Video Telephony for the Deaf: Analysis and Development of an Optimised Video Compression Product

Laura J Muir
Lecturer and PhD Student
The Robert Gordon University
Schoolhill, Aberdeen, UK.
+44 1224 262000
l.muir@rgu.ac.uk

Iain E G Richardson
Lecturer and PhD Supervisor
The Robert Gordon University
Schoolhill, Aberdeen, UK.
+44 1224 262000
i.g.richardson@rgu.ac.uk

ABSTRACT
The multimedia capability of video telephony and video conferencing systems has many applications and benefits. This paper describes research and development that aims to optimise video compression systems for a specific application - personal communication at a distance for deaf people. Results of eye movement tracking experiments and proposals for image content prioritisation based on these results are presented. The requirement for an appropriate quality assessment methodology is also addressed.

Categories and Subject Descriptors
H.4.3 [**Communications Applications**]: – *computer conferencing, teleconferencing and video conferencing.*

General Terms
Algorithms, Design, Human Factors.

Keywords
Eye movement tracking, sign language, video conferencing, video telephony, image content prioritisation.

1. INTRODUCTION
Video communication has the potential to change the way deaf people communicate since there are currently limited opportunities for direct personal communication at a distance. Deaf people use a rich combination of visual communication methods. These include sign language (hand gestures, facial expressions, body movements and eye movements), lip reading and finger spelling (where individual letters are signed, usually at rapid speed). Video communication of sign language and lip-reading and finger spelling relies on quality real-time transmission, which is currently unreliable and jerky. Studies have shown that accurate sign language communication places specific demands on a visual media application [1].

The aim of this project is to research and develop an optimised video compression solution, for video conferencing systems, that will enable deaf people to communicate at a distance using sign language.

Preliminary research by a MSc student and supervisor, Dr Iain Richardson, at The Robert Gordon University [2] identified the scope for optimised image quality and frame-rate for sign-language users. This research also established useful and interested contacts in the deaf community and the umbrella organisations that support it. Lilian Lawson, Director of the Scottish Council on Deafness is an advisor to the research project.

Of significant benefit to the research is previous work [3] to establish quality metrics for video communication of sign language. The technical requirements (or quality profile) are of value for benchmarking and evaluation in addition to the qualitative evaluations obtained from end-users (representatives of the deaf community).

2. IMAGE CONTENT PRIORITISATION AND VIDEO COMPRESSION OPTIMISATION
Video compression optimisation across the entire video scene is adequate for general-purpose applications, however the specific information carrying content of video for the target user cannot be optimised in this way. Optimising the entire video scene does not take account of the information carrying importance of different temporal and spatial components. It may be possible to make better use of available transmission bandwidth by selective optimisation of key features of the video sequence. The main research objectives are: to identify important regions of sign language video sequence images for the target user, develop a method of content prioritisation and develop and test the performance of the modified CODEC against standard CODECs.

2.1 Image Content for Sign Language Communication
Experimental work has been conducted, in this research project, to identify the most important regions of the video sequence images for sign language communication. Gaze point data was obtained from eye movement tracking experiments. These were conducted using specialist equipment that tracks the movement of the eyes using corneal and pupil reflection techniques. Profoundly deaf (L1) volunteers from the Aberdeen and North East Deaf Society took part in experiments, which recorded eye movements while watching a sign language video sequence on a PC monitor. A PC-based eye-tracker was used to record, play back and save eye gaze data. The results, illustrated in Figure 1, indicate that viewers focus mainly on the mouth with some vertical excursions. The main point of attention is the face. Hand movements are observed in peripheral vision referenced by occasional glances. Figure 2 shows vertical excursions recorded

while watching the sign language clip. The vertical excursions are typically very short (less than 100ms) which means that the viewer may not reach full spatial resolution in that time period.

The results are of value to the research as they indicate which regions of the image should be sharpest for the user. This has implications for future work on prioritisation. This early experimental work indicates that the background is not important and that there would be value in reducing the bit rate of peripheral regions and giving priority to the central position.

2.2 Options for Image Prioritisation and Segmentation

Segmenting the video scene and employing object-based coding to selectively encode the most important regions would result in efficient use of bandwidth and enhanced quality of important video content for sign language video transmission. There are two main options for image content prioritisation - approximate segmentation and accurate segmentation. Approximate segmentation is simplistic segmentation along block boundaries and so there are no pixel boundary concerns. Accurate segmentation allows coding of arbitrary shaped objects at the expense of increased computational complexity. Previous work in this area has been done based on skin colour, movement and a combination of these [4]. Accurate automatic segmentation of video objects would present a considerable challenge to the research project.

An alternative to segmentation is foveated imaging. This involves progressively reducing image resolution away from one or more focal points.

Future work will investigate and explore automatic segmentation and the use of foveated imaging for image optimisation in this project.

2.3 Proposals for Video Compression Optimisation and Testing

International video compression standards, being investigated for optimisation, are MPEG-4, H.263 and the new standard H.264. The standards describe mechanisms for coding, multiplexing and presenting video images at low bit rates. The standards also describe tool sets (profiles) for coding video images to enable efficient storage, transmission over networks and viewing/manipulation by end-users. Future work will involve the development and testing of modified CODECs that give priority to important image content for sign-language users. The optimised CODECs will be tested to determine whether they out perform the standard CODECs.

In addition, there will be a requirement to develop a suitable method of comparing quality. Conventional quantitative comparisons, for example PSNR (Peak Signal to Noise Ratio) output, is of no value to this research. This type of quality assessment applies to quality across the entire video scene. Qualitative measurements are of more value as they indicate the quality perceptions of the user. Conventional qualitative methods [5], using quality scales for subjective assessment, are not appropriate, as they provide no real indication of the ease and quality of sign-language communication. A suitable quality measurement scheme will be developed to test the video output for different sign-language sequences. These will facilitate end-user acceptance testing and comparison with output from existing CODECs for the developed system. A method, which assesses reliability and ease of sign language communication, is the desired result rather than one that assesses optimal quality across the whole video scene.

3. CONCLUSION

This paper presents current research to identify important regions of sign language video sequence images for the target user. This will inform future work on image content prioritisation. Future work is planned to develop a suitable method of automatic image segmentation, develop and test the performance of the modified CODEC against standard CODECs and devise an appropriate method of quality assessment by the target user.

Provision of an optimised video compression solution that out-performs general standard based CODECs is the desired outcome of the research. This will benefit the deaf community and allow personal communication at a distance and contribute to the body of knowledge in the area of video compression.

4. ACKNOWLEDGEMENTS

This work would not been possible without the help of Lilian Lawson (Director of the Scottish Council on Deafness), Jim Hunter and other members of the Aberdeen & North East Deaf Society. The work of Steven Leaper, MSc student at The Robert Gordon University, is gratefully acknowledged as contributing to this paper.

5. REFERENCES

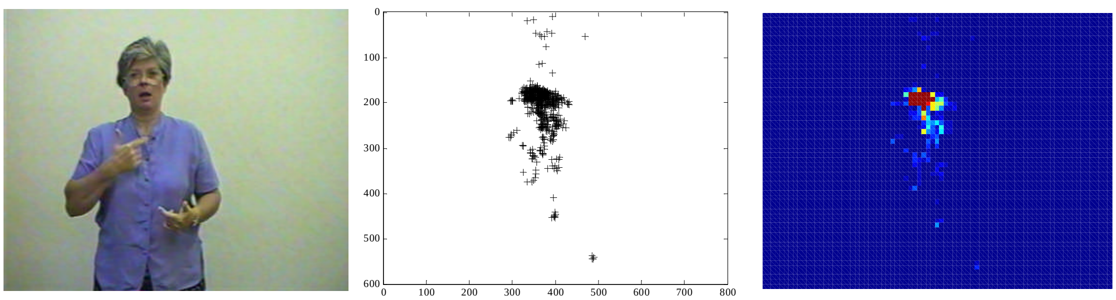
[1] ITU-T Q9/16, Draft Application Profile, Sign language and lip-reading real time conversation usage of low bit rate video communication, September 1998.

[2] VETTER, S.L., August 2000. Video Communications for Deaf People, MSc Thesis, School of Electrical and Electronic Engineering, The Robert Gordon University.

[3] HELSTROM,G., Delvert, J., November 1996. Quality Measurement for Video Communication of Sign Language. [www] http://www.omnitor.se/textversion/english/qualityonvideo.html.

[4] HABILI, N., July 2001. Automatic Segmentation of the Face and Hands in Sign Language Video Sequences, Department of Electrical and Electronic Engineering, Adelaide University.

[5] ITU, 2000. Methodology for the subjective assessment of the quality of television pictures, Recommendation ITU-T BT.500-10.



(a) Image *(b) Gaze Points* *(c) Gaze Point Densities*

Figure 1: Eye Movements

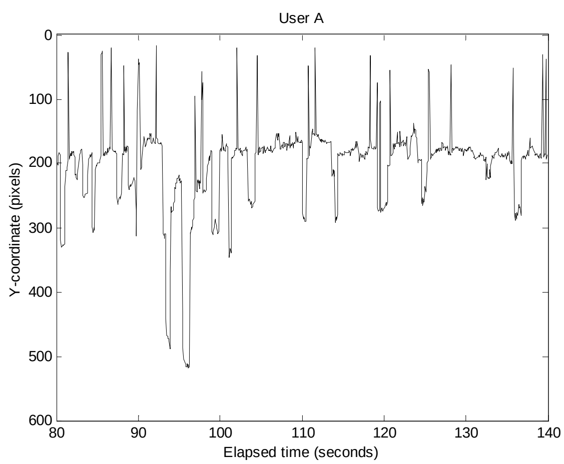


Figure 2: Vertical Excursions