DETECTING HAND-PALM ORIENTATION AND HAND SHAPES FOR SIGN LANGUAGE GESTURE RECOGNITION USING 3D IMAGES

Lalit K. Phadtare[1], Raja S. Kushalnagar[2], Nathan D. Cahill[3]

[1]Department of Electrical Engineering, Rochester Institute of Technology, NY, USA
[2]Department of Information and Computing Sciences, Rochester Institute of Technology, NY, USA
[3]School of Mathematical Sciences, Rochester Institute of Technology, NY, USA

ABSTRACT

Automatic gesture recognition, specifically for the purpose of understanding sign language, can be an important aid in communicating with the deaf and hard-of-hearing. Recognition of sign languages requires understanding of various linguistic components such as palm orientation, hand shape, hand location and facial expression. We propose a method and system to estimate the palm orientation and the hand shape of a signer. Our system uses Microsoft Kinect to capture color and the depth images of a signer. It analyzes the depth data corresponding to the hand point region and fits plane to this data and defines the normal to this plane as the orientation of the palm. Then it uses 3-D shape context to determine the hand shape by comparing it to example shapes in the database. Palm orientation of the hand was found to be correct in varying poses. The shape context method for hand shape classification was found to identify 20 test hand shapes correctly and 10 shapes were matched to other but very similar shapes.

Index Terms— ASL gesture recognition, palm orientation, hand shape recognition, shape context, virtual reality

1. INTRODUCTION

Gesture recognition for understanding sign languages still is a complex problem because of the large number of signs and the variety of different features defining each sign. Many methods for sign language gesture recognition look at subsets of these parameters [1] focuses on the problem of successfully tracking the hands and head position of the signer and [2, 3] classifies the hand gestures based on 3D model analysis, motion analysis, neural networks etc. An even more difficult problem is recognition of signs from a continuous image stream, popularly known as Continuous Sign Language Recognition or Motion Epenthesis [4, 3].

As opposed to many systems that focus on specific sign languages like ASL [5], BSL [6] and others our system focuses on the linguistic features of sign language that are common to many languages. The basic linguistic features, also known as production parameters [7] are:

- Hand Shape: the shape or configuration of the hands necessary to denote any sign.
- Palm Orientation: the direction in which the palm is rotated or pointed.
- Movement: the global or specific movement of hands.
- Location: the physical regions near the body where the sign is produced.
- Non-manual markers: any features not made by hands including shoulders, head or facial expression etc.

(a) HamNoSys Avatar 'Anna'

```
<sigml>
<hamgestural_sign gloss="film">
<sign_manual>
<split_handconfig>
<handconfig handshape="flat" extfidir="u"
palmor="d"/>
<handconfig handshape="finger2"
thumbpos="across"
extfidir="r" palmor="r"/>
</split_handconfig>
<split_location>
<location_hand digits="2" contact="touch"/>
<location_hand location="wristback"
side="palmar" contact="touch"/>
</split_location>
<wristmotion motion="swinging"/>
</sign_manual>
</hamgestural_sig
```

(b) SiGML for 'film'

Fig. 1. The HamNoSys avatar signing 'film' and the corresponding SiGML code. Avatar/code examples are taken from application provided by [8]

The proposed system is based on an existing transcription method known as Hamburg Sign Language Notation System, HamNoSys developed by the Universitäit Hamburg [9] and Signing Gesture Markup Language, SiGML [10] co-developed by Virtual Signing: Capture, Animation, Storage, and Transmission (VisiCAST) and Essential Sign Language Information on Government Networks (eSign) [11]. HamNoSys is a phonetic transcription system that is independent of any specific national finger-spelling system and thus can be used to transcribe any sign language system. SiGML is

978-1-4673-5600-8/12/$31.00 ©2012 IEEE

(a) RGB and Skeleton overlaid (b) Depth Image

Fig. 2. Example data obtained from the Kinect.

XML based mark-up language built upon HamNoSys. The SiGML code generated for signs is used to drive virtual avatars as shown in Fig. 1. The system in this paper can estimate a signer's palm orientation and hand shape in order to determine the HamNoSys transcription of these features and generate the corresponding SiGML notation that will ultimately drive an avatar mimicking the estimated features.

Most methods for gesture recognition use two dimensional images or videos. With the advent of inexpensive, consumer grade depth imaging devices like Microsoft Kinect it is possible to use 3D data to improve gesture recognition. The system proposed here uses MS Kinect and the OpenNI SDK. The Kinect provides RGB and depth values as well as the skeletal joint locations of human figures in the scene. Skeletal joint locations are determined using the parts-based model [12].

The rest of this paper is structured as follows. Section 2 discusses the system set-up and proposed algorithms for palm orientation detection and hand shape classification. Experimental results are discussed in section 3 followed by conclusions.

2. PROPOSED ALGORITHM

2.1. System Set-up

The set-up of the proposed system is as follows. The main image capturing device is MS Kinect. The Kinect is interfaced using the OpenNI library which is industry-led, not-for-profit organization formed to certify and promote the compatibility and interoperability of Natural Interaction (NI) devices like Kinect. The RGB, depth and skeletal data is read in from the Kinect using this interface and is read in Matlab where the palm orientation detection and hand shape classification is performed.

2.2. Palm Orientation

Fig. 3 describes the process of estimating the palm orientation, one of the linguistic features in sign language. The process starts by first acquiring the RGB and the depth frames from Kinect. The depth data stored as 11 bit number denotes the real world distance given in millimeters measured

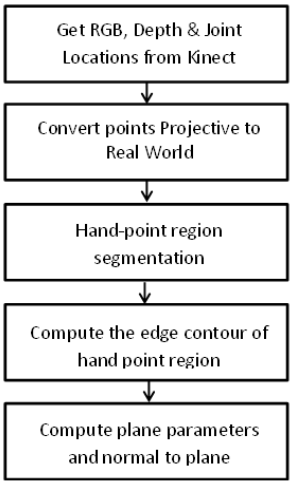


Fig. 3. Palm Orientation Algorithm.

from the center of the IR sensor. This is well explained in the Fig. 2b. The depth data thus gives the distance in real world along the *Z*-axis. The real world location of the points in *X* and *Y* axes along the rows and columns of image are obtained by calling the projective to real world conversion from the OpenNI API. We then identify the hand-point region i.e. the region around the hand joint location obtained from the skeletal tracking. In this region we apply a chain of morphological image processing methods to obtain a contour corresponding to the edge of hand palm region. First, we threshold the depth frame around the depth value corresponding to the hand joint so that the final frame contains only the depth value corresponding to the hands. Next we remove other false blobs in the image so that we only have a major blob corresponding to the hand point region. The contour obtained from this process corresponds to the outer edge of the hand. We assume that this contour lies in the plane in which the hand palm lies. To estimate the plane equation, we start forming the co-ordinate matrix from points on the contour. Let X, Y and Z = x, y and z co-ordinates of points. Then A is given by,

$$A = \begin{bmatrix} XYZ1 \end{bmatrix}^{\mathrm{T}} \quad (1)$$

We want to find a vector $\overline{w}$ that minimizes $\|A\overline{w}\|$. To avoid trivial solution, we constrain $\|\overline{w}\| = 1$, and solve using the method of Lagrange multipliers. This method establishes that $\overline{w}$ is equal to the eigenvector corresponding to the least eigenvalue of the matrix $A^{\mathrm{T}}A$. The normal to the palm is then given by the normal to the plane $\overline{w}$ and defined by $\langle w_1, w_2, w_3 \rangle$

2.3. Hand Shape Classification

To identify the hand-shapes we need a method that can classify the hand shapes according to the HamNoSys data set. We

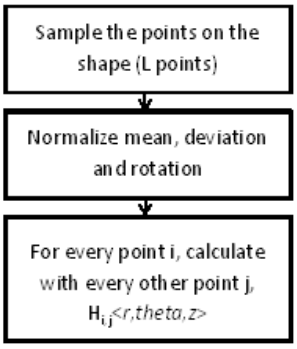


Fig. 4. Handshape classification.

propose a three dimensional extension of the shape context classification algorithm. The shape context algorithm was initially proposed by Belongie S. et. all.[13] for classifying shapes in two dimensional images. Shape contexts are calculated based on the contour of a shape which can thus define that shape. Fig. 4 describes the process of calculating the shape context. Construction begins by selecting a certain shape and capturing the three dimensional real world coordinates of the points defining the shape. The shape is further sampled to select a fixed number of points lying on the surface of the plane. The process of constructing a shape context begins by calculating the radial distance *r*, radial angle θ and the altitude z between a given point and each sample point. These distances are logarithmically binned to create a histogram corresponding to that point. We go on to create a similar histogram for each point in the set. The collective set of such histograms define the shape context of that shape. For classification we compare the test shape context with the shape context in the training set using the Chi-Square distance metric given as follows,

$$C_{i,j} = \frac{1}{2}\sum_{k=1}^{N}\frac{(h_i(k) - h_j(k))^2}{(h_i(k) + h_j(k))} \qquad (2)$$

where $h_i(k), h_j(k)$ are kth bin values of histograms of point i and j and N is total number of bins in the histogram.

3. EXPERIMENTAL RESULTS

We designed an experiment to test our system on images captured using Kinect. The algorithms for palm orientation and hand shape detection were run on a Windows machine with quad-core processor with 8 threads. Portions of the algorithm was also optimized to either run on all the 8 available threads of the CPU or on GPU using CUDA providing considerable time speed up.

Fig. 5 shows the output of palm orientation. Fig. 5a and b show the edge contour of the hand point region for two test cases. Fig.5c and d show actual points in real world in red and the points estimated using the plane we estimated in blue. The normal to the estimated plane marks the normal to the palm as shown with the blue vector in Fig.5c and d.

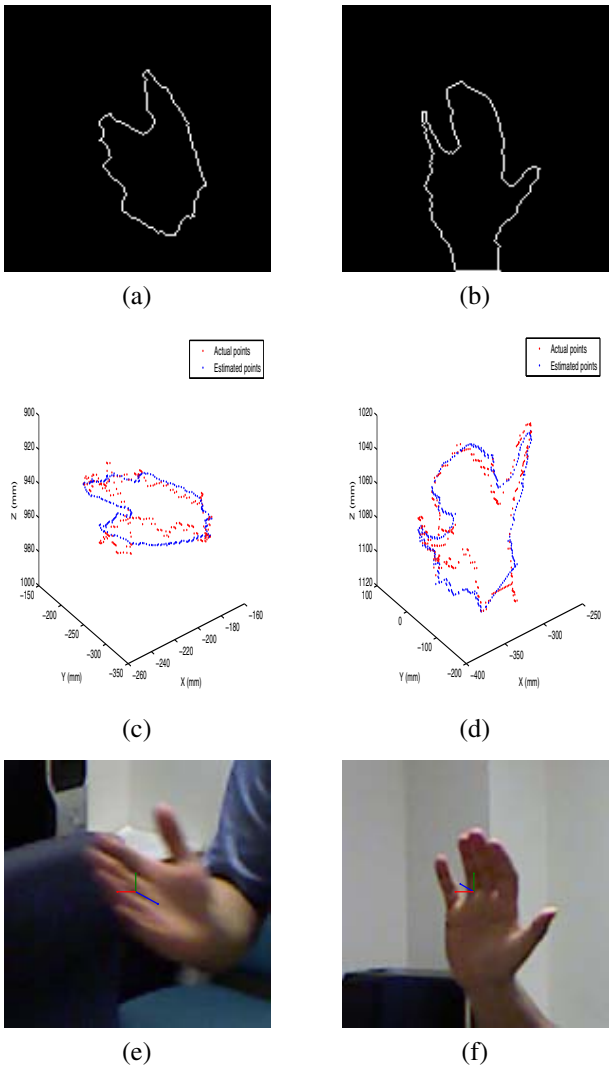
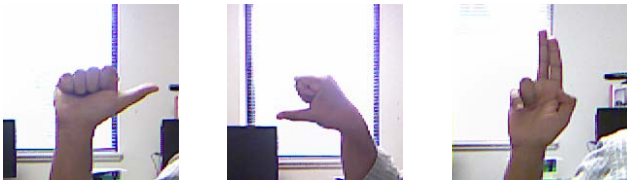


Fig. 5. (a),(b)-Edge of hand, (c),(d)-Actual and estimated points, (e),(f)-Palm Orientation mapped on RGB images.

The hand shape classification algorithm was implemented on a set 40 hand shapes selected from the Hamnosys set which consisted minimal to considerable variations with respect to number of straight fingers and thumb position. Fig. 6a, b and c shows the shapes correctly classified. The algorithm fails to differentiate the shapes with a subtle difference in their shape distributions. Such cases are shown in Fig. 7a and b. They correspond to letters "s" and "a" in ASL and the only variation is in the thumb position. From the test set, the algorithm could correctly classify 20 shapes. 10 shapes which were highly similar were confused with each other and it misclassified the remaining 10 shapes.



(a) (b) (c)

Fig. 6. Examples of correctly classified hand shapes

(a) (b)

Fig. 7. Test shapes that caused confusion because of highly similar shapes

4. CONCLUSIONS

We proposed a system for synthesis of ASL using a virtual avatar. To implement the system successfully we proposed a method to successfully extract features to identify hand palm orientation and hand shape which are two of the five important production parameters in signing in ASL. The palm orientation detection gave correct results to different hand poses. The shape context method displayed a good abilitiy to classify hand shapes. It can be further improved by increasing the shape sampling size and the number of bins. Being highly parallel in nature the process also has a good scope for high speed performance using a parallel implementation.

5. ACKNOWLEDGMENTS

The authors would like to thank Wei Yao, John Costanzo, and Harvey Rhody for helpful discussions. This research was funded in part by the RIT College of Science Dean's Research Initiation Grant Program.

6. REFERENCES

[1] O. Bernier and D. Collobert, "Head and hands 3d tracking in real time by the em algorithm," in *Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 2001. Proceedings. IEEE ICCV Workshop on*, 2001, pp. 75 –81.

[2] Ying Wu and T.S. Huang, "Hand modeling, analysis and recognition," *Signal Processing Magazine, IEEE*, vol. 18, no. 3, pp. 51 –60, may 2001.

[3] C. Nolker and H. Ritter, "Visual recognition of continuous hand postures," *Neural Networks, IEEE Transactions on*, vol. 13, no. 4, pp. 983 – 994, jul 2002.

[4] R. Yang, S. Sarkar, and B. Loeding, "Handling movement epenthesis and hand segmentation ambiguities in continuous sign language recognition using nested dynamic programming," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 3, pp. 462 –477, march 2010.

[5] T. Starner, J. Weaver, and A. Pentland, "Real-time american sign language recognition using desk and wearable computer based video," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 12, pp. 1371 –1375, dec 1998.

[6] S. Liwicki and M. Everingham, "Automatic recognition of fingerspelled words in british sign language," in *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, june 2009, pp. 50 –57.

[7] Kim B. Kurz, "An american sign language dictionary based on the shape of hands: A useful reference book," *Journal of Deaf Studies and Deaf Education*, vol. 17, no. 1, pp. 137, Winter 2012.

[8] "JNLP Applications," `http://vhg.cmp.uea.ac.uk/tech/jas/095i/`, 2012, [Online; accessed 12-October-2012].

[9] Leven R. Zienert H. Hanke T. Henning J. Prillwitz, S., *HamNoSys. Version 2.0. Hamburg Notation System for Sign Languages - An Introductory Guide.*, Signum Press, Hamburg, 1989.

[10] R. Elliott, J. R. W. Glauert, J. R. Kennaway, and I. Marshall, "The development of language processing support for the visicast project," in *Proceedings of the fourth international ACM conference on Assistive technologies*, New York, NY, USA, 2000, Assets '00, pp. 101–108, ACM.

[11] Glauert J.R.W. Kennaway J.R. Parsons K.J. Elliott, R., ," 2001, D5-2: SigML Definition Visi-CAST Project Working Document.

[12] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, june 2011, pp. 1297 –1304.

[13] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 4, pp. 509 –522, apr 2002.