia, Faculty of Applied Sciences, Dept. of Cybernetics
Univerzitní 8, 306 14 Pilsen, Czech Republic
zdkrnoul@kky.zcu.cz

ABSTRACT

This article presents recent progress with design of sign language synthesis and avatar animation adapted for the web environment. New 3D rendering method is considered to enable transfer of avatar animation to end users. Furthermore the animation efficiency of facial expressions as part of the non-manual component is discussed. The designed web service ensures on-line accessibility and fluent animation of 3D avatar model, does not require any additional software and gives a wide range of usage for target users.

Categories and Subject Descriptors

I.3.7 [**computer graphics**]: Graphics Systems—*Three- Dimensional Graphics and Realism*; H.5.1 [**information interfaces and presentation**]: Multimedia Information Systems—*Animation*

General Terms

Algorithms, Languages, Design

Keywords

Avatar, Sign language, Deaf, Web, Computer graphics

1. INTRODUCTION

Current assistive technologies allow automatic conversion of text to speech not only in audio form but there is ongoing research on sign language synthesis. Individuals with hearing disabilities will be able to use these technologies in their everyday lives. The first attempts to use the Internet and three-dimensional (3D) avatar animation indicate facilitation of access to these technologies: Vcom3D, SignSynth [1], eSIGN [3], SignStep [6].

We introduce new client-server framework for sign language synthesis incorporating up-to-date 3D computer graphics for the Internet, central administration of the system and easy use which allow signing avatar to be more accessible for the target users. The framework is experimentally implemented in the on-line sign language dictionary of Czech Sign Language (CSL) available at `http://signs.zcu.cz/`.

ASSETS'11, October 24–26, 2011, Dundee, Scotland, UK.
ACM 978-1-4503-0919-6/11/10.

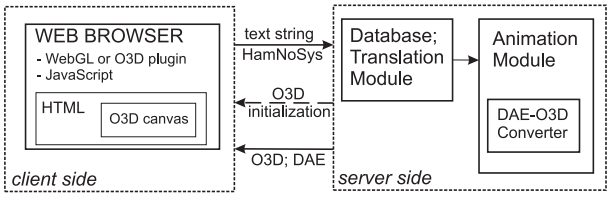

Figure 1: The schema of the web-based sign language synthesis system.

2. SYSTEM OVERVIEW AND USAGE

The web-based sign language synthesis system is composed of a client and server side part. The client side consists of a common web browser only. The server side receives all requirements of the client side, shares conversion methods and generates adequate answers. Currently the server side incorporates a animation module, a database for word-by-word transliteration, a cache of processed requirements and a data converter. The schema of the system is shown in Figure 1. The main features of this framework are:

- minimum requirements for installing additional software on the end user's computer;
- effective data transfer across the network;
- rapid response of the system and fluent animation of the 3D model.

Accessibility of the synthesis system as a web service has required a different representation of the 3D model and the rendering method in comparison with the off-line version [4]. The rendering method and graphic data are based on Google O3D[1] technology. O3D is an open-source Java Script (API) enabling to use signing avatar as interactive 3D applications that needs to run HTML based web browsers only (on Windows, Macintosh, Linux or Android platforms). This API was originally designed as a plug-in (JavaScript library) for the web browsers and now has evolved into the standalone JavaScript library using WebGL standard[2].

We support two forms of data transfer in the direction from the client side. The first form is standard text representation of a spoken language, such as word "house". These words are searched in the database and translated to the

[1] `http://code.google.com/p/o3d/`
[2] WebGL is a framework for the canvas HTML element providing a 3D computer graphics API.

suitable input of the animation module. The system currently provides automatic word-by-word transliterating spoken Czech into signs of CSL and does not produce a fluent sign language output. However additional translation modules can be incorporated into the server side. The second form of the data transfer consists of a stream of signs represented by the sign language notation system and directly connected to the animation module.

The current implementation of the animation module supports the input of one or more signs in the HamNoSys[3] notation. The detailed description of the conversion process is in [4]. The main modification of the animation module consists of its output. Instead of direct 3D rendering the module generates computer data files. We consider O3D file format and DAE (Collada) file[4].

Both file formats can be transfered in the direction to the client side. O3D file is obtained by simple conversion from the DAE file and is primarily used for 3D rendering in the web browsers. In contrast the DAE file can be optionally saved by the end user as "3D sign utterance" for other off-line applications. The animation module is implemented in the C programming language and provides very fast response.

3. CURRENT INVESTIGATION

We are currently concerned with proper animation of facial expressions as essential part of the sign language. The export of animation to DAE file format requires next changes in comparison with to the off-line version that implements 3D rendering via influence zones [4]. The Collada specification provides two different animation methods: the skin and morph controller [5]. The skin controller is primarily intended for animation of arms, fingers and neck. For animation of a human face we have to use the morph controller.

The morph controller is a 3D computer per-vertex animation using blend shapes (the morph targets). The morph target is a "deformed" version of the entire polygonal mesh, again stored as a series of their vertices. In every key-frame of animation the vertices are interpolated between these targets as the weighted average. When we apply it to the human face, one morph is used for neutral face ("basic face shape") and single or multiple morph targets are considered as individual facial expressions. For example for spoken English it requires to store at least 14 morph targets for each viseme[5]. Furthermore the sign languages include many more facial expressions. A large number of morph targets increase memory and rendering requirements on the client side as well as the data transfer because it need to transfer not only all morph targets but also their animation trajectories.

Principal component analysis (PCA) is a mathematical algorithm that uses orthogonal transformations for reduction of dimension and conversion apparently correlated data into an uncorrelated set of values so called principal components. PCA can be applied on the data measured at the face of a signing speaker via Motion capture systems or using video records via the active shape or appearance model (ASM, AAM) [2]. If we consider the basic face shape equal to "mean shape" analyzed by PCA and each of the principal components such as one morph target then principle is identical with the interpolation used by the morph controller. Initial results for isolated signs of CSL and mouthing indicate that to maintain 97.5% of the variance observed on the face can be animated just by nine morph targets.

[3] `http://www.sign-lang.uni-hamburg.de/projects/hamnosys.html`

[4] Collada is XML schema allowing full description of the sining avatar and flexible processing by the synthesis module.

[5] A viseme is a representational unit to classify speech sounds in the visual domain.

4. CONCLUSIONS

The article presents a framework for the web-based sign language synthesis system and its experimental implementation for Czech Sign Language (`http://signs.zcu.cz/`). The the system provides 3D rendering of the signing avatar in real time which is supported by most of the recent web browsers without any installation of additional software. The end users can more easily receive prearrange on-line contents containing together standard websites and the 3D signing avatar.

The proposed solution uses hardware accelerated graphics and does not cause fitful rendering. Furthermore efficient and precise rendering of the non-manual component is discussed. The initial analysis indicates potential to increase accessibility by reduction of necessary data. The synthesized utterance can also be optionally saved by the end user and used for other communication or educational purposes. Other specially designed mobile applications can be hereby considered. Next research have to be focused to get high quality signing that will be easily understandable.

5. ACKNOWLEDGMENTS

This research was supported by the Grant Agency of the Czech Republic, project No. GAČR 102/09/P609. and by the Grant Agency of Academy of Sciences of the Czech Republic, project No. 1ET101470416.

6. REFERENCES

[1] A. B. Grieve-Smith. Signsynth: A sign language synthesis application using web3d and perl. In *Gesture Workshop*, pages 134–145, 2001.

[2] M. Hrúz, Z. Krňoul, P. Campr, and L. Müller. Towards automatic annotation of sign language dictionary corpora. *Lecture Notes in Artificial Intelligence*, LNAI 6836, in press, 2011.

[3] J. R. Kennaway, J. R. W. Glauert, and I. Zwitserlood. Providing signed content on the internet by synthesized animation. *ACM Trans. Comput.-Hum. Interact.*, 14, September 2007.

[4] Z. Krňoul, J. Kanis, M. Železný, and L. Müller. Czech text-to-sign speech synthesizer. In *Proceedings of the 4th international conference on Machine learning for multimodal interaction*, MLMI'07, pages 180–191, Berlin, Heidelberg, 2008. Springer-Verlag.

[5] F. I. Parke and K. Waters. *Computer facial animation*. A. K. Peters, Ltd., Natick, MA, USA, 1996.

[6] L. van Zijl and J. Fourie. The development of a generic signing avatar. In *Proceedings of the IASTED International Conference on Graphics and Visualization in Engineering*, GVE '07, pages 95–100, Anaheim, CA, USA, 2007. ACTA Press.