Addressing the Problem of Hand Occlusion in Bimanual Hand Shape Recognition

Ibraheem Frieslaar, Mehrdad Ghaziasgar[1], 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
Tel: + (27) 46 603-8291, Fax: + (27) 46 636-1915
email: {ifrieslaar, mghaziasgar}@uwc.ac.za[1]; j.connan@ru.ac.za[2]

Abstract – Hand shape recognition is one of the fundamental features that characterise any sign language gesture. This plays an important role in all systems that attempt to interpret sign language. The SASL group has produced a state-of-the-art hand shape recognition and estimation system for South African Sign Language. It is capable of recognising and estimating hand shapes with an accuracy of 81% and 86% respectively. The system, however, only tracks the right hand. This research intends to extend to this system to track both hands simultaneously and implement hand-hand occlusion and collision detection and resolution. Hand-hand occlusion or collision prevents hand shape recognition systems from correctly tracking and distinguishing the left and right hands.

Index Terms— Collision, Hand Gestures, Hand Shape, Occlusion

I. Introduction

Hand gestures are used by people to explain speech, reinforce speech and express their internal emotions. Hand gestures are commonly used in sign language.

Hand shape recognition is one of the fundamental features that characterise any sign language gesture [1]. The other important features are: hand orientation; hand motion; hand location; and facial expressions.

Hand shape recognition plays an important role in all systems that attempt to interpret sign language. The South Africa Sign Language (SASL) Machine Translation system under development by the SASL group at the University of the Western Cape is an example of such a system [2].

The SASL group has produced a state-of-the-art hand shape recognition and estimation system for South African Sign Language [1]. The system is capable of recognising and estimating hand shapes using video-based input without the use of special equipment such as data gloves [3].

The system uses a web camera as input and image processing techniques to locate and track the hands. The hand shape is sent to a Support Vector Machine (SVM) for classification. The system has achieved an overall accuracy of 81% and 86% for recognition and estimation, respectively. However, there are drawbacks to the system. It is unable to detect and track both hands simultaneously. The problem of hand-hand occlusion and distinguishing the left hand from the right hand after occlusion has occurred has not been done.

This Work-in-Progress paper discusses the prospect of improving the current system and adding additional features such as: locating and tracking both hands simultaneously; solving the hand-hand occlusion problem; and recognising more hand shapes.

This work will ultimately be integrated into previous work, upper body pose recognition and estimation which was done by the SASL group. [4] [5].

This paper is organised as follows: Section II discusses related work in the field; Section III discusses the research goals; the paper is concluded in Section IV.

II. Related Work

This section discusses the Hand Shape Recognition and Estimation for the South African Sign Language system developed by Lei *et al*, which will be improved in this research. The second section focuses on Shamaie *et al* proposed solution to overcome hand-hand occlusion.

A. Hand Shape Recognition and Estimation for South African Sign Language

Li et al. created a novel system which recognises sign language hand shapes from 2D video input. The novelty of their system is the fact that it does not require a hand pose to be recognised at every frame [1].

Their system uses Haar Classifiers to determine the position of the face. The centre of the facial frame is usually situated on the nose. The skin colour distribution is extracted from the nose. This skin colour distribution is used to identify all skin pixels of an individual. Back projection is applied to these pixels in order to determine the hand and eliminate noise. However, objects in the frame may have the same colour as the skin pixels which are regarded as noise. This type of noise is eliminated by using Gaussian Mixture Models (GMM) background subtraction. This separates the background and foreground, thus only the hands will be present in the image.

Hierarchical Chamfer matching is used to locate the hands. It is detected by the different scale sizes of the hand. Their system uses the face as a reference to determine the optimal scale size of the hand should be. Once the hand has been located it is set as the Region of Interest (ROI). Hand detection is stopped and the ROI is tracked using CAMShift. The skin pixels of the face are removed to prevent interference with the CAMShift tracking.

To gain distinct features from the different hand shapes, the hand contour is used as the feature vector. To extract the contours of all the blobs a chain code based algorithm is used. Thus, the contour with the largest amount of pixels is the hand. The hand image is normalised by rotating it, until it is on the same axis as the image axis. This solves the problem of misalignment invariance. The image is scaled down to a resolution of 20 X 30 pixels and sent to a SVM.

Their training data consisted of a total of 400 images. 10 videos were recorded with each video having a different hand shape.

The testing procedure consisted of five subjects with different skin colour. Each subject was asked to perform a transition between the 10 hand shapes 15 times in a sequential order.

The recognition accuracy was determined by comparing each hand shape input to the recognised result. The estimation accuracy was determined by the hand shape input and compared to the resulting 3D model.

Their system achieved a recognition and estimation accuracy of 81% and 86%. However, there are drawbacks to the system. It is unable to detect and track both hands simultaneously as is required by SASL. The problem of hand-hand occlusion has not been addressed.

B. Hand tracking in bimanual movements

Shamaie *et al* proposed a method for tracking hands in bimanual movements [6]. Their algorithm consists of tracking both hands and correctly distinguishing the left and right hands after occlusion occurs.

They make use of the Grassfire algorithm to extract the blobs in the image. These blobs are the hands. In an image with both hands present, the algorithm would detect two blobs. When hand-hand occlusion takes place only one blob is detected.

Once the hands have been located in the image, a box is drawn around it. Kalman filters are used to track these boxes containing the hands. They use Kalman filters to determine the velocity and acceleration of each side of the boxes. Having each hand's position, velocity and acceleration they are able to determine when the hands will occlude or collide with each other.

When the algorithm detects occlusion it locates only one blob and its velocity is zero, they prove that both hands have paused or have collided with each other. Once collision or occlusion has occurred they assume that the hands will move back to its original position. If occlusion has occurred and there was no pause they conclude that the hands as crossed each other.

Their results consisted of three different sets. Each set consisted of 1600 movements. Set one was to determine the performance of the system with different velocity of the hands. Moving the hands at fast pace resulted in 88.13% accuracy and moving the hands at a slow pace recorded 95.13% accuracy. In the second set they used different video angles and still the hands were tracked and reacquired correctly after occlusion. The third set demonstrated that the algorithm tracks both hands perfectly even when one hand is out of the frame and returns to the frame.

III. RESEARCH GOAL

The aim of this research is to improve on Li's work on hand shape recognition and estimation for South African Sign Language [1]. The improvements will aim to achieve the following goals:

1. Extend the implementation to detect and track both hands simultaneously.
2. Solve the hand-hand occlusion problems according to Shamaie *et al's* implementation :
   a) Reacquire and distinguish between the left and right hands after the hands have moved towards each other and/or pass each other.
   b) Reacquire and distinguish between the left and right hands after the hands pauses at the point of occlusion and continues in the original direction of motion.
   c) Reacquire and distinguish between the left and right hands after both hands pause at the point of occlusion and both hands returns to their original positions.
3. Recognise and estimate more hand shapes than the previous system.

IV. CONCLUSION

This paper proposes an improvement to the existing hand shape recognition and estimation system of the SASL group. Improvements include: locating and tracking both hands simultaneously; reacquiring and distinguishing between the left and right hands after the hands have moved towards each other and/or pass each other; and recognising and estimating more hand shapes than the previous system

V. REFERENCES:

[1] P. Li, M. Ghaziasgar and J. Connan, "Hand Shape Recognition and Estimation for South African Sign Language", Cape Town, 2011.

[2] J. Connan. Integration of Signed and Verbal Communication: South African sign language Recognition and Animation. [Online]. http://www.coe.uwc.ac.za/index.php/SASL.html

[3] A. Erol et al, "Vision-Based Hand Pose Estimation: A Review", *Computer Vision and Image Understanding*, vol. 108, no. 1-2, pp. 52-73, 2007.

[4] N. L. Naidoo, and J. Connan, "Gesture Recognition Using Feature vectors", *Proceedings of the Southern Africa Telecommunication Networks and Applications Conference*, *SATNAC*, 2008.

[5] I. Achmed and J. Connan, "Upper Body Pose Estimation Towards the Translation of South African Sign Language", *Proceedings of the Southern Africa Telecommunication Networks and Applications Conference*, *SATNAC*, pp. 427–432, 2010.

[6] A. Shamaie and A. Sutherland, "Hand Tracking in Bimanul Movements", *Image and Vision Computing*, vol. 23, no. 13, pp. 1131-1149, 2005.

Ibraheem Frieslaar is an M.Sc student at the University of the Western Cape Department of Computer Science. He is currently doing research on sign language synthesis and novel communication applications on mobile interfaces 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 up the South African Sign Language (SASL) research group. His interests are computer vision and machine learning.