

Learning Alphabets and Numbers from Hand Motion Trajectories for Visually Impaired Children

Rabin Yu Acharya, Chenyang Zhang, and Yingli Tian, *Senior Member, IEEE*

Department of Electrical Engineering

The City College of New York,

New York, NY 10031

rachary00@citymail.cuny.edu, czhang10@citymail.cuny.edu, ytian@ccny.cuny.edu

Abstract- Children with visual impairments face disproportionate challenges in learning and communicating with other kids. Integrating the learning with computer games will motivate and enhance them to learn alphabets and numbers. In this paper, we develop a computer vision-based system to recognize alphabets and numbers from hand motion trajectories captured by an RGB-D camera. While the user draws a number or a letter in front of the camera, the computer captures the hand motion trajectory and then automatically recognize it. An audio feedback is output to the blind user. Our proposed method is evaluated on a dataset contains 144 videos of 10 numbers (0-9) and 26 letters (A-Z), and achieves 86% accuracy for subject independent test.

I. INTRODUCTION

Language and communication ability plays an important role in the development of students' literacy skills and their inclusion and integration into society. According to the Braille Institute [1], there are nearly 5 million preschool-aged children and about 12.1 million children ages 6-17 have visual impairments. It is a very challenging task for blind children to learn handwriting because they do not get the rich visual feedback. Computer vision-based tools and techniques are needed to help and motivate these children with visual impairments to learn numbers and letters more easily. Research indicates that actions and gestures play an important role in the evolution of language [25].

In addition, the computer vision-based tools and techniques can also be used to enhance the interaction and communication between the children with visual impairments and normal vision [11-23]. For example, the prototype system proposed in this paper can be further extended as a computer game called "Dumb charades" for children with visual impairments. The computer or the user will act out a number or a letter which the opponent has to recognize. Dumb charades is popularly played as a word guessing game. In the form most played today, it is an acting game in which one player acts out a word or phrase, often by miming similar sounding words, and the other players guess the word or phrase. The idea behind this game is to use physical rather than verbal language to convey the meaning to another party. The eventually developed computer game can be played between a blind kid and a kid with normal vision, which will help the blind kid learn normal way of writing while enhancing his/her social skills.

The hardware needed for our system includes a computer and an RGB-D camera (ASUS Xtion Pro in our setup) [2]. Figure 1 shows two examples for online drawing the number "2" (top row) and the letter "S" (bottom row) by moving his/her hand in front of an RGBD camera. The hand motion trajectories are detected and tracked from the depth channel.

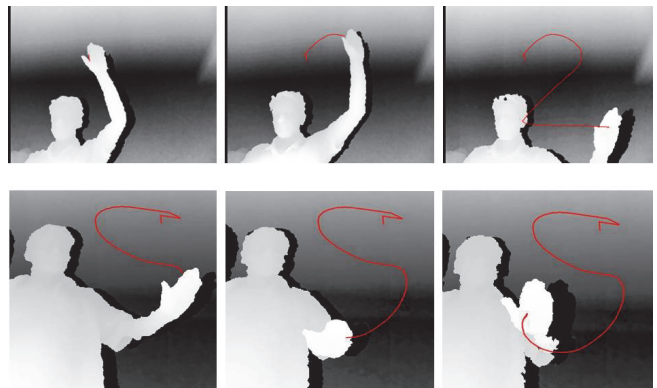


Figure 1: Users drawing the number "2" (top row) and the letter "S" (bottom row) by moving their hands in front of an RGBD camera. The hand motion trajectories are detected and tracked from the depth channel.

Handwriting in air without a pen is similar to handwriting in paper or a screen with the pen. Significant amount of work has been conducted for paper or screen handwriting recognition compared to air handwriting. Air handwriting recognition can be divided into offline (image-based) and online (real-time video-based) recognitions. Tier *et al.* surveyed different feature extraction methods for offline recognition of segmented characters [3]. Similarly, Santosh and Nattee have provided a comprehensive survey on various online handwriting recognition techniques [4].

For air handwriting recognition, Elmezain *et al.* presented a method for alphabet recognition from a motion of a single hand using Hidden Markov Models (HMM) after detecting face and hands from RGB videos [5]. Dutt and Dutt developed a system to recognize handwritten characters or symbols. Their system can handle transformation of translation, scaling or a combination of both [6]. S. Udhan *et al.* developed a basic method for recognizing the alphabets from hand motion trajectory by using artificial neural networks, however, a

sticker with a special color must be placed on user's hand for hand detection and hand trajectory tracking [7]. Escudeiro *et al.* developed a real-time bidirectional translator of Portuguese Sign Language system by employing the Microsoft Kinect and 5DT Sensor Gloves to capture and recognize the motion, shape and orientations of the hands [24].

Compared to the existing systems, our online alphabet and number recognition system is based on the depth video captured by an RGBD camera which is more robust to lighting changes and cluttered background. In addition, the RGBD camera automatically detects human body joints including hand positions.

II. PROPOSED METHOD

The proposed method for online alphabet and number recognition consists of three main stages: depth video preprocessing, feature extraction and representation, and classification.

A. Depth Video Preprocessing

We employ the depth video captured by an RGB-D camera. To obtain and track the user's hand position, Simple-Open library [8] is used which included the gesture recognition module to provide the skeleton joints tracked from depth images. The hand detection module is activated when the user his/her hand and works well for either right or left hand for the tracking process. Since the blind users cannot adjust their drawings based on the position and orientation of the camera, the position of the user is pre-determined in the setup to ensure the hand movement of the user stays within the camera view.

When the user draws a particular alphabet or a number in the air, the algorithm tracks the hand location and saves each tracked position in three dimensional coordinates (x, y, z) . The x -coordinate refers to the horizontal position, y -coordinate refers to the vertical position and the z -coordinate refers to the depth position of the hand's centroid (tracked hand).

B. Feature Extraction and Representation

The tracked hand coordinates for the hand movement are used as the trajectory features. To make the features independent of the size of the image formed, we first perform a feature normalization. The x and y coordinates are divided by the z -coordinate (the depth coordinate) since the depth controls the size of the hand trajectory image, e.g., farther distance between the hand and the camera, smaller size of the hand trajectory image is created. Furthermore, to make the features insensitive to the orientation and size of different letters/numbers, a spatial coordinate normalization is applied as shown in Figure 2.

As the example letter "S" shown in Figure 2, the image region of the letter is first calculated based on the minimum x and y coordinates of the hand trajectory. Then inspired by [9], the center of the image region (x_c, y_c) and the radius (r) of the inscribed circle of the letter is employed to create a new feature representation based on the distance and angle features.

The distance features $D = \{d_0, d_1, d_2, d_3, \dots, d_N\}$, where N is the total number of the coordinates of the hand trajectory and d_N is the distance between (x_N, y_N) and (x_c, y_c) divided by the radius r . Similarly, the angles are extracted to eliminate the effects of the orientation of different letters/numbers. In our method, all the angles are calculated based on a reference line which connects the first point of the hand trajectory and the center (x_c, y_c) . Thus the angle features $\theta = \{\theta_1, \theta_2, \dots, \theta_N\}$, where θ_N is the angle between the line formed by the trajectory point (x_N, y_N) and the center (x_c, y_c) and the reference line and θ_1 is 0 degrees. In our experiments, N is set as 150.

The distance and angle features are the concatenated as a feature vector to be used as the input for the recognition processing. These feature representation is also invariant to scale and orientation changes.

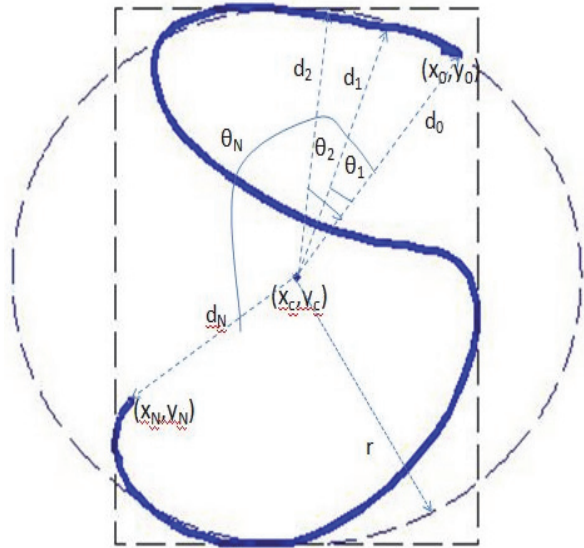


Figure 2: Spatial coordinate normalization to create orientation and scale invariant representation.

C. Classification

After extract the features for the hand drawing from the dataset we collected, Support Vector Machines (SVM) are employed as classifiers to recognize the hand writing letters or numbers. The labels of 10 digits (0-9) and 26 letters (A-Z) are represented by numbers 0 through 35 respectively. Linear scaling is also performed to avoid features in higher numeric

ranges dominating the ones in lower numeric ranges. In our experiment, the RBF kernel is used. The RBF kernel maps the data nonlinearly onto a higher-dimensional space. There are two parameters for the RBF kernel: Cost function (C) and sigma (γ). Cross-validation and Grid-search is performed to find the best value for C and γ [10]. In our experiment, C is set to 0.25 and γ is 0.0625.

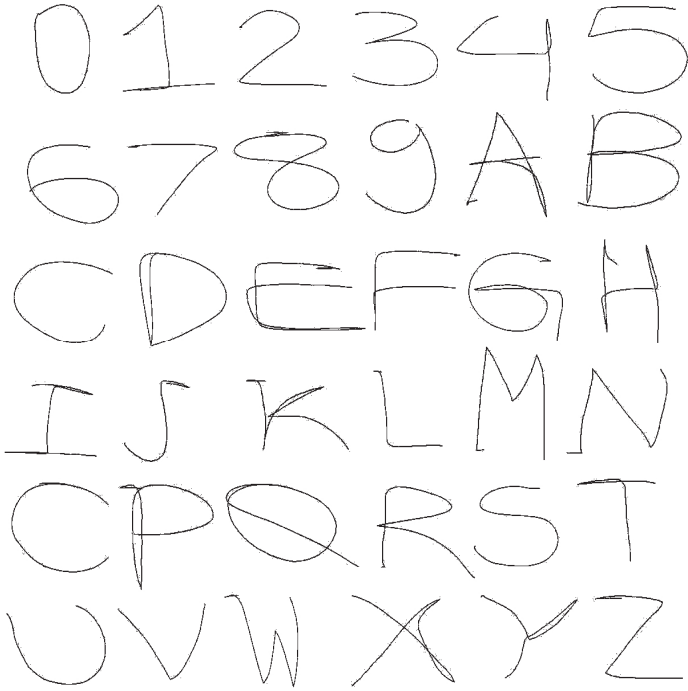


Figure 3: Hand trajectory examples of numbers and alphabets drawn by a subject in our dataset.

III. EXPERIMENT RESULTS

A. Dataset

To evaluate the proposed method, we collect a dataset contains 144 depth videos from 4 blindfolded subjects. For each subject, a total of 36 videos are captured (10 for each numbers (0-9) and 26 for each letters (A-Z)). Each subject was instructed to move his/her hands to draw a letter or a number starting from a neutral position. The hand is detected and the motion trajectory is tracked when the drawing begins. After the subject finished the drawing, he/she was instructed to keep their hand static for a couple of seconds at the end position of the drawing. For each drawing trajectory, 150 sets of coordinates were extracted. They were recommended but not required to draw around 30cm by 30cm of each letter or number. For most of the drawings, these 150 sets are enough to represent a number or a letter. Since the extracted features are

independent of the shape and orientation, specific instructions are provided to draw some letters or numbers. For example, to distinguish the number "0" from the letter "O", the subjects are instructed to make an oval shape for "0" and a circular shape for "O". To distinguish between "3" and "E", the subjects are instructed to make "E" with angled (pointed) lines. For drawing "7" they are instructed to make the down-stroke slanted. Hand trajectory examples of numbers and alphabets drawn by a subject in our dataset are listed in Figure 3.

Some researchers evaluated the efficient feedback of handwriting for visually impaired people [26, 27]. Crossan and Brewster [26] presented several studies by examining trajectory playback to transfer shape information to visually-impaired people. Plimmer *et al.* developed a system to train blind or visually impaired children to improve their handwriting by providing audio playback cues or verbal feedback from the teacher. To handling cases where the blind child needs feedback when the classifier fails to recognize, the system could provide more audio feedback that implies the character using context or suggestions like "Did you draw the letter "S"?" to achieve better user experiences.

B. Experiment Results

In our experiment, the videos from three subjects are used for training while the videos with the fourth subject are used for testing. Since there is no same subject appears in both training and testing data, the experiment is subject independent. In addition, out of the three subjects used for training, one subject is left-handed and drew all the characters and numbers using her left hand while all other subjects used their right hand. Therefore, our system works for both right-handed and left-handed users.

The experimental results demonstrated that the system works very reliably and is able to recognize the set of defined commands in real-time. The SVM-based classification achieves an average accuracy at 86%. Most wrong classified letters/numbers are from 3, E, 1, 4, J.

IV. CONCLUSION AND FUTURE WORK

In this paper, we have proposed an effective computer vision-based prototype system for recognizing alphabets and numbers to motivate and enhance the learning and communication of children with visually impairments. The proposed system is robust to light changes and cluttered background. This system is also independent of the users. The experiment results demonstrate the efficiency and effectiveness of the proposed system.

Our future work will focus on extending the system for recognizing lowercase letters and words as well as sequential recognition by taking signing history/context into consideration. We will also conduct more extensive testing and address the significant human interface issues associated human computer interaction for children with visual impairments.

ACKNOWLEDGMENTS

This work was supported in part by NSF grants EFRI-1137172, IIP-1343402, and IIS-1400802.

REFERENCES

- [1] K. Wynn, K. Psychological foundations of number: numerical competence in human infants. *Trends in Cognitive Sciences* 2, 296-303, 1998.
- [2] http://www.asus.com/Multimedia/Xtion_PRO/
- [3] O. Trier, A. Jain and T. Taxt, "Feature Extraction Methods for Character Recognition- A Survey", *Pattern Recognition*, Vol. 29, No. 4, 1996, pp. 641-662.
- [4] K. Santosh and C. Nattee, A comprehensive survey on on-line handwriting recognition technology and its application to the Nepalese Natural Handwriting. Kathmandu Univ J Sci Eng Technol 5(1):31-55, 2009.
- [5] M. Elmezain, A. Al-Hamadi, G. Krell, S. El-Etriby, and B. Michaelis, "Gesture Recognition for alphabet from Hand Motion Trajectory Using Hidden Markov Models". *IEEE International Symposium on Signal Processing and Information Technology*, pp:1192-1197, 2007.
- [6] V. Dutt and S. Dutt, "Hand Written Character Recognition Using Artificial Neural Network", *Advances in Computing*, Scientific & Academic Publishing, pp. 18-23, 2011.
- [7] S. Udhan, P. Futane and R. Dharaskar. Hand Motion Tracking for Alphabet Recognition using ANN. *International Journal of Computer Applications* 83(4):5-10, 2013.
- [8] Simple-OpenNI, <http://code.google.com/p/simple-openni/>
- [9] B. Chakraborty and G. Chakraborty, A new feature extraction technique for on-line recognition of handwritten alphanumeric characters. *Inform Sci* 148:55-70, 2002.
- [10] C. Hsu, C. Chang and C. Lin, A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University, 2003.
- [11] X. Yang, S. Yuan, and Y. Tian, "Assistive Clothing Pattern Recognition for Visually Impaired People," *IEEE Transactions on Human-Machine Systems*, Vol. 44, No. 2, pp234-243, April, 2014.
- [12] C. Yi, Y. Tian, and A. Arditì, "Portable Camera-based Assistive Text and Product Label Reading from Hand-held Objects for Blind Persons," *IEEE/ASME Transactions on Mechatronics*, Vol. 19, No. 3, pp808-817, June 2014.
- [13] S. Wang, H. Pan, C. Zhang, and Y. Tian, RGB-D Image-Based Detection of Stairs, Pedestrian Crosswalks and Traffic Signs, *Journal of Visual Communication and Image Representation (JVCIR)*, Vol. 25, pp263-272, 2014.
- [14] Y. Tian, X. Yang, C. Yi, and A. Arditì, "Toward a Computer Vision-based Wayfinding Aid for Blind Persons to Access Unfamiliar Indoor Environments," *Machine Vision and Applications*, Issue 24, No. 3, pp. 521-535, April 2013.
- [15] F. Hasanuzzaman, X. Yang, and Y. Tian, Robust and Effective Component-based Banknote Recognition for the Blind, *IEEE Transactions on Systems, Man, and Cybernetics--Part C: Applications and Reviews*, Jan. 2012.
- [16] C. Zhang, M. Shabbir, D. Stylianou, and Y. Tian, Computer Vision-based Mathematics Learning Enhancement for Children with Visual Impairments, *International Workshop on Biomedical and Health Informatics (BHI)*, 2012.
- [17] L. Tian, C. Yi, and Y. Tian, Detecting Good Quality Frames in Videos from Mobile Camera for Blind Navigation, *Journal of Computer Vision and Image Processing*, Vol. 5, No. 10, October 2015.
- [18] C. Yi and Y. Tian, Assistive Text Reading from Natural Scene for Blind Persons, Chapter in book *Mobile Cloud Visual Media Computing: From Interaction to Service*, by Springer, G. Hua and X. Hua (Eds), 2015.
- [19] Y. Tian, RGB-D Sensor-Based Computer Vision Assistive Technology for Visually Impaired Persons, Chapter in book *Computer Vision and Machine Learning with RGB-D Sensors*, by Springer, L. Shao, J. Han, P. Kohli, Z. Zhang (Eds), 2014. ISSN 2191-6586, ISBN 978-3-319-08650-7
- [20] X. Rong, B. Li, J. P. Mũno, J. Xiao, A. Arditì and Y. Tian, Guided Text Spotting for Assistive Blind Navigation in Unfamiliar Indoor Environments, *12th International Symposium on Visual Computing (ISVC)*, 2016.
- [21] B. Li, J. P. Mũnoz, X. Rong, J. Xiao, Y. Tian, and A. Arditì, ISANA: Wearable Context-Aware Indoor Assistive Navigation with Obstacle Avoidance for the Blind, *Fourth International Workshop on Assistive Computer Vision and Robotics (ACVR) in conjunction with ECCV 2016*.
- [22] C. Creusot and A. Munawar, "Real-time barcode detection in the wild," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 239-245, 2015.
- [23] A. Diallo, H. Tang, Y. Sheikh, M. Hebert, E. Molina and Z. Zhu, Accessible Gaming with Gestures Using a Close-Range Depth Sensor, *30th Annual International Technology and Persons with Disabilities Conference*, 2015.
- [24] P. Escudeiro, N. Escudeiro, R. Reis, J. Lopes, M. Norberto, A. Baltasar, M. Barbosa, J. Bidarra, *Virtual Sign – A Real Time Bidirectional Translator of Portuguese Sign Language*, 6th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Infoexclusion (DSAI), 2015.
- [25] E. A. Cartmill, S. Beilock, and S. Goldin-Meadow, A word in the hand: action, gesture and mental representation in humans and non-human primates, *Philosophical Transactions of the Royal Society B*, Volume 367, issue 1585, 2012.
- [26] A. Crossan and S. Brewster, Multimodal Trajectory Playback for Teaching Shape Information and Trajectories to Visually Impaired Computer Users. *Transactions on Accessible Computing*, 2008.
- [27] B. Plimmer, A. Crossan, S.A. Brewster, and R. Blagojevic, Multimodal collaborative handwriting training for visually-impaired people, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2008.