Improving fluency in sign language to text systems

Sam Black
524689

Supervisor
Prof Martin Russell

Second Report

Table of Contents

Introduction

Communication via sign language is more universal than the spoken word; we all understand what is meant by someone pointing, covering our ears over with our hands, or motioning somewhere.

Automatic sign language recognition (ASLR) systems lag some way behind other recognition systems for the simple reason that gesture permutations are nearly infinite; ask someone to move their hand from above their head to by their waist and you'll get variations every time. Capturing the motion of the person's arm in an non-intrusive way also limits the usefulness of a system, with the least intrusive systems being the most complex to implement.

To account for the first problem, a statistical model of the various motions in the form of a Hidden Markov Model (HMM) is used; this model maps the most likely transition of the person's arms signifying a word, and because it works using probabilities, it can be designed to allow for minor variations between each arm movement, user or application. The second problem is more difficult to overcome, and will be the subject of many other research projects and advances in technology.

Whilst research in ASLR has been conducted, previous systems to convert sign language to text have concentrated on translating individual letters rather than using whole words; whilst this limits the corpus needed, it is uncomfortable and inconvenient to spell each word rather than just sign it (imagine phonetically spelling each word). Most systems currently cannot translate sign language to text in real time fluently, requiring the user to sign each letter or word discretely, reducing the usefulness of the system in a real world setting.

Thus, the aim of the system is to use HMM to implement a fluent British sign language (BSL) to text translator for a limited corpus, such as an information help point for the University campus (for example, “Where is the Guild?”).

Research

Automatic sign language recognition (ASLR) is a combination of gesture recognition and facial recognition (Edwards, 1997). Various methods exist to capture the data, most systems utilise an image capture system that extracts the vector data. The unobtrusive nature of this method is the major factor to its wider adoption and basis of research. Image capture is less accurate than using a motion tracking system, such as using a data glove or optical motion tracking system, as there is a lower amount of noise and segmentation of the data (Dreuw, Rybach, Deselaers, *et al*, 2007).

Accuracy in the system is approaching that of automatic speech recognition (ASR) systems of 85% - 90% (Holt, Hendriks, and Andringa, pp. 8), but lower word error rates (WER) can be achieved by using multiple layers of HMMs. Layered HMM are best for systems which require low response times with a large corpus (Zhang, Yao, Jiang, *et al*, 2005).

Whilst other systems are relatively successful with identifying finger spelt words (Travieso, Alonso, and Ferrer, 2003), discrete isolated words signed (Grobel and Assan, 1997) with no attempt at grammatically correct sentence structure (Akmeliawati, Ooi and Kuang, 2007), few are successful in fluent sentence recognition with a grammar. Computer processed visual data is still susceptible to noise and interference, such as the background being of similar colour to the users hands, adverse lighting conditions or other people in motion providing spurious inputs (Je, Kim and Kim, 2007). Computer vision based systems offer the most comprehensive coverage of a sign language user's movements, but are computationally expensive to calculate the vectors off the user's body and limbs (Holt, Hendriks, and Andringa, pp 11).

Outcomes

To construct a fluent, real-time ASLR system based on HMM, using a motion tracker to collect the sign data for PCA. As an extension, enhance the system to work with a camera or webcam to capture data.

Requirements and Progress to date

The requirements to achieve the outcomes and their current status;

1. Create the corpus to be signed; Complete
2. Create grammar, models and pronunciation dictionary for HTK; In progress
3. Record the signing using a motion tracker; Complete
4. Compile the vector data from the motion tracker; In progress
5. Conduct principle component analysis (PCA) on the vectors; PCA software written
6. Run training data through HTK; HTK Python wrappers written
7. Improve the grammar, models and pronunciation dictionary; Incomplete
8. Re-run the training data to test improvements; Incomplete

Plans

Priorities

The immediate priorities are to annotate, check and separate the collected vector data and to process this using PCA over the Christmas break.

As shown in the gantt chart in Appendix 1, next semester will be devoted primarily to training and testing the HMM, improving the grammar, language models and pronunciation dictionary.

Improving the codebase for real-time use, and implementing OpenCV for image capture to fulfil the extension requirements are coding priorities for next semester.

Costing

The project has cost £50 to date in hiring two BSL signers to record the vector data. Whilst no further expenditure is expected, the BSL signers may need to be rehired to redo bad data, missing information or other unforeseen circumstance.

References

Akmeliawati, R. Ooi, M. P. L. and Kuang, Y. C. (2007). **Real-Time Malaysian Sign Language Translation using Colour Segmentation and Neural Network.** Proceedings of Instrumentation and Measurement Technology Conference, Warsaw, Poland, 1-3 May 2007.

Dreuw, P. Rybach, D. Deselaers, *D. et al.* (2007). **Speech Recognition Techniques for a Sign Language Recognition System.** In Interspeech, pages 2513-2516, Antwerp, Belgium, August 2007.

Edwards, A. D. N. (1997). If the glove fits: Progress in sign language recognition. **Ability**, 22: pp. 12-13

Grobel, K. and Assan, M. (1997) **Isolated Sign Language Recognition using Hidden Markov Models.** Systems, Man, and Cybernetics, 'Computational Cybernetics and Simulation'. IEEE International Conference 12-15 Oct. 1997. Volume 1, pp. 162-167

t. Holt, G. Hendriks, P. and Andringa T. (unknown date). **Why don't you see what I mean?** [online]. http://ict.ewi.tudelft.nl/pub/gineke/ASR05.pdf [Accessed 07 October 2008]

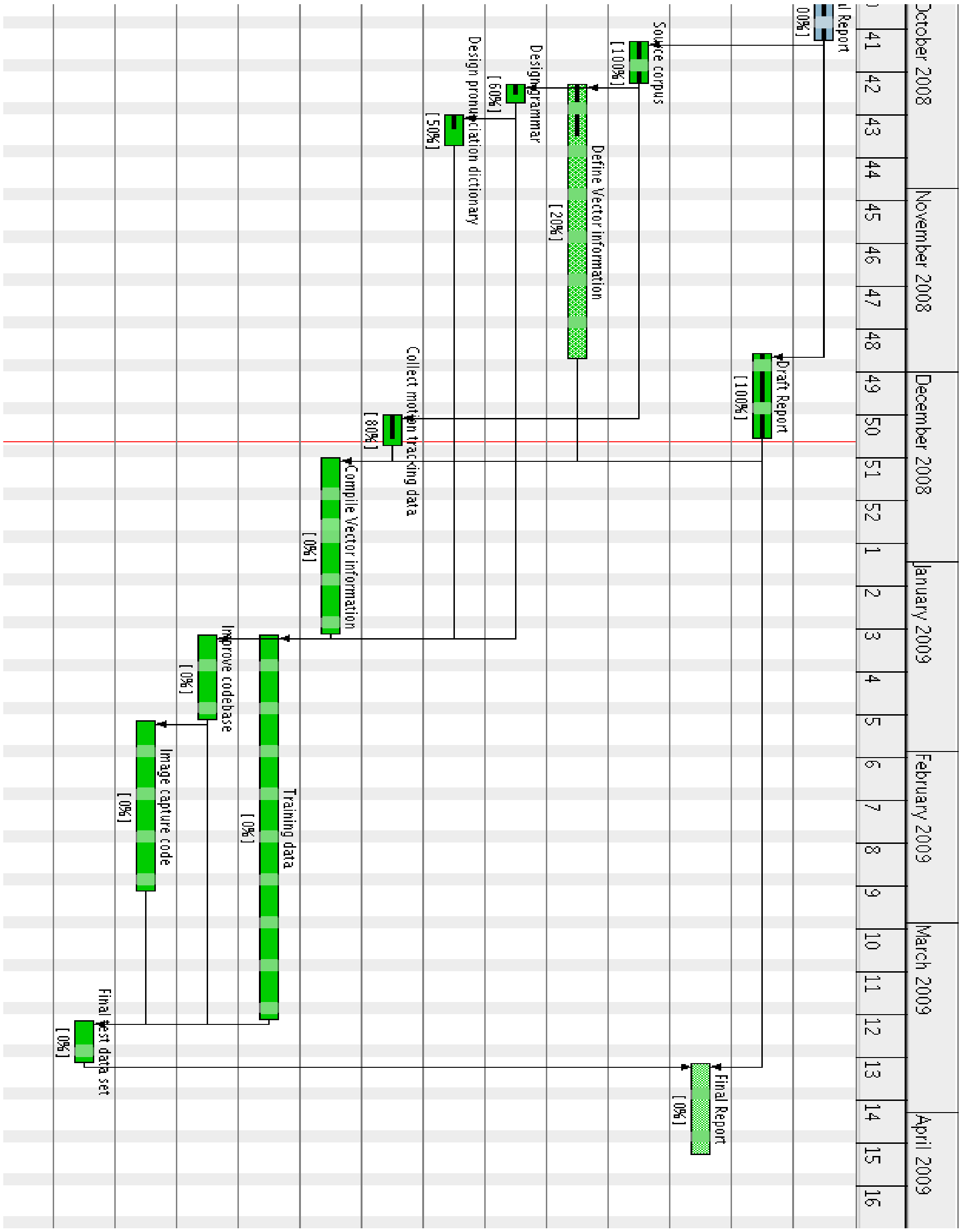
Je, H. Kim, J. and Kim, D. (2007). **Vision-Based Hand Gesture Recognition for Understanding Musical Time Pattern and Tempo.** In Industrial Electronics Society, 33rd Annual Conference of the IEEE, 5-8 Nov. 2007, pp. 2371-2376

Travieso, C.M. Alonso, J.B. and Ferrer, M.A. (2003). **Sign language to text by SVM.** Proceedings of the Signal Processing and Its Applications, Seventh International Symposium 1-4 July 2003, Volume 2, pp. 435-438

Zhang, C. Yao, H. Jiang, F. *et al* (2005). **Multilayer Method Based On Multi-Resolution Feature Extracting and MVC Dimension Reducing Method for Sign Language Recognition.** Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21 August 2005

Appendices

Appendix 1: Gantt Chart



Appendix 2: HTK Model

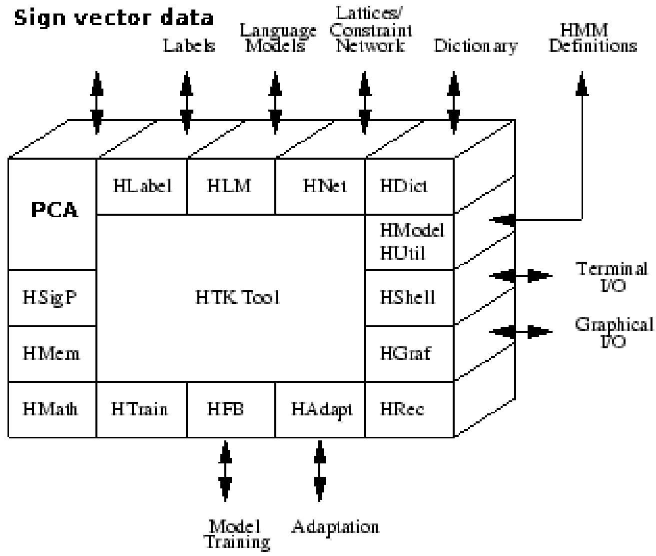


Figure 1: HTK tool software design, with speech analysis replaced with sign vector data and PCA. (Based on the HTK Tutorial, pp. 24).

Appendix 3: Access to source code

The HTK source code, research, reports, language models, grammars, dictionaries and scripts are accessible using source code management software package GIT (http://git.or.cz/) by;

```
git clone http://repos.lapwing.homelinux.org/personal-uni.git
```

and can be update by

```
git pull
```

in the directory **personal-uni/**.

Source code for this project is in the **personal-uni/ee4p/** directory.

Redmine (http://redmine.org/) is used to keep track of issues, work to do and a simple wiki. This is available from http://samwwwblack.lapwing.org/ .