Proceedings of the 9th Workshop on the Representation and Processing of Sign Languages, pages 135–138
Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020

Machine Learning for Enhancing Dementia Screening in Ageing Deaf Signers of British Sign Language

Xing Liang, Bencie Woll, Epaminondas Kapetanios, Anastasia Angelopoulou, Reda Al batat
IoT and Security Research Group, University of Greenwich, UK
Cognitive Computing Research Lab, University of Westminster, UK
Deafness Cognition and Language Research Centre, University College London, UK
x.liang@greenwich.ac.uk, b.woll@ucl.ac.uk, (kapetae, agelopa)@westminster.ac.uk, w1601767@my.westminster.ac.uk

Abstract

Real-time hand movement trajectory tracking based on machine learning approaches may assist the early identification of dementia in ageing deaf individuals who are users of British Sign Language (BSL), since there are few clinicians with appropriate communication skills, and a shortage of sign language interpreters. In this paper, we introduce an automatic dementia screening system for ageing Deaf signers of BSL, using a Convolutional Neural Network (CNN) to analyse the sign space envelope and facial expression of BSL signers recorded in normal 2D videos from the BSL corpus. Our approach involves the introduction of a sub-network (the multi-modal feature extractor) which includes an accurate real-time hand trajectory tracking model and a real-time landmark facial motion analysis model. The experiments show the effectiveness of our deep learning based approach in terms of sign space tracking, facial motion tracking and early stage dementia performance assessment tasks.

Keywords: Real-Time Hand Tracking, Facial Analysis, British Sign Language, Dementia, Convolutional Neural Network

1. Introduction

British Sign Language (BSL), is a natural human language, which, like other sign languages, uses movements of the hands, body and face for linguistic expression. Identifying dementia in BSL users, however, is still an open research field, since there is very little information available about the incidence or features of dementia among BSL users. This is also exacerbated by the fact that there are few clinicians with appropriate communication skills and experience working with the BSL-using population. Diagnosis of dementia is subject to the quality of cognitive tests and BSL interpreters alike. Hence, the Deaf community currently receives unequal access to diagnosis and care for acquired neurological impairments, with consequent poorer outcomes and increased care costs (Atkinson et al., 2002). In this context, we propose a methodological approach to initial screening that comprises several stages. The first stage of research focuses on analysing the motion patterns of the sign space envelope in terms of sign trajectory and sign speed by deploying a real-time hand movement trajectory tracking model (Liang et al., 2019) based on OpenPose[1] library. The second stage involves the extraction of the facial expressions of deaf signers by deploying a real-time facial analysis model based on dlib library[2] to identify active and non-active facial expressions. Based on the differences in patterns obtained from facial and trajectory motion data, the further stage of research implements both VGG16 (Simonyan and Zisserman, 2015) and ResNet-50 (He et al., 2016) networks using transfer learning from image recognition tasks to incrementally identify and improve recognition rates for Mild Cognitive Impairment (MCI) (i.e. pre-dementia). Performance evaluation of the research work is based on data sets available from the Deafness Cognition and Language Research Centre (DCAL) at UCL, which has a range of video recordings of over 500 signers who have volunteered to participate in research. Figure 1 shows the pipeline and high-level overview of the network design.

The paper is structured as follows: Section 2 gives an overview of the related work. Section 3 outlines the methodological approach followed by Section 4 with the discussion of experimental design and results. A conclusion provides a summary of the key contributions and results of this paper.

2. Related Work

Recent advances in computer vision and greater availability in medical imaging with improved quality have increased the opportunities to develop machine learning approaches for automated detection and quantification of diseases, such as Alzheimer's and dementia (Pellegrini et al., 2018). Many of these techniques have been applied to the classification of MR imaging, CT scan imaging, FDG-PET scan imaging or the combined imaging of above, by comparing patients with early stage disease to healthy controls, to distinguish different types or stages of disease and accelerated features of ageing (Spasova et al., 2019; Lu et al., 2018; Huang et al., 2019). In terms of dementia diagnosis (Astell et al., 2019), there have been increasing applications of various machine learning approaches, most commonly with imaging data for diagnosis and disease progression (Negin et al., 2018; Iizuka et al., 2019) and less frequently in non-imaging studies focused on demographic data, cognitive measures (Bhagyashree et al., 2018), and unobtrusive monitoring of gait patterns over time (Dodge et al., 2012). These and other real-time measures of function may offer novel ways of detecting transition phases leading to dementia, which could be another potential research extension to our toolkit, since the real-time hand trajectory tracking sub-model has the potential to track a patient's daily walking

[1] https://github.com/CMU-Perceptual-Computing-Lab/openpose
[2] http://dlib.net/

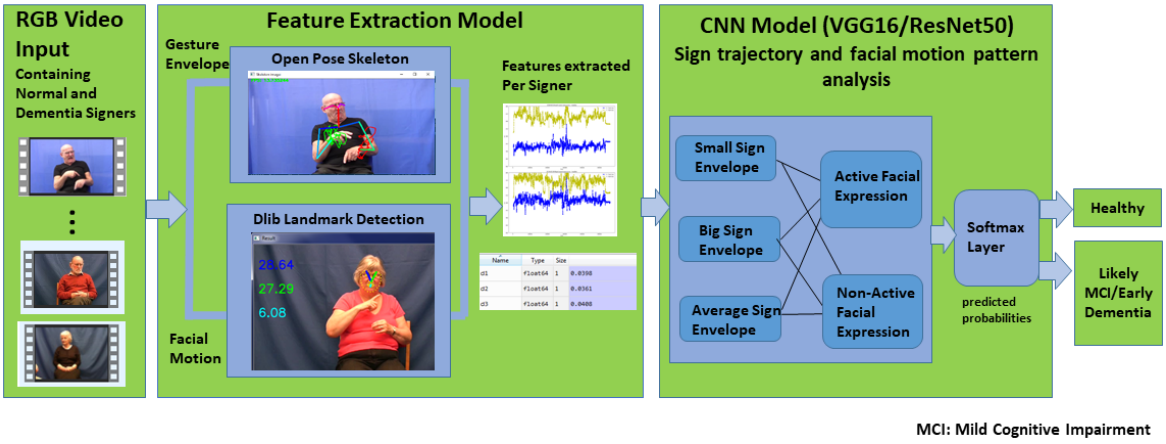

Figure 1: The Proposed Pipeline for Dementia Screening

pattern and pose recognition as well.

3. Methodology

In this paper, we present a multi-modal feature extraction sub-network inspired by practical clinical needs, together with the experimental findings associated with the sub-network. The input to the system is short term clipped videos. Different extracted motion features are fed into the CNN network to classify a BSL signer as healthy or atypical. Performance evaluation of the research work is based on data sets available from the BSL Corpus[3] at DCAL UCL, a collection of 2D video clips of 250 Deaf signers of BSL from 8 regions of the UK; and two additional data sets: a set of data collected for a previous funded project [4], and a set of signer data collected for the present study.

3.1. Dataset

From the video recordings, we selected 40 case studies of signers (20M, 20F) aged between 60 and 90 years; 21 are signers considered to be healthy cases based on their scores on the British Sign Language Cognitive Screen (BSL-CS); 9 are signers identified as having Mild Cognitive Impairment (MCI) on the basis of the BSL-CS; and 10 are signers diagnosed with MCI through clinical assessment. We consider those 19 cases as MCI (i.e. early dementia) cases, whether identified through the BSL-CS or clinically. As the video clip for each case is about 20 minutes in length, we segmented each into 4-5 short video clips - 4 minutes in length - and fed the segmented short video clip to the multi-modal feature extraction sub-network. In this way, we were able to increase the size of the dataset from 40 to 162 clips. Of the 162, 79 have MCI, and 83 are cognitively healthy.

3.2. Real-time Hand Trajectory Tracking Model

OpenPose, developed by Carnegie Mellon University, is one of the state-of-the-art methods for human pose estimation, processing images through a 2-branch multi-stage CNN (Cao et al., 2017). The real-time hand movement

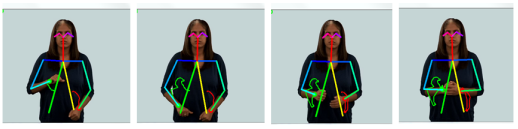

Figure 2: Real-Time Hand Trajectory Tracking for the Sign FARM

trajectory tracking model is developed based on the OpenPose Mobilenet Thin model (OpenPoseTensorFlow, 2019). A detailed evaluation of tracking performance is discussed in (Liang et al., 2019). The inputs to the system are brief clipped videos. We assume that the subjects are in front of the camera with only the head, upper body and arms visible; therefore, only 14 upper body parts in the image are outputted from the tracking model. These are: eyes, nose, ears, neck, shoulders, elbows, wrists, and hips. The hand movement trajectory is obtained via wrist joint motion trajectories. The curve of the hand movement trajectory is connected by the location of the wrist joint keypoints to track left- and right-hand limb movements across sequential video frames in a rapid and unique way. Figure 2, demonstrates the tracking process for the sign FARM. As shown in Figure 3, left- and right-hand trajectories obtained from the tracking model are also plotted by wrist location X and Y coordinates over time in a 2D plot. Figure 3 shows how hand motion changes over time, which gives a clear indication of hand movement speed (X-axis speed based on 2D coordinate changes, and Y-axis speed based on 2D coordinate changes). A spiky trajectory indicates more changes within a shorter period, thus faster hand movement.

3.3. Real-time Facial Analysis Model

The facial analysis model was implemented based on a facial landmark detector inside the Dlib library, in order to analyse a signer's facial expressions (Kazemi and Sullivan, 2014). The face detector uses the classic Histogram of Oriented Gradients (HOG) feature combined with a linear classifier, an image pyramid, and a sliding window detection scheme. The pre-trained facial landmark detector is used

[3]BSL Corpus Project, https://bslcorpusproject.org/.

[4]Overcoming obstacles to the early identification of dementia in the signing Deaf community

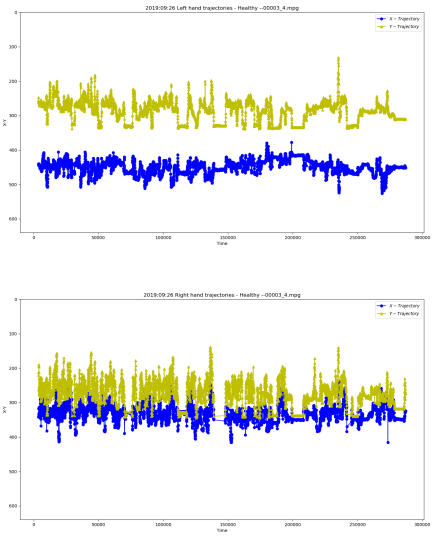

Figure 3: 2D Left- and Right- Hand Trajectory of a Signer

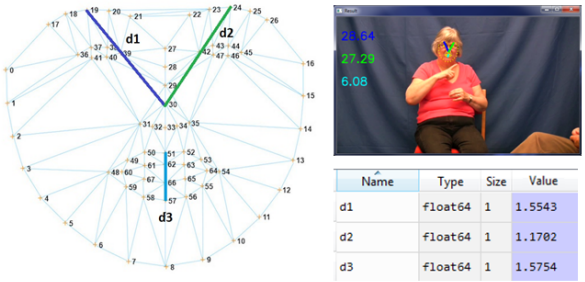

Figure 4: Facial Motion Tracking of a Signer

to estimate the location of 68 (x,y) coordinates that map to facial features (Figure 4). The facial analysis model extracts subtle facial muscle movement by calculating the average Euclidean distance differences between the nose and right brow as d1, nose and left brow as d2,and upper and lower lips as d3 for a given signer over a sequence of video frames (Figure 4). The vector [d1, d2, d3] is an indicator of a signer's facial expression and is used to classify a signer as having an active or non-active facial expression.

$$d1, d2, d3 = \frac{\sum_{t=1}^{T} |d^{t+1} - d^{t}|}{T} \quad (1)$$

where T = Total number of frames that facial landmarks are detected.

4. Experiments and Analysis

4.1. Experiments

In our approach, we have used VGG16 and ResNet-50 as the base models, with transfer learning to transfer the parameters pre-trained for the 1000 object detection task on the ImageNet dataset to recognise hand movement trajectory images for early MCI screening. We run the experiments on a Windows desktop with two Nvidia GeForce GTX 1080Ti adapter cards and 3.3 GHz Intel Core i9-7900X CPU with 16 GB RAM. In the training process, videos of 40 participants have been segmented into short clips with 162 segmented cases, split into 80% for the training set and 20% for the test set. To validate the model performance, we also kept 6 cases separate (1 MCI and 5 healthy signers), segmented into 24 cases for performance validation. Due to the very small dataset, we train ResNet-50 as a classifier alone and fine tune the VGG 16 network by freezing the Convolutional (Covn) layers and two Fully Connected (FC) layers, and only retrain the last two layers. Subsequently, a softmax layer for binary classification is applied to discriminate the two labels: Healthy and MCI, producing two numerical values of which the sum becomes 1.0. During training, dropout was deployed in fully connected layers and EarlyStopping was used in both networks to avoid overfitting.

4.2. Results Discussion

During test and validation, accuracies and receiver operating characteristic (ROC) curves of the classification were calculated, and the network with the highest accuracy and area under ROC (AUC), that is VGG 16, was chosen as the final classifier. Table 1 summarises the results over 46 participants from both networks. The best performance metrics are achieved by VGG16 with test set accuracy of 87.8788%, which matches validation set accuracy of 87.5%. In Figure 5, feature extraction results show that in a greater number of cases a signer with MCI produces a sign trajectory that resembles a straight line rather than the spiky trajectory characteristic of a healthy signer. In other words, signers with MCI produced more static poses/pauses during signing, with a reduced sign space envelope as indicated by smaller amplitude differences between the top and bottom peaks of the X, Y trajectory lines. At the same time, the Euclidean distance d3 of healthy signers is larger than that of MCI signers, indicating active facial movements by healthy signers. This proves the clinical observation concept of differences between signers with MCI and healthy signers in the envelope of sign space and face movements, with the former using smaller sign space and limited facial expression.

5. Conclusions

We have outlined a methodological approach and developed a toolkit for an automatic dementia screening system for signers of BSL. As part of our methodology, we report the experimental findings for the multi-modal feature extractor sub-network in terms of sign trajectory and facial motion together with performance comparisons between different CNN models in ResNet-50 and VGG16. The experiments show the effectiveness of our deep learning based approach for early stage dementia screening. The results are validated against cognitive assessment scores with a test set performance of 87.88%, and a validation set performance of 87.5% over sub-cases.

Astell, A., Bouranis, N., Hoey, J., Lindauer, A., Mihailidis, A., Nugent, C., and Robillardi, J. (2019). Technology and dementia: The future is now. *In: Dementia and Geriatric Cognitive Disorders*, 47(3):131–139.

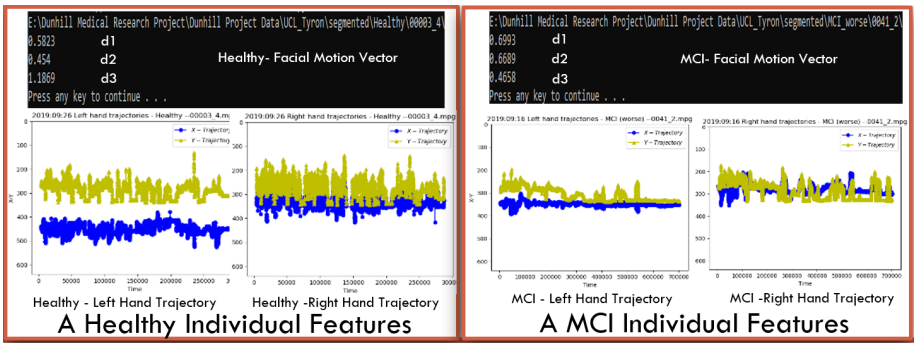

Figure 5: Experiment Finding

Table 1: Performance Evaluation over VGG16 and RestNet-50 for early MCI screening

Method	40 Participants 21 Healthy, 19 Early MCI			6 Participants 5 Healthy, 1 Early MCI	
	Train Result (129 segmented cases)	Test Result (33 segmented cases)		Validation Result (24 segmented cases)	
	ACC	ACC	ROC	ACC	ROC
VGG 16	87.5969%	**87.8788%**	**0.93**	**87.5%**	**0.96**
ResNet-50	69.7674%	69.6970%	0.72	66.6667%	0.73

Atkinson, J., Marshall, J., Thacker, A., and Woll, B. (2002). When sign language breaks down: Deaf people's access to language therapy in the uk. *In: Deaf Worlds*, 18:9–21.

Bhagyashree, S. I., Nagaraj, K., Prince, M., Fall, C., and Krishna, M. (2018). Diagnosis of dementia by machine learning methods in epidemiological studies: a pilot exploratory study from south india. *In: Social Psychiatry and Psychiatric Epidemiology*, 53(1):77–86.

Cao, Z., Simon, T., Wei, S., and Sheikh, Y. (2017). Real-time multi-person 2d pose estimation using part affinity fields. *In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Dodge, H., Mattek, N., Austin, D., Hayes, T., and Kaye, J. (2012). In-home walking speeds and variability trajectories associated with mild cognitive impairment. *In: Neurology*, 78(24):1946–1952.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *In: Proceedings of Computer Vision and Pattern Recognition (CVPR)*.

Huang, Y., Xu, J., Zhou, Y., Tong, T., Zhuang, X., and ADNI. (2019). Diagnosis of alzheimer's disease via multi-modality 3d convolutional neural network. *In: Front Neuroscience*, 13(509).

Iizuka, T., Fukasawa, M., and Kameyama, M. (2019). Deep-learning-based imaging-classification identified cingulate island sign in dementia with lewy bodies. *In: Scientific Reports*, 9(8944).

Kazemi, V. and Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. *In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Liang, X., Kapetanios, E., Woll, B., and Angelopoulou, A. (2019). Real Time Hand Movement Trajectory Tracking for Enhancing Dementia Screening in Ageing Deaf Signers of British Sign Language. *Cross Domain Conference for Machine Learning and Knowledge Extraction (CD-MAKE 2019). Lecture Notes in Computer Science*, 11713:377–394.

Lu, D., Popuri, K., Ding, G. W., Balachandar, R., Beg, M., and ADNI. (2018). Multimodal and multiscale deep neural networks for the early diagnosis of alzheimer's disease using structural mr and fdg-pet images. *In: Scientific Reports*, 8(1):5697.

Negin, F., Rodriguez, P., Koperski, M., Kerboua, A., Gonzàlez, J., Bourgeois, J., Chapoulie, E., Robert, P., and Bremond, F. (2018). Praxis: Towards automatic cognitive assessment using gesture. *In: Expert Systems with Applications*, 106:21–35.

OpenPoseTensorFlow. (2019). https://github.com/ildoonet/tf-pose-estimation.

Pellegrini, E., Ballerini, L., Hernandez, M., Chappell, F., González-Castro, V., Anblagan, D., Danso, S., Maniega, S., Job, D., Pernet, C., Mair, G., MacGillivray, T., Trucco, E., and Wardlaw, J. (2018). Machine learning of neuroimaging to diagnose cognitive impairment and dementia: a systematic review and comparative analysis. *In: Alzheimer's Dementia: Diagnosis, Assessment Disease Monitoring*, 10:519–535.

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional net-works for large-scale image recognition. *In: Proceedings of International Conference on Learning Representations*.

Spasova, S., Passamonti, L., Duggento, A., Liò, P., Toschi, N., and ADNI. (2019). A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to alzheimer's disease. *In: NeuroImage*, 189:276–287.