Deep Learning Shape Trajectories for Isolated Word Sign Language Recognition

Sana Fakhfakh
L3S Laboratory, El Manar University Tunis, Tunisia
sana.fakhfakh@enis.tn

Yousra Ben Jemaa
L3S Laboratory, El Manar University Tunis, Tunisia
yousra.benjemaa@enis.tn

Abstract: *In this paper, we propose an efficient trajectories analysis solution for the recognition of Isolated Word Sign Language (IWSL). The key technique innovation in this work is the shape trajectories analysis based on the deep learning method and achieved impressive results on different IWSL data sets: German: Rheinisch Westfälische Technische Hochschule(RWTH): RWTH-Boston-50 and RWTH-Boston-104(95.83%), Signer-Independent Continuous Sign Language Recognition for Large Vocabulary Using Subunit Models (SIGNUM: 98.21%) and new Tunisian Sign Language database (TunSigns: 98%).*

Keywords: *Sign language, isolated word recognition, shape trajectory analysis, deep learning, RWTH-Boston dataset, SIGNUM corpora.*

Received January 17, 2020; accepted February 9, 2021
https://doi.org/10.34028/iajit/19/4/10

1. Introduction

The natural communication tools among the hearing impaired communities are presented as Sign Language (SL) in the world. Since, it presents the primary communication cue of the deaf community, it becomes a necessity to know and understand the characteristics of this specific language. With all technological progress, many systems are proposed in order to offer an automatic SL translation system. It concerns continuous signs or isolated word signs [1]. This work addresses the Isolated Word Signs Language (IWSL) only. All word gestures can be performed using the hand and/or the body and the face. IWSL recognition still faces two great challenges due to the high signers and gesture variability such as sign execution speed variation and signer's interchangeability. For instance, any word gesture can be performed by different deaf in various shapes, hand poses and speeds. To address these issues, the majority of existing works rely on motion as one of the most basic and helpful cues in designing a large dynamic SL vocabulary [5, 6, 7, 8, 9, 10, 16]. These methods are meant principally to modelling the motion data relying on spatio-temporal analysis [2]. Unfortunately, these methods ignored inter and intra motion characterization and excluded the semantic spatial relationships between all IWSL gestures. However, when applying dynamic gesture, we can state that the signer builds a spatial data which describe each new gesture as a specific signature based on fingers trajectories derived information. So, naturally, gesture recognition system is closely associated with the spatial trajectory data representation, precisely shape trajectory description.

Thus, in IWSL recognition system, it is important to appropriately explain each word sign based on movement and shape patterns. Therefore, in order to avoid the presented issues (speed variation and signer's interchangeability), we propose a new method based on trajectory analysis. First, we provide a powerful way to extract all word gestures trajectories. Second, we provide a shape analysis for all extracted trajectories for more appropriate recognition task. We propose to adopt spatial information related to extracted hand trajectory shape representation. Precisely, we propose a new system able to analyze trajectory as image through applying a deep learning concept. So, we propose to define each word sign gesture as images representing combination of all spatial motion trajectories. In this context, the spatial data are introduced and the time information, presented in the dynamic data, is totally eliminated.

After extracting trajectory information in our proposed system, each word gesture will be presented and trained as a picture. These two principal steps of feature extraction and classification are combined using high-level abstractions of data and using only one architecture presented as deep learning technique. We propose, here, to use the network structure of Convolutional Neural Network (CNN) which is one of the deep learning architecture.

It can automatically extract multiple features from low features to high ones. Currently, CNN is a state of the art of image pattern recognition. It improves performance and achieves a high recognition rate in different domains and for different databases [10]. The proposed approach, not only allows to solve problems of speed and signer's interchangeability variations, but

also ensures invariance to scale and rotation variations when extracting features of word gesture [15].

The rest of this paper is organized as follows. First of all in section 2 a quick literature glance is presented about trajectory concept and SL recognition methods. Section 3 describes our proposed approach based on trajectories analysis and CNN technique. Then an experimental evaluation and performance analysis are conducted in Section 4. Finally, section 5 concludes this paper.

2. Related Works

Trajectory processing is widely used in different applications such as action recognition [3], real time character recognition [13], internet of Things [9].

In SL application, the trajectory was also an interest of different works.

Many techniques are introduced in this context to model hand motion trajectory like number of minima in the velocity, the standard deviation of the speed and the average speed [4], combining hand orientation and hand location information [19], conic presentation [5], Kernel Principal Component Analysis (KPCA) [11], But these approaches should not be viewed as performant with complex trajectories.

Other works proposes to introduce trajectory information acquired from acquisition devices [12] or Kinect camera [14] to find more pertinent motion features with ignoring background and clothing conditions and take into consideration complexes gestures [17]. Although these approaches give performant results, they are always surrounded by too many material acquisition constraints and then considered so expensive.

Work in [10] proposes to integrate CNN with tracking model to have a hand model characterization in order to achieve a good hand tracking system. The obtained results confirm the importance of this combination compared to existing methods. A 89.33% recognition rate is obtained using RWTH-BOSTON database. This proves the importance of introducing together shape and motion trajectory analysis.

According to these already existing works, the spatial shape trajectories analysis step was usually abandoned. There are no existing IWSL recognition systems based on trajectory shape analysis which treat extracted trajectories as a picture. This idea has two advantages. First, it introduces only spatial data and ignores temporal dimension constraints. Second, it takes advantage of CNN in image pattern recognition task. So, we are the first who propose it and we will demonstrate its performance and robustness.

3. Proposed Approach

Our system is based on two important steps:

1. *Trajectory extraction step*: it consists on extracting trajectories related to each word sign. These extracted trajectories are plotted and saved as an image. So, a new image-trajectories database is collected instead of the IWSL gesture database.
2. *Deep trajectory shape analyzing step*: the created database is well worn as input to convolution neural network architecture. The CNN is applied to analyze and recognize the obtained gesture shapes-trajectory.

3.1. Trajectory Extraction Step

This step aims to extract trajectories which can describe each word sign gesture. All details about trajectory extraction step are presented in [7]. In this contribution, motion and shape cues are considered together when presenting trajectory of gesture dynamic. So, each isolated word gesture is presented as moving points in the time. The proposed points' extraction strategy is based on finger flexion concept [8] which relies on a vision-based approach without environmental (clothing, background, dominant hand) and devices constraints. In fact, two principal levels, static and dynamic, are proposed in order to extract trajectories.

In static-level, 17 points are extracted from the head/hands region in the first frame of video gesture. Then, in dynamic-level, 17 trajectories related to the 17 proposed points in the rest of the video gesture are generated based on particular filter. Finally a trajectory matrix is generated for each word gesture. A Removing of Redundant Frame (RRF) step is also proposed to eliminate redundancy and reduce time processing [7].

In addition, in order to have invariant system to clothes condition, we propose to add a wrist detection step before extracting the 17 proposed key points and after head/hands regions localisation. The wrist line detection process is presented in [6]. It is based on 5 steps as illustrated in Figure 1.

a) Input image. b) Hand rotated in the vertical direction. c) Hand bounding and wrist line finding only in the 3 lower parts after division step. d) Output image after removing all pixels below the wrist line detected

Figure1. Wrist detection process.

3.2. Deep Trajectory Shape Analyzing

The proposed method uses a CNN as a powerful method to extract features of the input word trajectory in one hand and as a classifier in the other hand [20]. In the training process by CNN, we propose to plot the

extracted trajectory and create one image related to each word sign's trajectory. This way offers the possibility to work only with spatial data and eliminate the time parameter presented in the video of word gesture. The input of the CNN is then an image of size 32*32 and its outputs, after classification, are the scores for all categories of IWSL gestures. The overall architecture of the proposed CNN is illustrated in Figure 2. It consists of 7 layers: three convolution layers, three pooling layers and fully connected layer (Softmax: 8 classes).

The convolution layers and the pooling layers are used for feature extraction followed by fully connected layers used as classifier.

The activation function of the network is the Rectified Linear Unit (ReLU) function and each convolution layer is followed by a 2*2 max pooling layer.

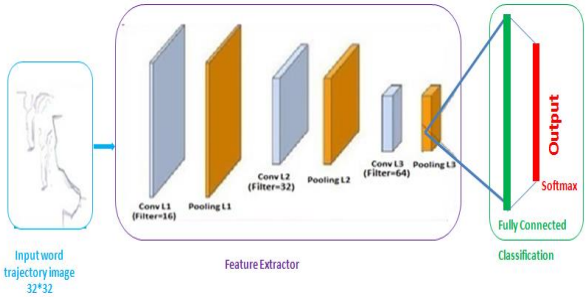


Figure 2. The proposed CNN architecture.

4. Performance Analysis

4.1. Evaluation on Our Dataset

4.1.1. The Proposed Tunisian IWSL Database

On account of the absence of public database for the Tunisian SL, we have created a new isolated word database. We named it TunSigns. Here, we propose to treat only family theme which is composed of 25 gestures. The proposed database is collected with complex environment by 9 different signers who with different ages and gender. So, 225 videos were generated.

Figure 3 illustrates some signers presented in the TunSigns database. Figure 4 presents the gestures signature description related to family theme of the Tunisian SL.

Figure 3. Examples of TunSigns database signers.

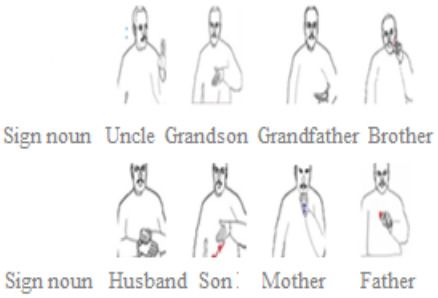


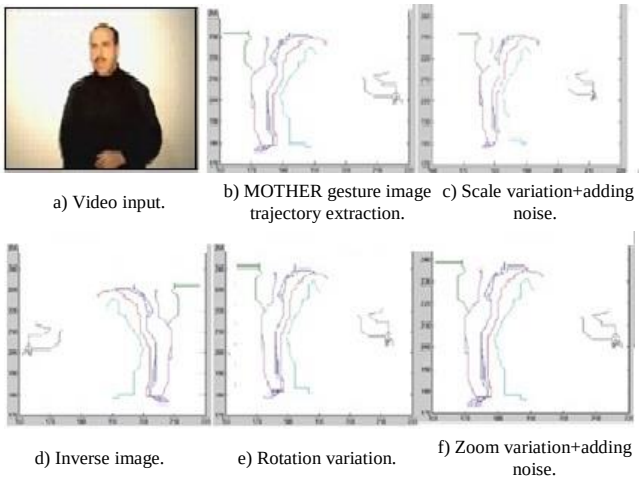
Figure 4. Gesture content: family theme of the TunSigns.

4.1.2. Trajectory Database Preparation

In this section we propose to treat only 8 gestures: mother, father, husband, brother, grandfather, son, uncle and grandson. In order to build a powerful image classifier, we augmented the number of images in our dataset via random transformations. In general, training the network with a dataset that already contains those transformations allows the network to better generalize the features of each class and recognize them under those conditions.

In addition, we propose to create a new independent shape trajectory database in order to be used in future works directly as a reference database containing shape trajectories Tunisian gestures. Many transformations are proposed based on 3 principal steps (See Figure 5):

- Step 1: we generated 7 transformations for each plotted image by applying different operations such as: rotation, scaling, zoom, adding random noise, cropping random parts, changing the brightness, changing the contrast.
- Step 2: we propose to invert all collected images to take into consideration when the gestures are executed by right or left hand.
- Step 3: we apply these two steps with and without removing redundant frame [7]. So the number of collected images is multiplied by 2. Finally, a 2016 ((9*7*2*2)*8) images are generated to constitute the database related to 8 gestures.



a) Video input. b) MOTHER gesture image trajectory extraction. c) Scale variation+adding noise.

d) Inverse image. e) Rotation variation. f) Zoom variation+adding noise.

Figure 5. Trajectory transformations for MOTHER gesture.

4.1.3. Test Protocol

Performance criteria used here are CCR, recall and precision. In addition to F score in order to evaluate the overall recall and precision performances calculated as follows in Equation (1):

$$F\ score = \frac{Recall * Precision}{Recall + Recision} * 2 \quad (1)$$

The images of the database are split to 70% for the training phase (1408 images: 176 images for each sign) and 30% for the test phase (608 image, 76 images for each sign). The CNN trained model related to the proposed 8 gestures for 25 epochs is presented in Figure 6.

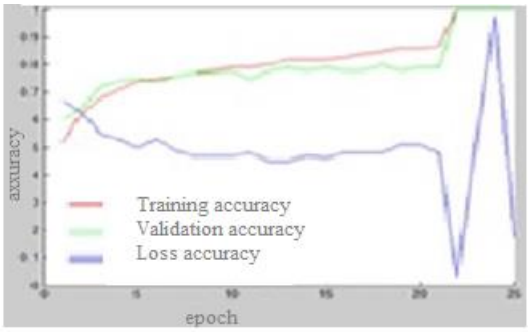


Figure 6. Accuracy of CNN trained model versus epochs.

The first 100% validation accuracy is obtained in epoch 22 and it persists in epoch 23, 24, and 25.

This demonstrates the effectiveness of our training process model. In addition to that, we show an augmentation of loss values after epoch number 22 and a brisk diminution in epoch 25. This can be a sign to arrest training process at an earlier epoch. So the choice of 25 epochs is confirmed.

4.1.4. Experimental Results

As already precised, we suggest to evaluate our proposed system based on Recall, Precision, F score and CCR metrics since it can be defined as the important evaluation index in gesture recognition system. Table 1 illustrates the obtained results when applying our proposed architecture to 25 epochs.

Table 1. Classification performances of the proposed system.

Gesture	Recall	Precision	F score
Mother	100%	100%	100%
Father	100%	100%	100%
Husband	100%	98%	98.98%
Brother	100%	96%	97.95%
Grandfather	100%	97%	98.47%
Son	100%	100%	100%
Uncle	100%	97%	98.47%
Grandson	100%	96%	97.95%
CCR Mean	98%		
F score Mean	98.97%		

The reached CCR is 98%. These satisfactory performances are supported by the fact that images containing trajectory gesture represent a robust feature able to describe gesture precisely.

In addition, the proposed CNN support all brisk changes related to gesture motion characteristics such as rotation, scale, brightness, contrast, noise, trajectories variations.... Also, the F scores related to 8 gestures have great values with F score mean 98.97%. All mean recall rates are the same, which means that our proposed system's inter-class classification is balanced.

4.2. Evaluation on Public Datasets and Comparison with Existing Works

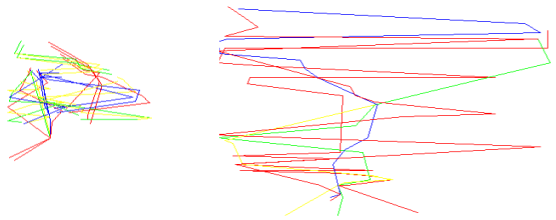
4.2.1. Comparison with [10]: Benefit of Shape Trajectory Analysis

To highlight the superiority of the proposed approach and to prove shape trajectory analysis step performance, we propose to compare our approach to the recent work of [10]. There are some challenges in this work, which becomes more suitable to compare it to our approach. First, this work projects the same general word sign gestures definition. The correct recognition of a sign contains place when it is based, since start to the end, on the hands shape together with the hand movement description. Second, we applied the same techniques differently as CNN and particle filter in order to have a good hand tracking and hand description. In fact, work [10] don't use shape trajectories in order to recognize word sign gestures. As presented above in the related works section, in hand tracking phase, [10] proposes to use particle filter. Also CNN pre-trained hand models related to left and right hand were combined with the extracted hand motion data in order to extract square hand region based on the hand predicted position. Second, a Hand Energy Image (HEI) is applied in the segmented regions as hand characteristics. To have a faithful comparison, we apply the same test conditions used in

[10]. It proposes to use RWTH-Boston-50[1] corpus in testing step and RWTH-Boston-104[2] training corpus in training step. Four cameras are used when collecting the proposed datasets: one in front of the signer and the others in lateral positions. The RWTH-BOSTON-50 corpus constitutes of 50 isolated SL. The RWTH-BOSTON-104 contains 104 continuous sign. The two proposed datasets are performed in the training and testing data by three different signers (two females and one male) dressed differently.

In addition, all video files presented in the two proposed datasets are captured with a variety of sizes and speeds. That offers the possibility to prove again speed variation and signer's interchangeability challenges. Also, only 15 gestures are employed in [10]. The choice of these gestures is based initially on the use of both left and right hand when applying the sign. To apply our proposed approach, an extraction of the 15 word gestures from training data becomes a necessity because each video data is presented as sentence in RWTH-Boston-104 training corpus. The extraction of all word gestures step is based on the existing "ground-truth" data which illustrate each start and end gesture position. This is suitable for our proposed deep learning approach to maintain a good shape trajectory analysis.

After words gestures extraction step and as described in section 3.1, a pre-treatment step was applied in order to extract all words trajectories related to the two used datasets. A sample of obtained trajectories is shown in Figure 7. Second, a deep trajectory shape analysis based on our proposed CNN model as described in section 3.2 is applied after creating a new shape-trajectories database. Finally, each shape-trajectory word gestures is presented by 240 images; 3600 images in the total (both test and train) after applying all three transformation steps presented in section 4.1.2.



a) BOOK gesture trajectory image. b) WRITE gesture trajectory image.

Figure 7. Examples of trajectory images.

As presented in Table 2, a 95.83% recognition rate is obtained after applying our proposed approach compared to 89.33% with the work of [10].

[1]https://www-i6.informatik.rwth-aachen.de/aslr/database-rwthboston-50.php

[2]https://www-i6.informatik.rwth-aachen.de/aslr/database-rwthboston-104.php

Table 2. Performances of the 2 approaches in terms of CCR for the RWTH-Boston database.

Approach	CCR: RWTH-Boston database
[10]	89.33%
Proposed approach	95.83%

Our approach outperforms the approach of [10] with 6.5%. In fact, the work of [10] is influenced by different factors like the changeability of hand sizes and the variety of the distances to the camera in the used database. This proves:

- The gap of [10] approach around the scale invariance and the robustness of our system to geometry variations (rotation, scale, translation), environment conditions (colors, lighting, contrast, background, user clothes ...) and prediction errors.
- The invariance of our approach to speed and signers interchangeability conditions.

These performances highlight:

- The importance of presenting each gesture with a pertinent trajectory. That implicitly describes the hand pose each time.
- The presence of a semantic relationship between all extracted gesture trajectories proved by a spatial data representation.
- The importance of applying a shape trajectory analysis, with introducing the CNN as feature detector and as classifier, which can extract the semantic relationship between all trajectories: precisely between all pertinent extracted key points.

These performances prove that our system introduces a natural deaf reasoning which automatically and implicitly describes a geometric and a semantic pattern when presenting each IWSL gesture signature.

4.2.2. Comparison with [7]: Benefit of Deep Learning Technique with CNN

To prove, once more, the importance of deep learning shape trajectory analysis step. We compare our proposed approach to the work of [7]. In fact, motion and shape hand modalities are introduced together to models a visual word gesture presentation. Thus, the proposed process is based on natural finger flexion function and shows that the gesture dynamism can be mirrored using hand key points trajectory picture. Three principal contributions are introduced and detailed in this work [7]:

- *Eliminate redundancy*: in order to eliminate frame redundancy and decrease time processing based on pixels number counting related to connected components.
- *The static level*: executed in the first frame and occupied two principal stages in order to take into consideration all complex environmental

conditions(background, lighting, clothing, colors,...) without instrument acquisition:

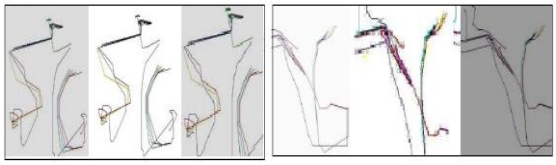
- *Region of Interest detection (ROI)*: occupied to extract head and hands regions. In this case, three steps are introduced. First, skin color detection is effectuated based on colors space YCbCr robust to lighting conditions. Second, closing and opening techniques are introduced to eliminate the obtained small regions after skin colors extraction step. Only the three biggest segmented regions which are generally linked to head and hands regions are reserved. Viola and Jones technique is also applied in order to identify face region and guarantee a performing ROI localization stage. The two rest regions are considered as hands objects.
- *Key point extraction*: 17 interest points are proposed which are: 3 gravity centers for the head and hands objects in addition to 14 points related to left and right hands which are extracted using the convexity defect concept (10 fingers tip and four wrist line hand extremity positions).

- *The dynamic-level*: occupied with the extraction of Key Point Trajectory Matrix (KPTM) related to the 17 introduced key points. 17 particular filters are used in tracking phase.

All extracted KPTM matrixes are trained using SVM classifier. We propose to apply two tests. The first is similar to the work of [7] by applying the same test conditions and Signer-Independent Continuous Sign Language Recognition for Large Vocabulary Using Subunit Models (SIGNUM) databases [18]. The second is similar to our proposed approach presented in section 3.2 using TunSigns databases.

Test 1 using SIGNUM databases. Only 8 gestures taken from SIGNUM corpora are introduced in our evaluation step: book, bus, sell, butter, apple, banana, computer, and pizza.

After trajectory extraction step, as presented in the work of [7], a generation of shape-trajectory database becomes a necessity. In this case, each gesture is presented with 140 images (see Figure 8).



a) Book Gesture trajectories images. b) Butter gesture trajectories images.

Figure 8. Examples of trajectories images from the new shape-trajectory SIGNUM database.

Table 3. Classification performances the two approaches for SIGNUM and TunSigns databases.

Approach	SIGNUM	TunSigns
[7]	CCR: 95.41%	CCR: 95.21%
Proposed approach	CCR: 98.21% F-score: 99.10%	CCR: 98 % F-score: 98.97%

As shown in Table 3, a 95.41% recognition rate is obtained when applying the approach proposed in [7] versus 98.21% when adding our deep learning shape trajectory analysis step with the 8 gestures taken from the SIGNUM database. So an amelioration of 3% is observed. Test 2 using TunSigns database. Here, we propose to train the KPTM with a Support Vector Machine (SVM) classifier related to the 8 proposed isolated Tunisian words as presented in the work [7]. Also, as shown in Table 3, a 95.21% recognition rate is obtained when applying the approach proposed in [7] versus 98% when adding our deep shape trajectory analysis step with the CNN technique. An amelioration of 3% is observed proving the capacity of CNN to strongly characterize shape trajectories related to the dynamic pattern, especially the IWSL gesture.

5. Conclusions

In this paper, we have presented a deep learning shape trajectory analysis for isolated word signs language recognition concept.

With minimum of constraints and with natural environment conditions, our proposed system has achieved high performances with the new proposed isolated word TunSigns database (98 %) as well as other databases such as RWTH-Boston (95.83%) and SIGNUM (98.21%) and outperforms the state-of-the-art methods [7, 10]. This is due to the advantage of introducing deep shape trajectory analysis step based on natural deaf reasoning, principally, on the relationship between the signer's fingertips and their brain.

In the future, we will try to improve the same idea in IWSL domain with introducing geometry space concept [3] and take advantage about geodesic distance as elastic metric.

References

[1] Agrawal S., Jalal A., Tripathi R., "A Survey on Manual and Non-Manual Sign Language Recognition for Isolated and Continuous Sign," *International Journal of Applied Pattern* Recognition, vol. 3, pp. 137-145, 2016.

[2] Balaji S. and Karthikeyan S., "A Survey on Moving Object Tracking Using Image Processing," *in Proceedings of International Conference on Intelligent Systems and Control*, Coimbatore, pp. 469-474, 2017.

[3] Ben-Tanfous A., Drira H., and Ben-Amor B., "Coding Kendall's Shape Trajectories for 3D Action Recognition," *in Proceedings of IEEE Computer Vision and Pattern Recognition*, Salt Lake City, pp. 2840-2849, 2018.

[4] Bhuyan M., Bora P., and Ghosh D., "Trajectory Guided Recognition of Hand Gestures Having Only Global Motions," *World Academy of*

Science, Engineering and Technology, vol. 21, pp.753-764, 2008.

[5] Boulares M. and Jemni M., "3d Motion Trajectory Analysis Approach to Improve Sign Language 3dbased Content Recognition" *Procedia Computer Science*, vol. 13, pp. 133-143, 2012.

[6] Fakhfakh S. and Ben-Jemaa Y., "Hand and Wrist Localization Approach for Features Extraction in Arabic Sign Language Recognition," *in Proceedings of IEEE/ACS 14th International Conference on Computer Systems and Applications*, Hammamet, pp. 774-780, 2017.

[7] Fakhfakh S. and Ben-Jemaa Y., "Gesture Recognition System for Isolated Word Sign Language Based on Key-Point Trajectory Matrix," *Computación y Sistemas*, vol. 22, no. 4, pp. 1415-1430, 2018.

[8] Gopura R., Bandara D., Gunasekera N., Hapuarachchi V., and Ariyarathna B., "A Prosthetic Hand with Self-Adaptive Fingers," *in Proceedings of 3rd International Conference on Control, Automation and Robotics*, Nagoya, pp. 269- 274, 2017.

[9] Li G., Wu H., Jiang G., Xu S., and Liu H., "Dynamic Gesture Recognition in the Internet of Things," *IEEE Access*, vol. 7, pp. 23713-23724, 2019.

[10] Lim K., Tan A., Lee C., and Tan S., "Isolated Sign Language Recognition using Convolutional Neural Network Hand Modeling and Hand Energy Image," *Multimedia Tools and Applications*, vol. 78, no. 14, pp. 19917-19944, 2019.

[11] Lin W. and Hsieh C., "Kernel-Based Representation for 2d/3d Motion Trajectory Retrieval and Classification," *Pattern Recognition*, vol. 46, pp. 662- 670, 2013.

[12] Mohandes M. and Deriche M., "Arabic Sign Language Recognition by Decisions Fusion using Dempster-Shafer Theory of Evidence," *in Proceedings of Computing, Communications and IT Applications Conference*, Hong Kong, pp. 90-94, 2013.

[13] Noubigh Z. and Kherallah M., "A Survey on Handwriting Recognition Based on the Trajectory Recovery Technique," *in Proceedings of 1st International Workshop on Arabic Script Analysis and Recognition*, Nancy, pp. 69-73, 2017.

[14] Sidig A. and Mahmoud S., "Trajectory based Arabic Sign Language Recognition," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 4, pp. 283-291, 2018.

[15] Singh K., David N., Hsiao C., Jacob M., Patel K., and Magerko B., "Recognizing Actions in Motion Trajectories Using Deep Neural Networks," *in Proceedings of AAAI Conferenceon Artificial Intelligence and Interactive Digital Entertainment*, pp. 211-217, 2016.

[16] Teow M., "Understanding Convolutional Neural Networks using A Minimal Model for Handwritten Digit Recognition," *in Proceedings IEEE 2nd International Conference on Automatic Control and Intelligent Systems*, Kota Kinabalu, pp. 167-172, 2017.

[17] Tukhtaev S. and Whangbo T., "A Combined Method of Skin-and Depth-based Hand Gesture Recognition," *The International Arab Journal of Information Technology*, vol. 17, no. 1, pp. 99-134, 2020.

[18] Von-Agris U. and Kraiss K., "Towards a Video Corpus for Signer-Independent Continuous Sign Language Recognition," *in Proceedings of International Workshop on Gesture in Human-Computer Interaction and Simulation*, Lisbon, pp. 10-11, 2007.

[19] Wang H., Stefan A., Moradi S., Athitsos V., Neidle C., and Kamangar F., "A System for Large Vocabulary Sign Search," *in Proceedings of European Conference on Computer Vision*, Berlin, pp. 342-353, 2012.

[20] Zhang Q., Wu Y., and Zhu S., "Interpretable Convolutional Neural Networks," *in Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, pp. 8827-8836, 2018.

Sana Fakhfakh She obtained the Master in computer science and multimedia from Institute of Computer Science and Multimedia of Sfax (ISIMS) in 2013 and Ph.D. degree in computer systems engineering in February 2020 from the National Engineers School of Sfax (ENIS). His research interests include signal processing, image and video processing, and machine learning.

Yousra Ben Jemaa She obtained the engineering degree from Tunisia Polytechnic School in 1997 and the Master in signal processing from SUPELEC in France in 1998 She received the Ph.D degree in Electrical Engineering in February 2003 and the HDR in Telecommunications in June 2012 from the National Engineers School of Tunis (ENIT). She joined National Engineers School of Sfax (ENIS) in 2000 where she is actually a full professor.