An Integrated Sign Language Recognition System

Warren Nel[1], Mehrdad Ghaziasgar[1] and James Connan[2]
Department of Computer Science
University of the Western Cape[1], Private Bag X17 Bellville, 7535, South Africa
Tel: + (27) 21 959-3010, Fax: + (27) 21 959-3006
and Department of Computer Science
Rhodes University[2], P O Box 94 Grahamstown, 6140, South Africa
Tel: + (27) 46 603-8291, Fax: + (27) 46 636-1915
email: {2806381, mghaziasgar}@uwc.ac.za[1], J.Connan@ru.ac.za[2]

Abstract—Five fundamental parameters uniquely characterize sign language gestures. These are hand shape, location, orientation, and motion, as well as facial expressions. The South African Sign Language group at the University of the Western Cape has created several independent systems that recognize each of these parameters individually. This research aims to make use of parallel processing techniques to integrate the existing systems into a single sign language gesture recognition system using the Graphics Processing Unit (GPU) Application Programming Interface (API), called Compute Unified Device Architecture (CUDA).

Index Terms—South African Sign Language, Recognition, Integration, Parallel Programming

I. Introduction

Sign languages are languages that make use of manual, facial, and other body movements as a means of communication. They are the primary means of communication amongst many deaf people. Contrary to common belief, sign languages are not the signed equivalent of spoken languages [1]. It is, therefore, little known that many deaf people are completely unable to understand spoken languages of any form. A communication barrier exists between the hearing and deaf.

The South African Sign Language (SASL) research group at the University of the Western Cape (UWC) has made significant contributions towards the innovation in technologies to aid translation between SASL and English. The group is in the process of developing a machine translation system that can automatically translate phrases between SASL and English.

Research has shown that there are five fundamental parameters that uniquely characterize any sign language gesture [1]. These are: hand shape, location, orientation and motion, as well as facial expressions. The recognition of SASL phrases from video input such as from a mobile phone camera requires the recognition of all five parameters from the video input. A variety of projects within the SASL group have successfully achieved the recognition of four of the five parameters, namely hand shape [2], location [3] and motion [4], and facial expressions [5]. These systems are currently independent.

The challenge remains to create a system that integrates existing components to produce a coherent output that fully describes the sign language gestures in terms of these parameters. This can be used to derive the meaning of the sign language phrases and carry out a translation into English.

This research proposes an integrated system that incorporates the methodologies of the existing recognition systems within the group into a single recognition unit. A recent study carried out within the group showed that the use of parallel processing techniques on the GPU can significantly speed up the processing time of image processing systems [6]. The proposed system will use parallel processing to achieve real-time speeds.

The rest of this paper is organised as follows: Section II discusses the existing recognition systems developed within the group; Section III discusses the goals and objectives of this research; this paper is then concluded in Section IV

II. Related Work

This section describes the existing recognition systems developed at the SASL research group which are to be integrated into a single system. The sections that follow describe the existing systems that carry out recognition of hand location, hand shape and facial expressions, respectively. These are the systems that will be integrated.

A. *Acnmed's Hand Location Recognition System*

Achmed developed a system that determines the location of a signer's hands [3]. He conducted a comparison between an example-based and a novel learning-based approach to the problem. He found that the learning-based approach performed at a much higher level of accuracy than the example-based approach. Only the learning-based approach is considered in this research.

Achmed's learning-based system makes use of face detection to determine the presence of a human and to normalize the location of the person in the video. Skin detection is applied to eliminate non-skin pixels. Frame differencing is applied to the skin images to arrive at only skin pixels that have moved. Morphological operations are applied to eliminate small isolated skin regions and enhance large skin regions. The resulting image contains regions of moving skin in the frame.

A Support Vector Machine (SVM) is trained on all such images to be able to determine the position of the hands. The system was trained and tested on videos of six different signers each performing 20 SASL signs once. The system achieved an accuracy of 81% across all signs.

B. *Li's Hand Shape Recognition System*

Li developed a system that recognizes the hand shapes of a signer [2]. The system is able to accurately track the hand of the signer and recognize the hand shape performed. To track the hand, Li makes use of CAMShift tracking. Hierarchical Chamfer matching is used to locate the hand, only once, to initialize the CAMShift tracking window. This significantly enhances the tracking speed.

The skin segmentation method used is the same as Achmed and frame differencing is similarly applied to obtain moving skin pixels inside the tracking window. Contour detection on these pixels is used to locate the hand within the tracking window. A minimum bounding box is used to segment the hand contours from the greater CAMShift tracking window. Normalization is carried out by rotating and aligning the principal axis of the bounding box to either the x- or y-axis of the image. The resulting image is used as a feature vector. An SVM was trained to recognize various hand shapes.

The system was trained to recognize 10 SASL hand shapes. The system was trained on 40 example images for each hand shape.

The system was tested on a simple and complex background on six test subjects. Each subject was asked to perform each of the 10 hand shapes in succession with instructions to hold each hand shape before a transition. This yielded six videos of the 10 hand shapes on a simple background and six on a complex background. Overall, the system was found to achieve a recognition accuracy of 78.5% across all signers, hand shapes and both backgrounds.

C. Whitehill's Facial Expression Recognition System

Whitehill developed a robust facial expression recognition system [5]. The system uses the Facial Action Coding System (FACS), a system that describes facial expressions in terms of motions of 44 muscle groups in the face also referred to as action units (AUs) [7]. It recognizes various facial expressions as combinations of AUs.

Images read in are first scanned to detect the location of the face. The facial frame is normalized by setting its width and height to a set value. Square regions of size 24 pixels around the eyebrows and mouth were cropped out of the image. Each cropped region was converted into a Gabor representation using a bank of 40 Gabor filters. Feature vectors were determined as the complex magnitude of the Gabor jets. A series of SVM classifiers were trained, each classifiers detecting the presence of exactly one AU, regardless of whether or not they were combined with other AUs. Only 27 AUs were considered.

Ten-fold cross-validation using a total of 580 images of 76 human subjects performing six human facial expressions was used to train and test the performance of the system. Testing aimed to determine the accuracy with which the system could correctly determine the AUs associated with a test image of a facial expression. The facial expressions were: anger, disgust, surprise, joy, fear, sadness. The system was found to achieve an overall recognition accuracy of 91.41% across all AUs.

III. Research objectives

This research aims to combine the aforementioned SASL projects into a single robust sign language translation system. The discussion of the methodologies brings to light the fact that these systems have many components in common. For example: they all make use of face detection and Li and Achmed make use of the same background subtraction method.

The implementation of all three systems as one unit entails identifying all such common components and implementing these components in such a way as to eliminate redundancy to enhance processing speed. The combined methodology will be a new methodology in its own right.

The combined system will be implemented using parallel processing techniques with an aim to achieve real-time processing speeds.

Finally, the system will be required to produce output in a standard sign language format, specifically Sign Writing Markup Language (SWML). The resulting output describes the sequence of gestures in a video in terms of the changes in each of the recognized parameters. This can subsequently be translated into English text.

IV. Conclusion

This paper proposed a sign language gesture recognition system that will be using a parallel processing approach. This approach will make use of the GPU API called CUDA. Upon the success of this project, the SASL group of UWC will have a satisfactory sign language gesture recognition system which will also serve as a significant milestone for the group.

References

[1] W.C. Stokoe. “Sign language structure: an outline of the visual communication systems of the American deaf. 1960”. *J. Deaf Stud. Deaf Educ.*, 10(1):3–37, 2005.

[2] P. Li. "Hand Shape Estimation for South African Sign Language". Master's thesis, University of the Western Cape, 2012.

[3] I. Achmed. "Upper Body Pose Estimation Towards the Translation of South African Sign Language". In Proc. *Southern Africa Telecommunication Networks and Applications Conference*, pp. 363–366, 2009.

[4] C. Rajah. "Chereme-Based Recognition of Isolated, Dynamic Gestures from South African Sign Language with Hidden Markov Models". Master's thesis, University of the Western Cape, 2006.

[5] J. Whitehill. "Automatic real-time facial expression recognition for signed language translation". Master's thesis, University of the Western Cape, 2006.

[6] D. Brown, M. Ghaziasgar and J. Connan. "Faster Upper Body Pose Estimation Using CUDA". In Proc. *Southern Africa Telecommunication Networks and Applications Conference*, 2011.

[7] P. Ekman and W.V. Friesen. “Facial Action Coding System”. *Consulting Psychologists Press*, Stanford University, 1977.

Warren Nel is currently a Telkom Centre of Excellence M.Sc. student at the University of the Western Cape. He is currently doing research on sign language synthesis and novel communication applications for the Deaf and hearing impaired.

Mehrdad Ghaziasgar is the project manager of the South African Sign Language (SASL) research group. His interests are internet programming, mobile computing, computer vision and machine learning.

James Connan heads the South African Sign Language (SASL) research group. He has a wide range of interests that include: computer vision and machine learning.