

Real Time Hand Movement Trajectory Tracking for Enhancing Dementia Screening in Ageing Deaf Signers of British Sign Language

Xing Liang, Epaminondas Kapetanios, Bencie Woll, Anastassia Angelopoulou

▶ **To cite this version:**

Xing Liang, Epaminondas Kapetanios, Bencie Woll, Anastassia Angelopoulou. Real Time Hand Movement Trajectory Tracking for Enhancing Dementia Screening in Ageing Deaf Signers of British Sign Language. 3rd International Cross-Domain Conference for Machine Learning and Knowledge Extraction (CD-MAKE), Aug 2019, Canterbury, United Kingdom. pp.377-394, 10.1007/978-3-030-29726-8_24 . hal-02520065

HAL Id: hal-02520065

https://inria.hal.science/hal-02520065

Submitted on 26 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution| 4.0 International License

Real Time Hand Movement Trajectory Tracking for Enhancing Dementia Screening in ageing Deaf Signers of British Sign Language

Xing Liang[1][0000−0002−6630−298X], Epaminondas Kapetanios[1][0000−0002−0617−2183], Bencie Woll[2][0000−0002−3300−4775], and Anastassia Angelopoulou[1][0000−0003−1453−492X]

[1] Cognitive Computing Research Lab, University of Westminster, UK
[2] Deafness Cognition and Language Research Centre, University College London, UK
x.liang@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. Unlike other computer vision systems used in dementia stage assessment such as RGB-D video with the aid of depth camera, activities of daily living (ADL) monitored by information and communication technologies (ICT) facilities, or X-Ray, computed tomography (CT), and magnetic resonance imaging (MRI) images fed to machine learning algorithms, the system developed here focuses on analysing the sign language space envelope (sign trajectories/depth/speed) and facial expression of deaf individuals, using normal 2D videos. In this work, we are interested in providing a more accurate segmentation of objects of interest in relation to the background, so that accurate real-time hand trajectories (path of the trajectory and speed) can be achieved. The paper presents and evaluates two types of hand movement trajectory models. In the first model, the hand sign trajectory is tracked by implementing skin colour segmentation. In the second model, the hand sign trajectory is tracked using Part Affinity Fields based on the OpenPose Skeleton Model [1, 2]. Comparisons of results between the two different models demonstrate that the second model provides enhanced improvements in terms of tracking accuracy and robustness of tracking. The pattern differences in facial and trajectory motion data achieved from the presented models will be beneficial not only for screening of deaf individuals for dementia, but also for assessment of other acquired neurological impairments associated with motor changes, for example, stroke and Parkinsons disease.

Keywords: Segmentation· Hand Tracking· OpenPose · Sign Language · Dementia · Time-series Data Analytics

1 Introduction

Most of the world's developed societies are experiencing an ageing trend in their populations[3]. Ageing is correlated with increased prevalence of cognitive impairments such as dementia, stroke and Parkinsons disease. With this in mind, researchers are working urgently to develop effective technological tools that can help doctors undertake, as precise as possible, early identification of cognitive decline. In order to capture change and to monitor behavioural patterns of ageing individuals, there have been many studies of patient monitoring and surveillance with the main focus on using ICT facilities to recognise difficulties with ADL. [4–9]. The ADL framework, using sensors, Internet of Things (IoT) and other emerging technologies, provides cost-efficient solutions for in-home or nursing-home monitoring, and can alert health-care providers to significant changes in ADL behaviours which may indicate cognitive impairment. With mounting of ICT facilities, such frameworks usually have a complex structure and need to be evaluated over an extended period of time to be useful for clinicians to detect health deterioration in patients.

Improvements in medical imaging quality and the greater availability of brain imaging data sets have increased opportunities to develop machine learning approaches for automated detection, classification and quantification of diseases. Many of these techniques have been applied to the classification of brain MRI or CT scans, comparing dementia patients to healthy controls, and to distinguish different types or stages of dementia and accelerated features of ageing [10]. As recently addressed in [11], a powerful data-driven machine learning algorithm based on a mixture of linear z-score models is used to identify the exact form and stage of Alzheimer's disease and frontotemporal dementia (FTD) from brain scans alone using an MRI image database. However, the use of neuroimaging to diagnose cognitive impairment and dementia relies on the availability of the advanced hardware and computational power of computing platforms, which results in a high cost for image interpretation.

With the rapid development of artificial intelligence technology, deep learning neural networks have begun to be applied to the automatic detection and classification of acquired neurological impairments using 3D information acquired by RGB-D cameras. [12] proposed an automatic computer-assisted cognitive assessment method for older adults using gesture recognition by means of the Praxis test which is a gesture-based diagnostic test that has been accepted as diagnostically indicative of cortical pathologies such as Alzheimer's disease. An Alzheimer's patient has to imitate the doctors gestures in doing simple movements such as waving; indicating actions, like going to sleep; rotating hands or upper body. A Deep Convolutional Neural Network (CNN) coupled with Long Short Term Memory (LSTM) is adopted to jointly perform gesture classification and fine grained gesture correctness evaluation using an RGB-D gesture video dataset recorded by Kinect v2. [13] uses a Recurrent Neural Network (RNN) with Parametric Bias to detect action anomalies of Alzheimer's patients. Supervised learning is used for action recognition by comparing the L2 distance between

pre-trained action and evaluated action. By detecting anomalous actions which do not follow the predefined actions, a patient's dementia stage can be evaluated.

The British Deaf community uses British Sign Language (BSL) as their preferred language. BSL is a natural language and, like other sign languages, uses movements of the hands, body and face for linguistic expression. BSL is unrelated to English, and has a very different grammar and lexicon. Because there are few health staff with appropriate language skills, and a shortage of BSL interpreters, the Deaf community receives unequal access to diagnosis and care for acquired neurological impairments [14], with consequent poorer outcomes and increased care costs. Inspired by the emerging and innovative technologies described above, we propose a method focusing on the analysis of the sign space envelope (the area in front of the signers upper body and head in which signs are located) and facial expressions of signers, using normal 2D videos to develop an automated screening toolkit for dementia in the ageing deaf population, thereby making possible more efficient use of the limited number of clinicians with appropriate skills and experience in diagnosis in the deaf population and ensuring early screening and provision of appropriate services and interventions [15].

Clinical observation suggests that there may be differences between signers with dementia and healthy signers in the envelope of sign space (sign trajectories/depth/speed) and movements of the face, with signers who have dementia using restricted sign space and limited facial expression compared to healthy deaf controls. Therefore the first phase of research is focusing on analysing the sign space envelope in terms of sign trajectory and sign speed, together with the facial expressions of deaf individuals, Data on healthy older signers is taken from standard 2D videos freely available from the BSL Signbank dataset [17] and compared to those with mild cognitive impairment and early stage dementia to identify changes in signing associated with dementia.

In this paper, we present two methods of real-time hand trajectory tracking models deployed in order to obtain the sign space envelope. In the first model, the hand sign trajectory is tracked by implementing skin colour filtering and morphology operations, before using contour extraction to track hand blob trajectories based on contour centroids. The second model is based on the OpenPose library for real time multi-person keypoint detection. 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 4 or 7 (Figure 3) across sequential video frames. The remainder of this paper is organised as follows. Section 2 presents the formulation and the methodology of our pipeline where two methods are evaluated using our datasets. Section 3 presents the experimental analysis, results and discussions. Finally, Section 4 concludes the study and discusses about future work.

2 Methodology

In this work, we are interested in providing a more accurate segmentation of objects of interest in relation to the background, so that accurate real-time hand

trajectories (path of the trajectory and speed) can be achieved. These segmented patches and their associated trajectories and speed of movement will be used in future work as features in a machine learning model for the classification of the sign space used by a BSL signer as healthy or atypical. Performance evaluation of the research work will be based on data sets available from the Deafness Cognition and Language Research Centre (DCAL) at UCL, which has a range of video recording of over 500 signers who have volunteered to participate in research.

Figure 1 shows the pipeline of the two methods we have applied to evaluate the datasets and future work of the machine learning model. The highlighted section and the two dashed boxes indicate the two methods for the gesture tracking given RGB video stream as input. We present results for two different baselines for feature extraction: one based on image processing methods and the other on deep learning models. Each method is discussed in more details in the following sub-sections and for each developed method we assume that the subjects are in front of the camera with only the upper body visible. The input to the system is short-term clipped videos.

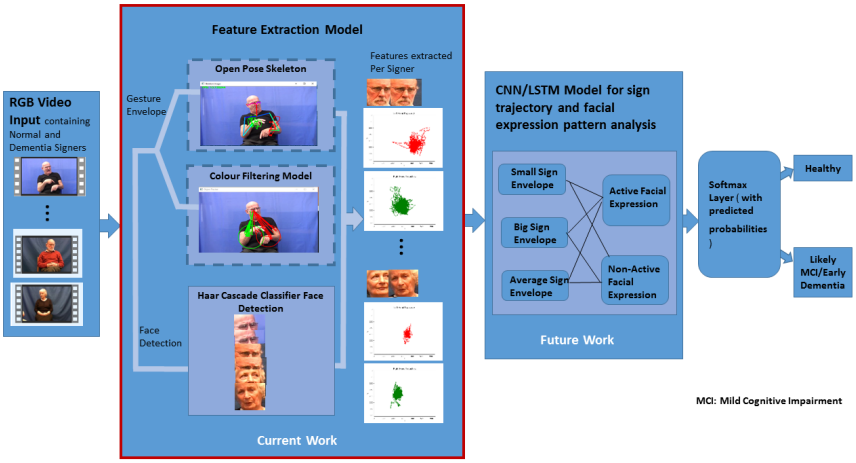


Fig. 1. The Proposed Pipeline for Dementia Screening

2.1 Datasets

British Sign Language Corpus a collection of video clips of 250 Deaf signers of BSL from 8 regions of the UK [16]. **BSL Cognitive Screen norming data** video interviews with 250 signers aged 50-90, and video recordings of a range of language and cognitive tasks (picture descriptions and memory tasks). **Video**

recordings of case studies of signers with acquired neurological disorders including dementia, left- and right-hemisphere stroke, Parkinsons disease, motor neuron disease and progressive supranuclear palsy. **BSL Signbank** standard 2D videos of single lexical signs, from an online sign dictionary [17].

2.2 Colour Filtering Models in HSV/YCrCb/Lab Colour Space

The first model for feature extraction is based on image processing method by skin colour segmentation. As shown in Figure 2, firstly face detection is performed using Haar cascade classifiers [18] for facial expression analysis. Secondly, skin colour information is used as a powerful descriptor to identify the human hands [19, 20], because human skin has a colour distribution that differs significantly (although not entirely) from background objects. Participants clothes and background have to be carefully selected to avoid similarity to skin colour. As both hands and face can be detected due to their colour similarity, but only hand blob tracking is focused on in the current stage, a rectangular box is drawn around the face (previously detected by Haar cascade classifiers). The next step is to apply skin colour thresholds to detect hand location by filtering out the skin colour distribution characteristics. As an image can be presented in a number of different colour models, such as HSV, YCrCb and CIELab, multiple colour filtering models with multi-colour thresholds for skin segmentation are used in this approach. A video frame is converted from RBG format to HSV/YCrCb/Lab format, before applying the appropriate skin segmentation thresholds.

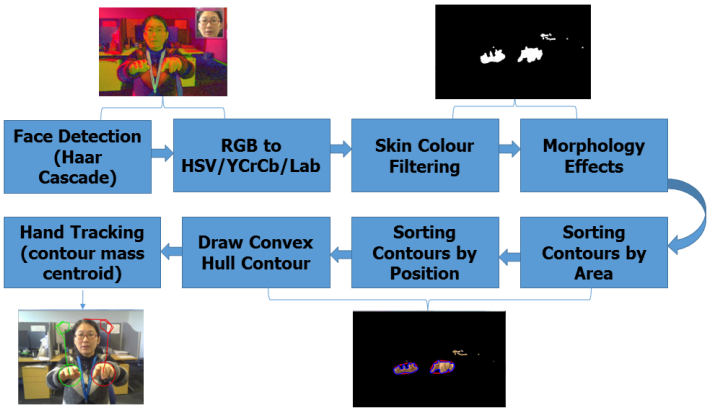


Fig. 2. Real time Hand Tracking Algorithm Based on Skin Colour Segmentation

RGB to HSV model HSV (Hue Saturation Value) is a model representing colour space, similar to the RGB (Red Green Blue) colour model. Since the

Hue channel models the colour type, it is very useful in image processing tasks that need to segment objects based on colour. Variation in Saturation goes from unsaturated (representing shades of grey) to fully Saturated (no white component). Value channel describes the brightness or intensity of the colour. In our experiments, the thresholds used for skin segmentation in the HSV model are: $0 \leq H \leq 20, 48 \leq S \leq 255, 80 \leq V \leq 255$.

RGB to YCrCb model The video frame is also converted to YCrCb format for skin segmentation. In the YCrCb model, Y is the Luminance (brightness) component. Cr (Red-difference) and Cb (Blue-difference), as colour difference signals, represent the Chrominance component. In our experiments, the thresholds used for skin segmentation in the YCrCb model are: $0 \leq Y \leq 255, 133 \leq Cr \leq 173, 77 \leq Cb \leq 127$.

RGB to CIELab model The video frame is also converted from CIELab format for skin segmentation. Lab colour space is defined by the International Commission on Illumination. It expresses colour as three numerical values, L for lightness and a and b for the greenred and blueyellow colour components. In our experiments, the thresholds used for skin segmentation in the CIELab model are: $20 \leq L \leq 220, 128 \leq a \leq 245, 130 \leq b \leq 255$. In order to measure the performance between the segmented skin colour region obtained by the three different colour models and the ground truth, we applied the S⊘rensen Dice coefficient, as a standard segmentation performance metric, to all three colour models. The S⊘rensen Dice index, measures the spatial overlap between two segmentations, the A and B regions (in our case these are the ground truth image and each segmented image taccording to he three colour models), and is defined as

$$Dice = \frac{2 \mid A \cap B \mid}{\mid A \mid + \mid B \mid} \tag{1}$$

We also used a second segmentation metric known as the Jaccard similarity coefficient which measures the number of pixels common to both the ground truth and the segmented regions, divided by the total number of pixels present across both regions.

$$Jaccard = \frac{\mid A \cap B \mid}{\mid A \cup B \mid} \tag{2}$$

After skin colour segmentation, which captures only the values between the lower and upper thresholds for skin detection, morphology operations are applied to the binary mask in order to get rid of the noisy specks. This procedure consists firstly of Erosion (to remove pixels at the boundaries of an object in the image), followed by Closing (i.e. Dilation followed by Erosion), and Dilation again (to add pixels to the boundaries of an object in the image). Basically in the morphology approach, by removing the pixels at the boundaries of an object and adding them back, small white noisy specks are eroded. Clearer hand blobs are obtained, as shown in Figure 2. After these steps, contour extraction is applied using the inbuilt OpenCV function [21]. The output of the contour function is a 2-dimensional array containing the list of x, y coordinates for all the contours, an array of points that are part of a curve and have the same pixel intensities.

Sorting contours by areas helps to extract the largest two contours (i.e. the two hands). At the same time, by sorting the largest two contours by position using the x coordinate, both hands are detected from left to right. Then a convex hull and a normal contour are drawn on the hand contour. Finally the hand trajectory is tracked by connecting its contour mass centroid, while the tracking time is recorded for the purpose of sign speed analysis.

2.3 OpenPose Skeleton Model

In the second model, hand movement trajectory tracking is based on the OpenPose library. OpenPose, developed by Carnegie Mellon University, is one of the state-of-the-art methods for human pose estimation. It processes images through a two-branch multi-stage CNN. The first branch takes the input image and predicts the possible locations of each keypoint in the image with a confidence score (the confidence map). The second branch predicts a set of 2D vector fields that encode the location and orientation of limbs over the image domain (the part affinity fields). Finally the confidence maps and the affinity fields are parsed by greedy inference to output the 2D keypoints for all people in the image [1].

OpenPose consists of three different blocks: body/foot detection; hand detection; and face detection. The core block is the combined body/foot keypoint detector, which provides a 15-,18-, or 25-keypoint body/foot keypoint estimation[22]. The computational performance on body keypoint estimation is invariant to the number of detected people in the image. It can be used on various platforms, including Ubuntu, Windows and Mac, and also has been implemented in different deep learning frameworks such as Tensorflow and Torch. In this paper, the hand tracking model implementation is based on the OpenPose Mobilenet Thin model in Tensorflow [23] on the Windows CPU/GPU platform, for 18 keypoints of body part keypoint estimation (including eyes, nose, ears, neck, shoulders, elbows, wrists, hips, knees and ankles) as shown in Figure 3 [1, 2]. These 18 joint coordinates are able to track limb and body movement in a rapid and unique way. For our purpose, only 14 upper body part joints of the signer in the image are outputted from the OpenPose skeleton model, since only the upper body of a singer is involved in signing. These are: eyes, nose, ears, neck, shoulders, elbows, wrists, and hips, as illustrated in Figure 4. Wrist keypoints 4 and 7 are utilised for left and right hand tracking respectively, corresponding to the joints motion trajectory as shown in Figure 4.

3 Evaluation of Results

The results presented in this section are from initial stage data analysis, mainly based on real time web camera capture of data and standard 2D videos from BSL Signbank [17]. Section 3.1 evaluates the skin filtering results for different colour models using video frames from BSL Signbank. To demonstrate the model capability of hand tracking, section 3.3 uses not only the videos from BSL Signbank but also real time web camera capture data as the input. Hand tracking

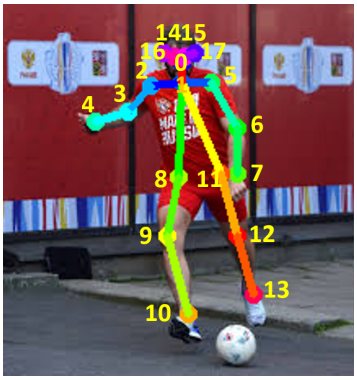


Fig. 3. OpenPose Skeleton 18 Body Joints [1, 2]

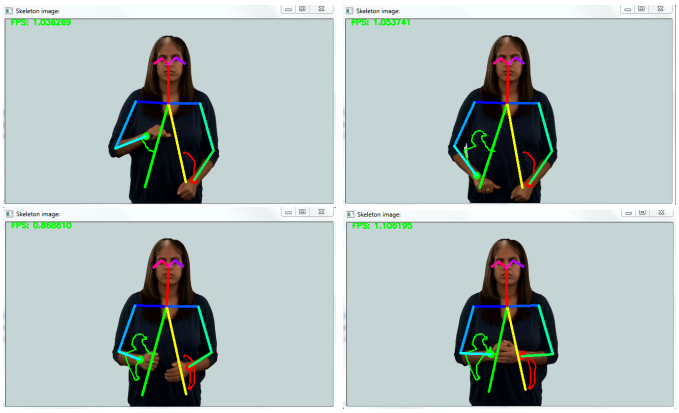


Fig. 4. OpenPose Skeleton Model Hand Trajectory Tracking

trajectories from real time web camera capture are compared with ground truth collected by a magnetic positional tracker (Polhemus 3Space Fastrak tracking instrument). For two hand trajectory tracking, the Polhemus tracking instrument reports the positional coordinates of each hand with 60 updated coordinate points per second, a static accuracy of 0.08 cm, and resolution of 0.0005 cm/cm of range as indicated in product specifications [24].

The first hand tracking model (colour segmentation based) was developed and tested on a desktop machine, 8 GB RAM 3.00 GHz Intel Core i5-4590S CPU processor. This method was implemented in Python 3.6.5 and OpenCV 3.3.1. The second hand tracking model (OpenPose skeleton based) was developed and tested on the same CPU desktop, and on a GPU desktop with two NVIDIA GeForce GTX 1080Ti adapter cards and 3.3 GHz Intel Core i9-7900X CPU with 16 GB RAM. The second model was implemented in Tensorflow 1.11, Python

3.6.5, OpenCV 3.3.1 for the CPU environment and Tensorflow 1.12, Python 3.6.8, OpenCV 3.4.2 for the GPU environment.

3.1 Colour Models Evaluation

Figures 5 and 6 show the skin segmentation comparisons between the different colour models. In each colour model, the colour thresholds play an important part in segmentation. For colour thresholds presented in section 2.2, Figures 5 and 6 show that HSV and CIELab outperform the YCrCb colour models, with better skin filtering results [20] and less error mapping. Table 1 shows quantitative results for Figure 6 for all three colour models based on the two segmentation metrics as discussed in section 2.2.

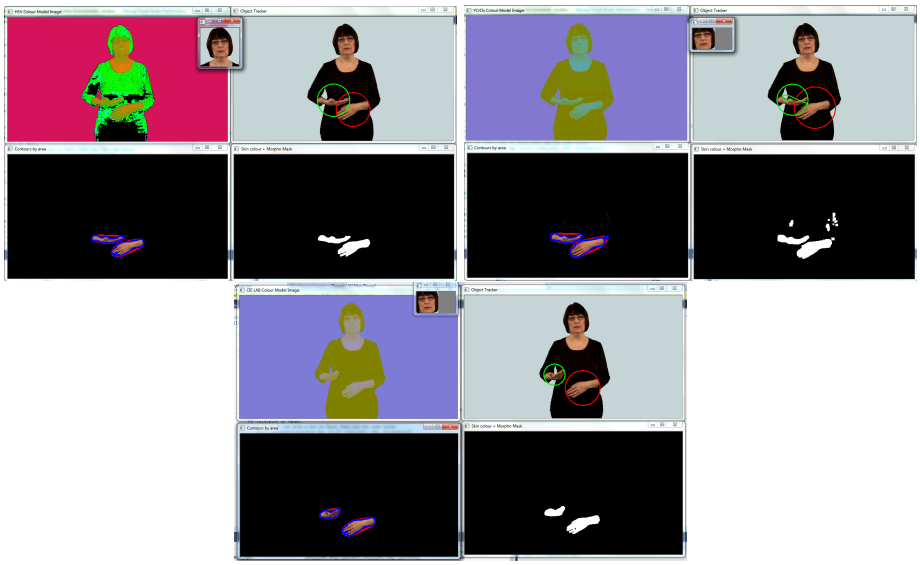


Fig. 5. Comparisons between multiple colour filtering models for skin segmentation (from left to right: (a) HSV, (b) YCrCb, (c) CIELab)

3.2 Real Time Tracking Trajectory Evaluations

Figure 7 shows 2D hand tracking trajectory results from the real time hand tracking demo (Figure 2). Three signs, differing in location: CLOUD, PICTURE, and SAILOR are clearly tracked, based on skin colour segmentation. Figure 8 is the 3D real time hand tracking trajectory. Hands are tracked not only based on 2D coordinates, but also in time, with the purpose of tracking the speed of hand movement. In the left hand trajectory (Figure 7 and Figure 8), there is an

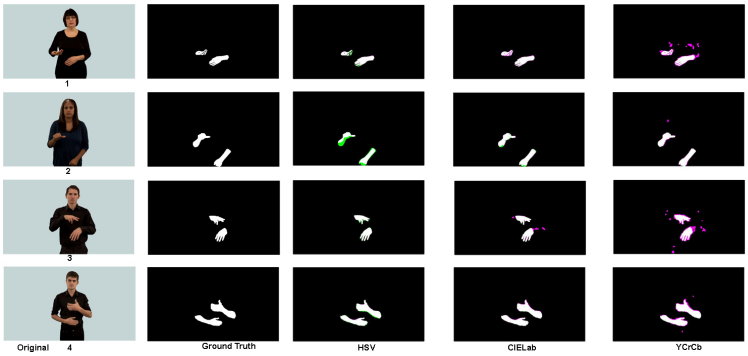


Fig. 6. Different colour models and their associated error map. From left to right column: Original image, Ground Truth, HSV, CIELab, YCrCb and their associated error map (green and magenta pixels). Green pixels indicate False Negatives and magenta pixels indicate False Positives

Table 1. Jaccard and Dice scores for Signing images

Signing Image	Colour Models	Dice score	Jaccard score
1	HSV	**0.87508**	**0.77914**
	CIELab	0.86467	0.7616
	YCrCb	0.70551	0.55474
2	HSV	0.78313	0.64356
	CIELab	**0.88988**	**0.8016**
	YCrCb	0.88428	0.79256
3	HSV	**0.88683**	**0.83695**
	CIELab	0.83965	0.72362
	YCrCb	0.70207	0.54591
4	HSV	0.90132	0.82037
	CIELab	**0.90239**	**0.82215**
	YCrCb	0.86992	0.76978

clear match between 2D and 3D trajectory. Figure 9 shows how speed of hand motion changes over time in a 2D plot, which gives a clear indication of how hand movement speed over time (X-axis speed based on 2D coordinate changes, and Y-axis speed based on 2D coordinates changes). By introducing another dimension in time (milliseconds), hand movement speed pattern can be easily identified to analyse acquired neurological impairments associated with motor symptoms (i.e. slowered movement) such as in Parkinson's disease. A longer trajectory within a shorter period shown in the right hand 3D trajectory (green Diagram in Figure 8) indicates faster hand movement.

Figure 10 are the 2D plot (x-axis vs. time and y-axis vs. time) comparing hand movement tracking in a Deaf individual with Mild Cognitive Impairment (MCI) and a healthy individual. It shows that the MCI signer's trajectory resembles a

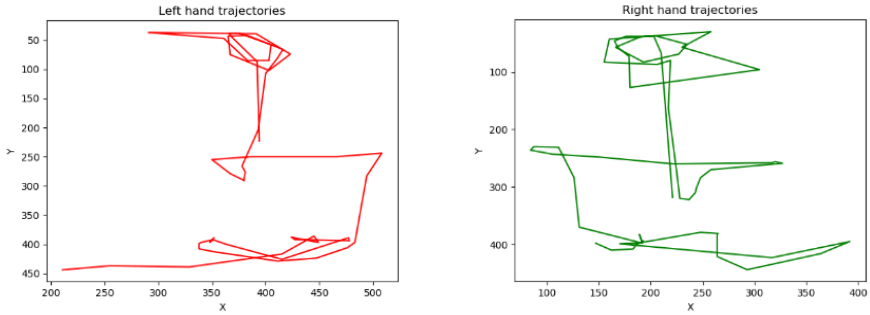


Fig. 7. 2D Real Time Hand Tracking Trajectory

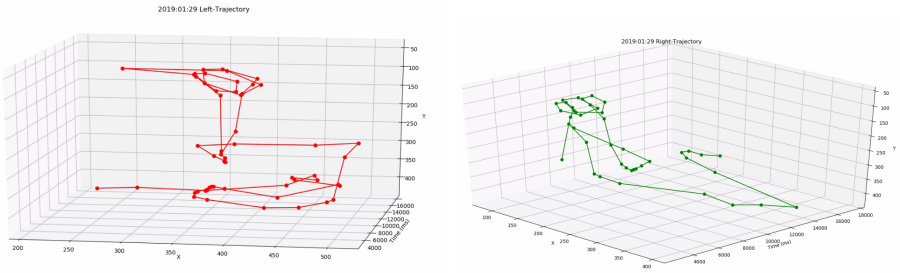


Fig. 8. 3D Real Time Hand Tracking Trajectory

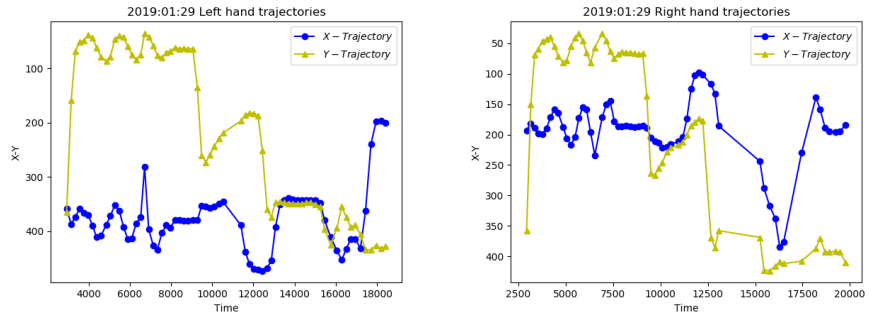


Fig. 9. 2D Real Time Hand Tracking Trajectory over Time

straight line rather than the up and down trajectory characteristic of a healthy individual, indicating that the MCI signer produced more static poses/pauses

during signing. Moreover, the X and Y trajectory lines of the signer with MCI are closer to each other as a result of a limited sign space envelope.

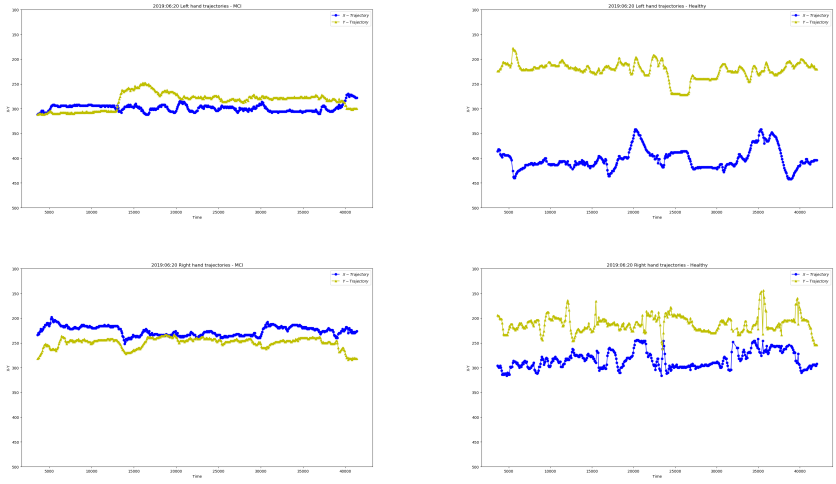


Fig. 10. 2D Real Time Hand Tracking Trajectory between MCI and Healthy individual (from left to right: (a) MCI-Left, (b) Healthy-Left, (c) MCI-Right, (d) Healthy-Right)

3.3 Comparisons between Two Tracking Models

In order to compare the two proposed real time hand tracking models, firstly data from real time web camera capturing were analyzed. The hand tracking trajectory obtained from the tracking model was then compared with its ground truth as collected by the Polhemus magnetic tracker. As shown in Figure 11, two receivers of the magnetic tracker are attached to both wrists and used to track the ground truth trajectory at the same time as the tracking model performs tracking. So far we are measuring the ground truth and its trajectory data obtained with each tracking model individually in a qualitative way. Figure 12 shows the differences between the tracking models and the ground truth in the hand trajectory of the sign DEAF. They clearly indicate that, on the right figure, the tracking trajectory is closer to its ground truth: that is, the skeleton tracking model performs better in terms of accuracy.

In order to compare the two proposed real time hand tracking models, videos of the same signs from BSL Signbank are also applied to each model. Figure 13 and Figure 14 show selected tracking results for the sign FARM. Comparing the sign trajectories in Figure 13 and in Figure 14, it can be seen that the OpenPose skeleton model is more accurate with respect to the ground truth

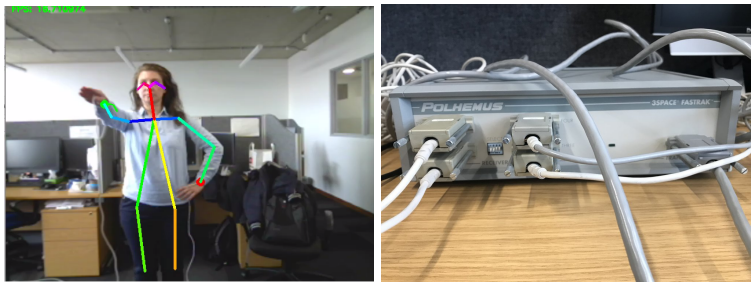


Fig. 11. Ground Truth Data Collection Setup

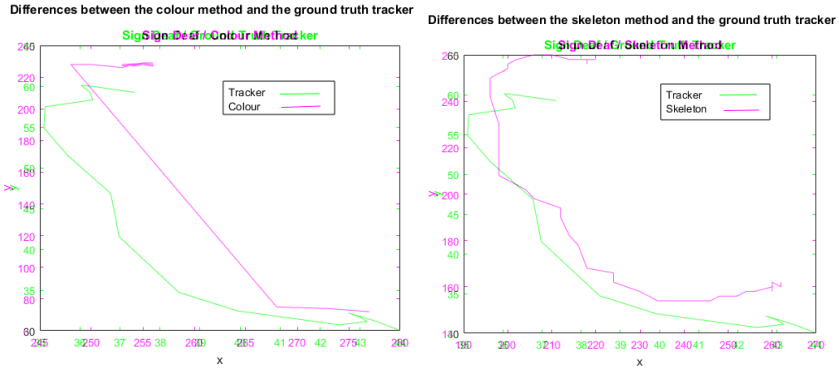


Fig. 12. Differences between Trajectory Obtained from Tracking Models and its Ground Truth

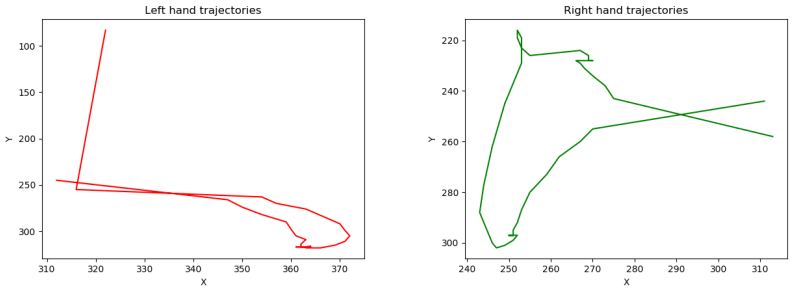


Fig. 13. 2D Sign Tracking Trajectory from Colour Filtering Model

trajectory. Figure 15 takes a closer look at the left hand trajectory of Figure 13. When in a case where the Haar classifier failed in face detection. This may have occurred because prominent black features are missing or because the image

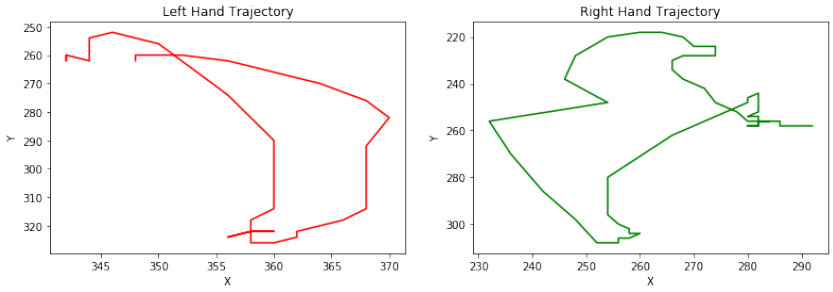


Fig. 14. 2D Sign Tracking Trajectory from OpenPose Skeleton Model

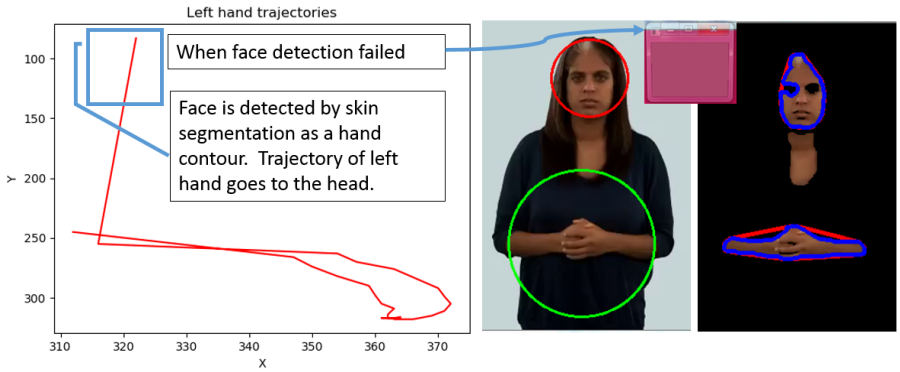


Fig. 15. Analysis of 2D Left Hand Sign Tracking Trajectory from Colour Filtering Model

is very bright, meaning that the mask could not be drawn on the face as no face detection bounding box was returned. The face was then detected by skin segmentation and was sorted as a hand contour due to its size. At the same time, when two hands were joined together, they were segmented and taken as a single hand contour. Consequently, the left hand tracking trajectory in Figure 15 is incorrectly connected to the head as highlighted in the blue box. Similarly, when a face is partially occluded or turnws to the side, a Haar classifier will fail in detection, and the skin segmentation model will generate an inaccurate trajectory. Despite the above mentioned drawbacks, the skin segmentation model is easy to implement, and performs a relatively fast and accurate tracking result under low operating system requirements.

CNN-based part detection using the OpenPose skeleton model is not influenced by the colour of the background and participants clothing. This makes it more robust in hand tracking. Figure 14 shows that details of changes in hand movement are also well tracked. This is because the model uses the wrist joint

for motion trajectory. Unlike the contour centroid that can be shifted as the gesture or posture changes, the PAF of the wrist joint is relatively stable.

The performance of the second model relies highly on the system's computational capability. In order to have a better knowledge of its tracking performance in speed, we applied the model in Windows with both a CPU and a GPU. The system specifications are listed in Table 2. The processing speed of the web camera input is slower than the video input. However, with utilisation of the GPU, overall performance speeds are greatly improved. In conclusion, the second model has significantly enhanced tracking accuracy and is more robust in tracking. To obtain the best performance especially for web camera capture, system computational capability plays a key role. As low processing capability causes loss of tracking points, this will add errors to real time trajectory tracking.

Table 2. OpenPose Tracking Model Performance between GPU and CPU

	System Specifications	Average FPS
CPU	Win7, Intel Core CPU@3.00GHz, RAM 8GB	1.2 (Video) 0.9 (Web Camera)
GPU	2 NVIDIA GeForce GTX 1080Ti, Win10, Intel Core CPU@3.30GHz, RAM 16GB	60 (Video) 23 (Web Camera)

3.4 Discussion

In this project, we needed to decide on what would be the most promising approach to pursue in order to maximise the probability of extracting hand gestures and their trajectories that were as accurate as possible. Not investigating and comparing the main approaches in this context, would have been detrimental to the quality of the follow-up process of extraction of high level features from the related sign envelope. This, in turn, would have affected the quality of interpretation of the information provided by the sign envelope in relation to one that might be potentially atypical. In that sense, the main two approaches selected for comparison: skin colour filtering and pose (skeleton), can be viewed as representatives of larger families of algorithmic approaches: image processing techniques versus pre-trained machine learning (ML) models.

The experiments and comparisons verified that the pre-trained, ML based model is superior to the image processing one, in several aspects: simplicity in setting up the experiment, simplicity in capturing hand gesture trajectories as it is less sensitive to background and environment, increased speed and accuracy of measurement. Even if pre-trained ML-based models for skin colour filtering are used, which has not been the case in our methodology and comparisons, the pose (skeleton) based approach retains its superiority with regard to simplicity in

setting up the experiment and simplicity in capturing hand gesture trajectories as it is less sensitive to background and environment.

As far as accuracy is concerned, ground truth data other than the video-recorded signs have been used. In order to increase trustworthiness in our methodological approach and comparisons, tracking data from real participants articulating signs have also been captured as test data to be used for verifying the hypothesis that the pose (skeleton), pre-trained ML-based approach delivers more accurate hand trajectories. Accuracy is defined as the closest possible trajectory to the one captured by the hand tracker. This is also based on the assumption that a tracker's data trajectories are close to real signs; hence, the use of this data set as ground truth data. In that sense, the captured trajectories from image processing and the skin colour filtering approach significantly deviated from the tracker data based trajectories. As pointed out previously, even if such a pre-trained ML-based, skin colour filtering system does exist or will be developed, it is unlikely that better accuracy will be achieved, and if so, it would be at the expense of complexity and intrinsic vulnerability to errors. It is also worth mentioning that the tracker data have been captured twice, once with each approach. Hence, for the sake of fairness, we decided to use only the tracker data corresponding to either the skin colour filtering or pose (skeleton) approach, respectively. An average of the two trajectories could also be drawn and used as common ground truth data, however, no significant difference with the results and comparisons will be observed.

Finally, first preliminary comparisons with real patient data has confirmed the significance of this methodological approach and the comparison results in identifying the approach delivering most accurate hand trajectories possible. Although at a very early stage, it appears that there is difference between healthy Deaf individuals and those with early evidence of mild cognitive impairment.

4 Conclusions

Two types of real time hand movement trajectory tracking models have been introduced with the aim of enhancing dementia screening in ageing deaf signers of BSL. In the first model, hand sign trajectory is tracked by implementing skin colour filtering to track hand blob trajectories based on contour centroids. As an image can be presented in a number of different colour space models, multiple colour space filtering models with multi-colour thresholds (HSV/YCrCb/Lab) for skin segmentation are also addressed, to perform relatively accurate and fast hand tracking with low platform requirements. The second model is based on the OpenPose library for real time multi-person keypoint detection. The hand movement trajectory is obtained via wrist joint motion trajectory. It provides enhanced tracking accuracy and more robust tracking. To obtain the best performance, system computational capability plays an important role and as such the implementation has been performed on both CPU and GPU architectures. Based on the differences in patterns obtained from facial and trajectory motion data, further research work will implement machine learning and deep neural

network models (CNN/LSTM/Hybrid) for the incremental improvement of dementia recognition rates. The final screening toolkit will be trained and validated against behavioural cognitive screening tests designed for users of BSL. As the proposed system focuses on analysing the sign space envelope and facial expression of BSL signers using normal 2D videos without requiring any ICT/medical facilities setup, the proposed system will be more economical, simpler, more flexible, and more adaptable.

5 Funding

This work has been supported by the Dunhill Medical Trust Grant RPGF1802\37 UK.

References

1. Cao, Z., Simon, T., Wei, S.E., Sheikh, Y.: Realtime multi-person 2d pose estimation using part affinity fields. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7291-7299. IEEE Press, Honolulu (2017).doi: 10.1109/CVPR.2017.143.
2. Cao, Z., Hidalgo, G., Simon, T., Wei, S.E, Sheikh, Y.: OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. In: arXiv preprint arXiv:1812.08008 (2018).
3. Kleinberger, T., Becker, M., Ras, E., Holzinger, A,. Mller, P.: Ambient Intelligence in Assisted Living: Enable Elderly People to Handle Future Interfaces. In: International Conference on Universal Access in Human-Computer Interaction, pp. 103112, Springer, Berlin (2007).
4. Urwyler,P., Stucki,R., Rampa, L., Mri,R., Mosimann, U., Nef, T.: Cognitive impairment categorized in community-dwelling older adults with and without dementia using in-home sensors that recognise activities of daily living. In: Scientific Reports. Vol. 7, 42084 (2017).
5. Banerjee, T., Keller, J. M., Popescu, M., Skubic, M.: Recognizing complex instrumental activities of daily living using scene information and fuzzy logic. In: Computer Vision and Image Understanding. Vol. 140, pp. 6882(2015).
6. Negin, F., Cogar, S., Bremond, F., Koperski, M.: Generating unsupervised models for online long-term daily living activity recognition. In: 3rd IAPR Asian conference on Pattern recognition (ACPR), pp. 186190. IEEE Press, Kuala Lumpur (2015).
7. Sheriff, R.: Employing ICT in Smart Cities for the Health and Well-Being of Older People with Dementia. In: RISUD Annual International Symposium (RAIS) Smart Cities. Hong Kong (2016). doi: 10.13140/RG.2.2.24997.29923.
8. Enshaeifar, S., Zoha, A., Markides, A., Skillman, S., Acton, S.T., Elsaleh, T., Hassanpour, M., Ahrabian, A., Kenny, M., Klein, S., Rostill, H., Nilforooshan, R., Barnaghi, P.: Health management and pattern analysis of daily living activities of people with dementia using in-home sensors and machine learning techniques. In: PLoS ONE. Vol. 13: e0195605 (2018). doi:10.1371/journal.pone.0195605
9. Singh, D., Merdivan, E., Psychoula, I., Kropf, J., Hanke, S., Geist, M., Holzinger, A.: Human Activity Recognition using Recurrent Neural Networks. In: Lecture Notes in Computer Science LNCS 10410. Cham: Springer International, pp. 267-274 (2017).

10. Pellegrini,E., Ballerini, L., Hernandez, M., Chappell, F., Gonzlez-Castro, V., Anblagan, D., Danso, S., Maniega, S., Job, D., Pernet, C., Mair, G., MacGillivray,T., Trucco, E., Wardlaw, J,: Machine learning of neuroimaging to diagnose cognitive impairment and dementia: a systematic review and comparative analysis. In: arXiv: 1804.01961 (2018).
11. Young, A., Marinescu, R., Oxtoby, N., Bocchetta, M., Yong, K., Firth, N., Cash, D., Thomas, D., Dick, K., Cardoso, J., Swieten, J., Borroni, B.,Galimberti, D., Masellis, M.,Tartaglia, M.,Rowe,J., Graff, C., Tagliavini, F., Frisoni, G., Laforce Jr R., Finger E.,Mendona, A., Sorbi, S., Warren, J., Crutch, S., Fox, N., Ourselin, S., Schott, J., Rohrer, J.,Alexander, D., The Genetic FTD Initiative (GENFI), The Alzheimers Disease Neuroimaging Initiative (ADNI): Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. In: Nature Communications. Vol. 9, 4273 (2018). doi: 10.1038/s41467-018-05892-0.
12. Negin, F., Rodriguez, P., Koperski, M., Kerboua, A., Gonzlez, J. Bourgeois, J., Chapoulie, E., Robert, P., Bremond, F.: PRAXIS: Towards Automatic Cognitive Assessment Using Gesture. In: Expert Systems with Applications. Vol. 106, pp.21-35 (2018).
13. Iarlori, S., Ferracuti, F., Giantomassi, A., Longhi, S.: RGBD camera monitoring system for Alzheimers disease assessment using Recurrent Neural Networks with Parametric Bias action recognition. In: Proceedings of the 19th World Congress the International Federation of Automatic Control (IFAC), pp. 3863-3868. Cape Town (2014).
14. Atkinson JA, Marshall J, Thacker A, Woll B.: When sign language breaks down: Deaf people's access to language therapy in the UK. In: Deaf Worlds. Vol.18, pp.9-21 (2002).
15. Liang, X., Angelopoulou, A., Woll, B., Kapetanios E.: Enhancing Dementia Screening in ageing Deaf Signers of British Sign Language via Analysis of Hand Movement Trajectories. In: Workshop of RSLondonSouthEast2019. Royal Society, London (2019).
16. British Sign Language Corpus Project. https://bslcorpusproject.org/.
17. BSL SignBank. http://bslsignbank.ucl.ac.uk/.
18. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In:Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), pp.511-518. IEEE Press, Kauai(2001). doi:10.1109/CVPR.2001.990517.
19. Chai, D., Ngan, K.: Face Segmentation Using Skin-Color Map in Videophone Technology. In: IEEE Transactions on Circuits and Systems for Video Technology. Vol.9, pp. 551-564 (1999). doi: 10.1109/76.767122.
20. Angelopoulou, A ., Garcia Rodriguez, J., Orts-Escolano, S., Kapetanios, E., Liang, X., Woll, B., Psarrou, A.: Evaluation of different Chrominance Models in the Detection and Reconstruction of Faces and Hands using the Growing Neural Gas Network. In: Springer Journals of Pattern Analysis and Applications, pp.119 (2019).
21. OpenCV. https://opencv.org/.
22. OpenPose. https://github.com/CMU-Perceptual-Computing-Lab/openpose.
23. OpenPose in Tensorflow. https://github.com/ildoonet/tf-pose-estimation.
24. O'Suilleabhain, PE., Dewey, RB.,Validation for tremor quantification of an electromagnetic tracking device. In: Movement Disorders, 16: pp. 265 271 (2001).