Kinect Depth Stream Pre-processing For Hand Gesture Recognition

Muhammad Asad and Charith Abhayaratne,
Department of Electronic and Electrical Engineering,
The University of Sheffield, United Kingdom.

1. Introduction

Recent introduction of Microsoft Kinect sensor and its software have resulted in a trend of utilizing the real-time depth stream for robust and efficient hand gesture recognition methods.

However, the depth streams acquired from Kinect sensor suffer from:

- low resolution (640x480)
- random noise
- quantization errors

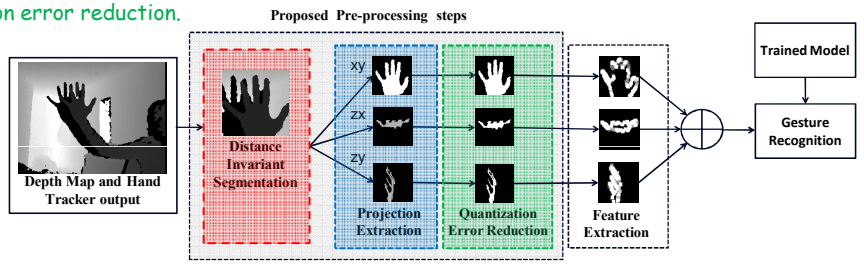
These limit the operating range and functionality of the most of the existing hand gesture recognition techniques.

In this work, we propose Kinect depth stream pre-processing techniques to overcome the quantization errors and assist in extraction of relevant features leading to robust hand gesture recognition.

2. System overview

The proposed Kinect depth stream pre-processing approach consists of three main steps, which are:

(i) Distance invariant segmentation;
(ii) Projection extraction and,
(iii) Quantization error reduction.



Flowchart showing the proposed pre-processing method within a hand gesture recognition system

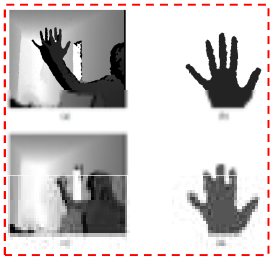
3. Methodology

1. Distance Invariant Segmentation

- Hand is tracked using OpenNI SDK tools
- If the hand region is an s x s square and the distance to the hand is p_z
- Then

$$s \propto \frac{1}{p_z} \rightarrow s = \frac{k}{p_z}$$

- k is a constant, experimentally found as k=108,000.
- Normalised hand region segmentation examples ....



- For Pz = 700mm, (a) depth stream I, (b) Segmented depth stream Is with size 156x156 pixels,
- and for Pz = 1700mm, (c) depth stream I, (d) depth stream Is with size 64x64 pixels

2. Projection Extraction

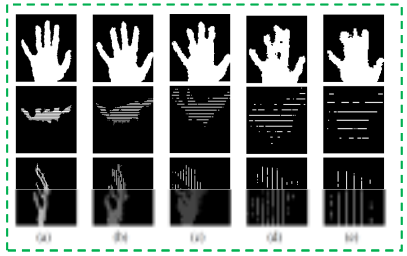
- Based on the work of Li et al. (IEEE CVPR 2010) and Okada (WSEAS 2002), 2 projections are extracted from the segmented hand region depth stream

(a) XY projection (front view), (b) ZX projection (top view), (c) ZY projection (side view).

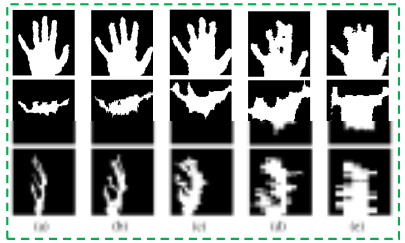
3. Quantization Error Reduction

- Pre-processing for quantization error reduction includes 2 steps
  1. Separable morphological closing using 1x3 and 3x1 structural elements.
  2. Adaptive linear interpolation (To fill in the discontinuities for p_z > 950mm. This is performed repeatedly for several steps. The number of steps is limited by the required complexity level.

- Examples of effect of pre-processing on projections from various distances are shown as follows:



WQR: Projections without quantization error reduction showing XY (top row), ZX (middle row) and ZY (bottom row) at Pz = (a) 700mm, (b) 950mm, (c) 1200mm, (d) 1450mm, (e) 1700mm



QR: Projections with the quantization error reduction step showing: XY (top row), ZX (middle row) and ZY (bottom row) at Pz = (a) 700mm, (b) 950mm, (c) 1200mm, (d) 1450mm, (e) 1700mm

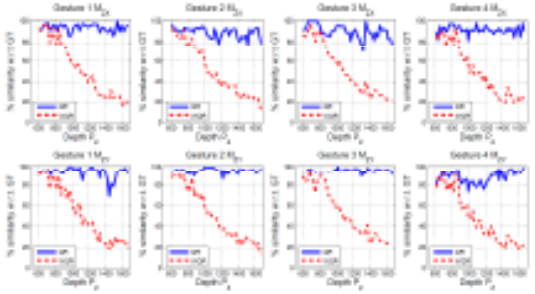
4. Validation results

- A dataset of horizontal swipe gesture was divided into four gesture stages and used for evaluation and validation of the proposed pre-processing methods.

Projections of four gesture stages used for evaluation showing XY (left), ZY (centre) and ZY (right) for stages of left-to-right swipe (a) stage 1, (b) stage 2, (c) stage 3, (d) stage 4.

-We captured a dataset containing 2000 samples for each of the four gesture stages, with p_z varying between 600mm and 1700mm

- The similarity between the projections and the ground truth (GT) were computed using computing the similarity percentage for using the quantization reduction step (QR) and without using the quantization reduction step (WQR)



- The hand gesture recognition accuracy rates for the two methods: using the quantization reduction step (QR) and without using the quantization reduction step (WQR)

Different gestures	QR	WQR
1	**98.60%**	97.95%
2	**99.25%**	99.05%
3	**83.20%**	77.85%
4	**88.15%**	86.70%

5. Conclusions

-A Kinect depth stream pre-processing method for Hand Gesture Recognition has been proposed in this work.

-It involves a distance invariant segmentation process along with a method to construct projections of hand in three different planes and pre-processing for quantization error reduction (QR) which includes morphological closing followed by adaptive linear interpolation.

- It results in consistent high similarity with the ground truth projection data resulting in distance invariant high performance leading to improvements in hand gesture recognition accuracy (improvement by 0.2% -5.35%)

Author Contact details:
Muhammad Asad is currently with City University London ,UK: (Email: muhammad.asad.2@city.ac.uk) Charith Abhayaratne : (Email: c.abhayaratne@sheffield.ac.uk web: http://svc.group.shef.ac.uk/shefhgr.html)