

# Video-Based Sign Language Content Annotation by Incorporation of MPEG-7 Standard

Rashad Aouf and Steve Hansen

<sup>1</sup> University of Western Sydney, School of Computing and Information Technology,  
Sydney, Australia  
[{r.aouf,s.hansen}@uws.edu.au](mailto:{r.aouf,s.hansen}@uws.edu.au)

**Abstract.** The advanced progress in multimedia technology increases the demand on delivering effective content in term of quality with the ability to describe content. From the W3C initiative into the web accessibility (WAI), there is a dedicated effort to make data accessible by every person even by people with disabilities. Accordingly, this paper balances the portion between the minimum bandwidth and the optimum required data to display customized video-based sign language. It also describes a systematic approach derived from the MPEG-7 multimedia content description standard to annotate sign language information. A new approach is proposed by this paper. It makes use of an “intermediary” signage object rather than immediate transmission of sign language video clips. Based on the signage object, this research analyses the components in order to enhance the display quality for video-based sign language with less data consumption by determining the accurate display parameters.

## 1 Introduction

Delivering multimedia data on a Web client is currently a significant area of research due to the bandwidth bottleneck [15] in particular with streaming video. The current compression technology (e.g. MPEG) proved its efficiency in minimizing the data rate that is relevant to ADSL and ISDN [15] [7] internet connection which are still expensive. However this technology is still not enough to deliver multimedia data with good quality for 56 kbps modem connection which is commonly being used to access internet. Meanwhile accessibility is still a current issue in the web environment in term of disability. The WAI [5] specified set of scenarios to access telecommunication systems by disabled people such as video captioning for deaf [4]. From the web accessibility issue and the enhancement in generic Web blended multimedia data presentation this research is looking in a systematic approach to provide effective sign language transmission with the minimum bandwidth needed.

Making use of the intermediary object, rather than immediate transmission of sign language video clips, provides access to sign language related information and allow enhancement in content (e.g. include graphics or animation) and minimize data (e.g. adjust display parameters – frame rates, resolution, and colors). This work make use of the MPEG-7 Multimedia Description Scheme [14] (MDS) to manage signage information.

This paper is organized as follows. Section 2 introduces sign language status in the web. Section 3 outlines the proposed signage information system. Section 4 explains in detail the system architecture. Section 5 presents a summary and conclusion, and declaration of future work.

## 2 Sign Language

Sign language [2] is the natural approach of communication for deaf and hearing impaired people and providing sign language information within the Web will assist these people to access telecommunication systems.

Nowadays studies that focus in digitally handling sign language in the web environment is extremely limited (ViSiCast [6], DePaul University Project [16] [17], Dublin City University Project [18], and Gallaudet University[19]) and need more contribution from researchers around the continents while the statistics show a trend for these people to use web services.

Sign language has wealth of information to be clearly shown in video-based sign language [1] without jerky object movements and low-temporal resolution [10]. The next sections of this paper show how it is possible to adjust display quality parameters (frame rates, resolution, and colors) in order to minimize data while respecting the acceptable appearance of sign language attributes within a video-based sign language.

## 3 Proposed System

Currently, it is aimed to integrate sign language information system into the Web environment making use of XML-based technology [3] [8] (i.e. MPEG-7 standard). The goal of this research is to produce a signage information system that will balance the portion between the minimum bandwidth and the optimum required data to be displayed in video-based sign language transmission.

As outlined above, an intermediary object is defined to store multimedia data extracted from the original message (i.e. video-based sign language). This object is wrapped with additional information related to sign language such as users' preferences (e.g. display parameters, sign language components, etc), speech acts (e.g. interrogation, exclamation, etc) and emotional state (e.g. happiness, sadness, etc).

The role of the system is to improve in the presentation of the original message in term of display quality with minimum data rate transmission by associating of other web objects such as graphics (i.e. style) that supports the content of the description.

## 4 System Architecture

The system is built on client/server environment. This paper focuses on the server side in term of describing data modularity. The server contains 4 modules: MPEG-7 Framework, AV Storage System, User Information Storage System, and Intermediate Signage Object Generation.

### 4.1 MPEG-7 Framework

This research examines the extension of MPEG-7 Multimedia Description Scheme (MDS) [14] tools to include sign language attributes. However, MPEG-7 is considered as a library of description tools whereas creation and application of the descriptors are outside its scope. Therefore, the application specific domain [12] has to extract and extend MPEG-7 data model and select an appropriate subset of MPEG-7

Framework. According to the MPEG-7 Requirements Document [13] MPEG-7 framework consists of Descriptors Ds (low level features such as colors, texture, etc), Description Schemes DSs (high level features such as audio visual segment, etc), and the Description Definition language DDL [9] (i.e. XML-Schema Document) to extend or define new DSs and Ds to describe sign language AV content.

MPEG-7 Framework is addressed by two associated areas: AV Storage System and Users' Information Storage System. More details are given in the following sections.

## 4.2 AV Storage System

This system is to store video based sign language content and their description documents. It consists of two main parts: the AV content and their associated MPEG-7 description documents.

The MPEG-7 description documents provide information about the AV content making use of MPEG-7 description schemes and descriptors related to the AV content structure.

However, the description process produced a document that describes the content in generic form except with the “TextAnnotation” DS [11], free text is used to annotate low level sign language information. High level sign language attributes require extending normalized DSs through the DDL schema (section 4.4).

## 4.3 User Information Storage System

The user information storage system is an essential module while making a decision about the display quality of the AV content. This model consists of two parts: User's Information and MPEG-7 Preferences.

The user's information is a set of data collected from the user and then transformed into MPEG-7 descriptions (i.e. User Interaction Tools). This information is used in order to customize the display quality and minimize data consumption.

A DDL document has been generated to extend existed DSs in order to define the description structure, relationship among description schemes or descriptors and to constraint the values. Section 4.4 shows two samples of MPEG-7 XML Schemas.

## 4.4 Intermediate Signage Object Generation

The intermediary signage object is the generic object, defined above, wrapped with additional attributes that are extracted to enhance the original signed message. Attributes that are related to sign language are defined in the DDL XML-based schema. Existed DSs are selected to be used with the DDL for extension in order to include high level interpretation of audiovisual sign language.

From the signer's face, high level sign language attributes can be extracted to define other related attributes such as style and speech action. The intermediary object that contains these attributes will be associated with other web objects (e.g. animation) in order to enhance in the appearance of the signed message. The other attributes will be treated to ensure clear appearance of sign language parameters in the video clips in complicated conditions such as busy background, clothes color not contrast with skin color or even other objects that hide parts of the signer.

The following fragment of MPEG-7 Schema Definition (DDL) shows how to extend the MovingRegion DS that defined as PersonType in the code above to include additional elements related to sign language such as the “EmotionalState” and the “SpeechActs” elements and attributes related to sign language video clips.

```
<complexType name="MovingRegionType" final="#all">
<complexContent><extension base="mpeg7:VisualDType">
<element name="EmotionalState">
<simpleType><restriction base="mpeg7:ModelStateType">
<pattern value="Happy|Sad|Angry"/>
</restriction></simpleType></element>
<element name="SpeechActs">
<simpleType><restriction base="mpeg7:TextType">
<pattern value="Interrogation| Exclamation| Af-
firmative"/>
</restriction></simpleType></element>
</extension></complexContent></complexType>
```

**Code 1.** Extends MovingRegion DS to include SignerFace

The following fragment of MPEG-7 Schema Definition (DDL) shows how to extend another DS called “MediaProfiles” that belongs to the MPEG-7 Content Management Tools [14] to include additional elements related to sign language display quality such as the “Resolution”, “FrameRate”, and “Color”.

```
<complexType name="MediaInformationType" final="#all">
<complexContent>
<extension base="mpeg7:MediaProfilesType">
<element name="Resolution">
<simpleType><restriction base="mpeg7:VectorType">
<length value="2" /></restriction>
</simpleType><element name="FrameRate">
<simpleType><restriction base="mpeg7:unsigned6">
<pattern value="5|10|15|25"/>
</restriction></simpleType></element>
<element name="Color"><simpleType>
<restriction base="mpeg7:ColorSpaceType">
<pattern value="RGB|GrayScale|BlackAndWhite" />
</restriction></simpleType></element>
</extension></complexContent></complexType>
```

**Code 2.** Extends the MediaProfiles DS within MediaInformaiton

Finally, the user’s preferences are also taken into account in order to determine the final MPEG-7 content description document. User’s preferences are sent to the multimedia servers to be combined with the original signage object and “actively” recommend customized signage content with minimum data rate transmission.

## 5 Summary and Conclusion

This paper gives an overview of a project being investigated into methodologies to integrate signage information into the web. Accordingly, sign language attributes are initiated according to their importance while addressing a signed message throughout

computer mediated communication. In addition, user's preferences are considered in order to deliver customized signage content in term of display quality editing, with minimum data transmission rates in conjunction with other web objects that can be interact with content and style. MPEG-7 standard provides an appropriate markup environment to annotate various related content depicted by the intermediary signage object. To extensively broaden the accessibility issue in the web, this project looks forward for deeper examination of signage image and video content such as automatic facial expressions extraction and analysis rather than embed them manually.

## References

1. Agrafiotis, D. and Canagarajah, N., Perceptually Optimized Aign Language Video Coding based on Eye Tracking Analysis. *Electronics letters*, 2003. **39**.
2. Aouf, R. and Hansen, S. Integration of Signage Information into the Web Environment. in Proceedings of the 9th International Conference on Computer Helping People with Special Needs, ICCHP. 2004. Paris, France: Springer-Verlag GmbH.
3. Aouf, R., Hansen, S., and Salter, G. Incorporation of Signage Information with Multimedia Data Processing by Making Use of XML Technology. in Proceedings of the International Conference on Computers in Education, ICCE. 2004. Melbourn, Australia.
4. Brewer, J., Scenarios of People with Disabilities Using the Web. 2000.
5. Brewer, J., WebAccessibility Initiative WAI. 2003.
6. Elliott, R., Glauert, J., Kennaway, J., and Marshall, I. The Development of Language Processing Support for the ViSiCAST Project. in Proceedings of the Fourth International ACM Conference on Assistive Technologies. 2000. Virginia, USA.
7. Foulds, R., A., Biomechanical and Perceptual Constraints on the Bandwidth Requirements of Sign Language. *IEEE*, 2004. **12**: p. 65-72.
8. Han, W., Buttler, D., and Pu, C., Wrapping Web data into XML. *SIGMOD*, 2001. **30**.
9. Hunter, J., An Overview of the MPEG-7 Descripiton Definition Language (DDL), in *IEEE Transactions on Circuits and Systems for Video Technology*. 2001. p. 765-772.
10. Manoranjan, M., D., Practical Low-Cost Visual Communication Using Binary Images for Deaf Sign Language. *IEEE*, 2000. **8**: p. 81-88.
11. Martinez, J., M., Overview of the MPEG-7 Standard V5.0. 2001, ISO/IEC JTC1/SC29/WG11 N4031 Coding of Moving Pictures and Audio: Singapore. p. 1-88.
12. Pereira, F., MPEG-7 Application Document. 1998, ISO/IEC JTC1/SC29/WG11 MPEG98/N2462 Coding of Moving Pictures and Audio: Atlantic City. p. 1-25.
13. Pereira, F., MPEG-7 Requirements Document V.15. 2001, ISO/IEC JTC1/SC29/WG11 N4317 Coding of Moving Pictures and Audio: Sydney. p. 1-30.
14. Salembier, P. and Smith, J., R., MPEG-7 Multimedia Description Schemes, in *IEEE Transactions on Circuits and Systems for Video Technology*. 2001. p. 748-759.
15. Sandini, G. and Nielsen, J., Image-based Personal Communication Using an Innvative Space-Variant CMOS Sensor. 1997.
16. University, D., The DePaul University American Sign Language Project. 2002, School of Computer Science, Telecommunication and Information Systems: Chicago.
17. University, D., The DePaul University American Sign Language Project Overview. 2002, School of Computer Science, Telecommunication and Information Systems: Chicago.
18. University, D.C., Research Studentships in the School of Computer. 2003, School of Computing.
19. University, G., Gallaudet University: E-Learning Online Graduate Degrees. 2004.