A Deep Learning Approach for Hybrid Hand Gesture Recognition

Diego G. Alonso(✉), Alfredo Teyseyre, Luis Berdun, and Silvia Schiaffino

ISISTAN (CONICET-UNCPBA), Campus Universitario, Tandil, Argentina
{diego.alonso,alfredo.teyseyre,luis.berdun,
silvia.schiaffino}@isistan.unicen.edu.ar

Abstract. Emerging depth sensors and new interaction paradigms enable to create more immersive and intuitive Natural User Interfaces by recognizing body gestures. One of the vision-based devices that has received plenty of attention is the Leap Motion Controller (LMC). This device models the 3D position of hands and fingers and provides more than 50 features such as palm center and fingertips. In spite of the fact that the LMC provides such useful information of the hands, developers still have to deal with the hand gesture recognition problem. For this reason, several researchers approached this problem using well-known machine learning techniques used for gesture recognition such as SVM for static gestures and DTW for dynamic gestures. At this point, we propose an approach that applies a resampling technique based on fast Fourier Transform algorithm to feed a CNN+LSTM neural network in order to identify both static and dynamic gestures. As far as our knowledge, there is no full dataset based on the LMC that includes both types of gestures. Therefore, we also introduce the Hybrid Hand Gesture Recognition database, which consists of a large set of gestures generated with the LMC, including both type of gestures with different temporal sizes. Experimental results showed the robustness of our approach in terms of recognizing both type of gestures. Moreover, our approach outperforms other well-known algorithms of the gesture recognition field.

Keywords: Hand Gesture Recognition · Deep Learning · Natural User Interfaces · Leap Motion Controller

1 Introduction

During the last decade, emerging technologies and new interaction paradigms motivated the creation of several commercial depth sensors such as Kinect, Real Sense, and Xtion. These devices enable the recognition of body gestures, which strongly enhances Human-Computer Interaction (HCI). In particular, the ability of understanding hand gestures has become a crucial component for the development of more immersive and intuitive Natural User Interfaces (NUIs) in many domains including virtual reality gaming, physical rehabilitation, and sign language recognition. At this point, one of the vision-based devices that has received

L. Martínez-Villaseñor et al. (Eds.): MICAI 2019, LNAI 11835, pp. 87–99, 2019.
https://doi.org/10.1007/978-3-030-33749-0_8

plenty of attention is the Leap Motion Controller (LMC), which captures the movements of human hands creating a virtual representation of them.

The LMC is a small sensor consisting of two monochromatic cameras and three infrared LEDs. It models the 3D position of hands and fingers and provides more than 50 features such as palm center, fingertips, finger bones, and hand orientation. Nevertheless, its Application Programming Interface (API) only provides support for recognizing three gestures. For this reason, in spite of the fact that the LMC provides such useful information, developers still have to deal with the tedious and difficult task of recognizing gestures. Thus, several researchers approached this issue of increasing the LMC gesture recognition capabilities by using machine learning techniques. These hand gesture recognition approaches can be grouped into two main categories: static and dynamic [3]. In brief, static hand gesture recognition considers single frames of data and dynamic hand gesture recognition considers sequences of frames of data.

For dynamic hand gesture recognition most commonly used techniques are Dynamic Time Warping (DTW) and Hidden Markov Models (HMM), while for static hand gesture recognition the conventional method is Support Vector Machines (SVM) [2]. However, the studies in the current state-of-art do not addressed the hybrid hand gesture recognition problem, which considers both static and dynamic gestures together. In addition, as far as our knowledge, there is no dataset generated with the LMC that contains both static gestures and dynamic gestures.

In this context, this paper presents a robust approach for hybrid hand gesture recognition with LMC using deep learning. We propose a neural network architecture that involves using a Convolutional Neural Network (CNN) for extracting features from the 3D data provided by the LMC, combined with a Long Short-Term Memory (LSTM) recurrent network for supporting the temporal dimension of the 3D data. It is important to clarify that, we consider real time scenarios in which the sequences of each gesture have not necessarily the same size. For this purpose, before feeding our CNN+LSTM model we augment the 3D data provided by the LMC with a resampling technique based on Fast Fourier Transform (FFT) algorithm. Thus, our approach is able to classify 3D data sequences of different sizes. It is noteworthy that, we also applied a centering transformation in order to make the sequences suitable for comparison. In short, we move the hand joints regarding the palm position so as to avoid heterogeneous hand part locations within the LMC detection field.

To evaluate our approach, we generated a balanced dataset consisting of 20 gestures including both static and dynamic types, and we performed a cross-validation scheme. Experimental results showed the robustness of the approach in terms of recognizing both types of gestures. Furthermore, our approach achieved higher accuracy rates than other well-known algorithms proposed in the current state-of-art for hand gesture recognition.

The remainder of this paper is organized as follows. Section 2 describes relevant related works of the current state-of-art. Section 3 introduces our proposed

approach. Section 4 discusses the experiments and results along with the lessons learned. Finally, Sect. 5 presents the conclusions and future work.

2 Related Work

In this section, we review relevant works of the current state-of-art of hand gesture recognition with LMC. In Subsect. 2.1, we focus on the approaches that tackle static hand gesture recognition. In Subsect. 2.2, we analyze the approaches that address dynamic hand gesture recognition.

2.1 Static Hand Gesture Recognition

A static gesture, also known as pose or posture, refers to only a single image or a single frame of data corresponding to one command such as stop [14]. However, we adapt this definition to the context of real-time HCI scenarios. In these scenarios, the devices establish a stream where the input data is provided continuously. Under this condition, we determine that a static gesture is one that given a sequence of frames could be identified by individually taking any frame of that sequence. Chuan et al. [4], recognized 26 letters of the American Sign Language (ASL), which were single-framed gestures, using a K-Nearest Neighbor (KNN) classifier and a Support Vector Machine (SVM) classifier. In particular, they extracted 9 features from the raw data of the LMC and performed a 4-fold cross-validation test obtaining an average recognition rate of 72.78% and 79.83% by KNN and SVM respectively. Similarly, Mohandes et al. [12] identified 28 letters of the Arabic Sign Language (ArSL). Their approach preprocessed the raw data and extracted 12 features in order to feed a Multilayer Perceptron neural network (MLP) and a Naive Bayes Classifier (NBC). The authors reported an average recognition accuracy of 98.3% for the NBC and 99.1% for the MLP technique in a 5-fold cross-validation test. Khelil and Amiri [6] trained a SVM for recognizing static gestures of the ArSL too, achieving a recognition accuracy of 91.3%.

Other approaches proposed using multiple sensors working together. For example, Marin et al. [9] analyzed the use of LMC and Microsoft Kinect, both individually and in joint, for 10 static gestures of the ASL. In particular, they trained a SVM classifier using three feature vectors for the LMC and two for the Kinect. Experimental results showed that better recognition accuracy is achieved by combining data from both devices. Later, the authors extended their work by proposing new features that improved the recognition accuracy [10]. Fok et al. [5] combined two LMCs for the recognition of 10 digits of the ASL. The authors proposed a Hidden Markov Model (HMM) of 9 states, which were empirically determined. Experimental results evidenced that the multi-sensor approach (accuracy: 93.14%) can deliver higher recognition accuracy than a single-sensor system (accuracy: 87.39%).

2.2 Dynamic Hand Gesture Recognition

Unlike previous works, in which the authors experimented with static gestures, other researchers evaluated their approaches analyzing dynamic gestures. A dynamic gesture, also known as move, is a temporal and spatial sequence of postures [15]. López (et al.) [7] introduced an approach for manipulating augmented objects based on a SVM classifier, which was trained with only a set of features provided by the LMC. Their approach was tested with 8 dynamic gestures performed by 12 participants and achieved an average recognition accuracy of roughly 80%. Chen et al. [1] took the projection of the palm position directly obtained from the LMC and trained a SVM classifier and a HMM. The authors built a dataset of 3600 samples for 36 dynamic gestures of the ArSL. The results showed that SVM outperformed HMM in recognition accuracy and recognition time. In addition, the authors obtained encouraging results for early gesture recognition. McCartney et al. [11] processed a sequence of frames with positional data and summarize it to a shorter sequence of lines and curves. With these new sequences they fed a HMM in order to detect start and stop points for dynamic gestures. A dataset of 9600 samples of 12 dynamic gestures was performed for the experiments. The obtained results with the introduced method were not as well as expected. However, the authors also proposed an approach based on image processing that consists on training CNNs. They achieved an accuracy rate of 92.4% using only samples corresponding to 5 of the 12 gestures. Lu et al. [8] fed a Hidden Conditional Neural Field (HCNF) classifier with single finger and double finger features extracted from the LMC. The authors generated two different datasets with dynamic gestures performed by 10 participants. The experimental results showed that the proposed approach achieved an average recognition accuracy of 95% for one dataset and 89.5% for the other.

This analysis of the current state-of-art provides evidence that these approaches, both those for static gestures and those for dynamic gestures, allow developers to create different ways of interacting through the LMC. Nevertheless, to the best of our knowledge, no approach considers simultaneously recognizing both types of hand gestures. In addition, although neural networks have been used for gesture recognition, no approach considers using a CNN+LSTM architecture for hand gesture recognition with the LMC.

3 Our Approach for Hybrid Hand Gesture Recognition

In general terms, we propose a 3-step pipeline that consists in Data Acquisition, Data Preprocessing, and Gesture Classification based on a Deep Learning Model. As it is shown in Fig. 1, once a user performs a gesture in the interaction field of the LMC the Data Acquisition step begins. In short, the tracking software of the LMC reads the sensor data and combines it with a skeletal model of the human hand. The data is assembled in the form of frames, which contain the information of several hand joints including fingers, palm, and wrist. Thus, a gesture is interpreted as a sequence of frames that describes the overall movement

of the different hand joints over time. The gesture data provided by the LMC is collected by our tool and passed as raw data to the next step of the pipeline.

The Data Preprocessing step aims to make the raw data suitable for comparison. As mentioned before, in real-time scenarios the gestures have different temporal sizes, that is why our approach resamples the sequence of frames to a predefined number of frames by using the Fast Fourier Transform (FFT) algorithm. Afterwards, our approach proceeds to crop some percentage of the beginning and the ending of the resampled sequence to avoid noisy data corresponding to the positions of the hand both before performing the gesture and after completing the gesture. This is an important step, particularly, when analyzing an stream with several gestures. Another crucial preprocessing transformation is centering, which consists in moving the trajectories of the hand joints regarding the palm position. This is applied because heterogeneous joint locations within the LMC interaction field negatively affect the recognition accuracy.

The gesture representation delivered by the Data Preprocessing step is used to feed a neural network classifier. In brief, the network architecture of our approach consists in a Convolutional layer for extracting meaningful features of the frames, a Long Short-Term Memory layer for modelling the temporal dimension of the data, and two fully connected layers for making easier the classification task. This architecture is also known as CLDNN [13]. Finally, our approach indicates the predicted gesture.

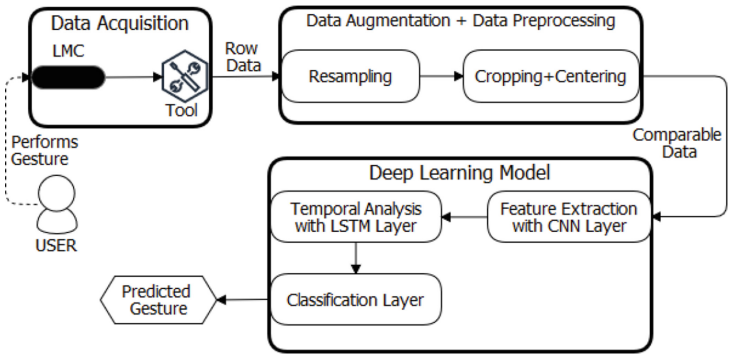


Fig. 1. Our Deep Learning approach for Hybrid Hand Gesture Recognition with LMC.

In the next subsections, we delve into each step of the proposed pipeline. In Subsect. 3.1, we describe the Data Acquisition with LMC and its features. In Subsect. 3.2, we explain the transformations applied in the Data Preprocessing step. In Subsect. 3.3, we introduce our Deep Learning Model architecture.

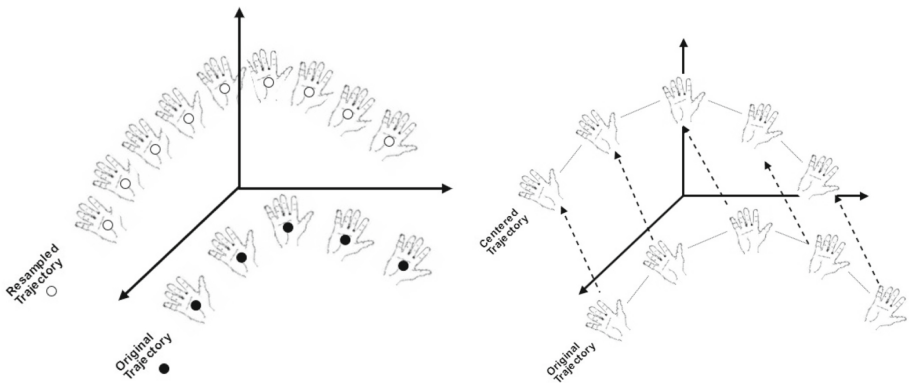
3.1 Data Acquisition with Leap Motion Controller

As mentioned before, the LMC is a vision-based device that consists of two cameras and three infrared LEDs. Among the vision-based approaches for data acqui-

sition, it is considered an active technique since it uses the projection of structured light [3]. The LMC interaction area is like an inverted pyramid because of the intersection of the binocular cameras' fields of view and its viewing range is about 2.6 ft. In short, the LMC's USB controller reads the sensor data into its own local memory, performs resolution adjustments, and streams it to the tracking software. Afterwards, the LMC applies advanced algorithms to the raw sensor data in combination with a model of the human hand for inferring the 3D positions and directions of the different joints. Some of these joints are the fingertips, the finger bones (distal, intermediate, proximal, metacarpal), the palm, the wrist, among others. For more details about the LMC see here[1].

3.2 Data Preprocessing

At first, considering the variations in the temporal sizes of the gestures, our approach resamples the gesture data. At this point, it uses a well-known resampling technique based on the FFT algorithm, which efficiently computes the Discrete Fourier Transform (DFT) of a sequence. Figure 2a shows a handwaving gesture consisting of five frames being resampled, in this case, up-sampled to nine frames. Then, our approach crops a percentage of the beginning and the ending of the resampled gesture. We empirically determined cropping 10% of the beginning and 10% of the ending of the gesture. Moreover, our approach centers the trajectories of the different hand joints regarding the palm position as it is shown in Fig. 2b. As it was mentioned above, this preprocessing transformation significantly improves the recognition accuracy of the approach, since it avoids the variations of the joints locations within the LMC interaction area.



(a) Resampling of a handwaving gesture (b) Centering of a handwaving gesture

Fig. 2. Data Preprocessing applied techniques.

[1] http://blog.leapmotion.com/hardware-to-software-how-does-the-leap-motion-controller-work/.

3.3 Deep Learning Architecture

Once the comparable data is available, our approach delivers it to a neural network classifier based on a CNN+LSTM architecture (see Fig. 3). First, the input data is divided in a predefined number of steps and passed to a Convolutional 2D layer that works as a feature extractor. Shortly, this layer applies a specified number of filters that scan the input by subregions and perform a set of mathematical operations, also known as convolutions, to generate convolved feature maps. After performing several tests we decided to use 64 filters of 2×2 size. This layer intends to learn complex feature representations from raw data**.** Since our input data are 1D signals, our approach transforms them to 2D virtual images. Then, a pooling layer summarizes the outputs of neighboring groups of neurons in the same kernel map reducing its dimensionality by performing a basic downsampling operation along the spatial dimensions. Since CNNs learn so fast, our approach has a Dropout layer (in our case of 0.3) that aims to improve the learning process by slowing it down. In order to pass the CNN feature outputs to the LSTM, our approach also has a Flatten layer that represents the features as one long vector. LSTM is a type of Recurrent Neural Networks (RNNs), which is specifically designed to maintain long-term dependencies. The LSTM (in our case of 256 units) can model the temporal structure of the input data being crucial for sequential prediction. Finally, our approach has two fully connected layers. The first one (in our case of 128 neurons) for reducing variation in the hidden states and the second one for the classifying task (in our case of 20 classes). It is important to note that, for the experiments, the batch size was set to 32 and the number of epochs was set to 30.

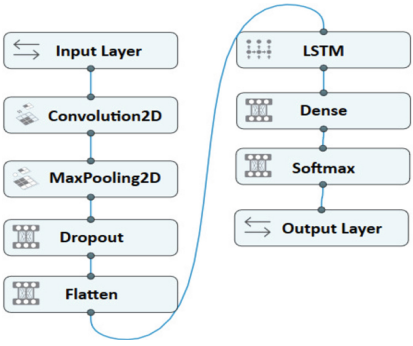


Fig. 3. Deep Learning Model Architecture.

4 Experiments

As mentioned before, to the best of our knowledge, no study considers simultaneously recognizing static and dynamic hand gestures. Therefore, we generated

a balanced dataset with the LMC containing both static and dynamic hand gestures. In order to validate our proposal, we compared its recognition accuracy with the recognition accuracy of three well-known algorithms of the hand gesture recognition field. In Subsect. 4.1, we describe the generated dataset and the selected techniques for comparison. In Subsect. 4.2, we present the experimental results and discussion.

4.1 Experimental Setup

For generating the dataset, we implemented a prototype tool that allows us to collect data directly from the LMC. In addition, this tool also has support for incorporating new recognition techniques and preprocessing capabilities. Thus, we created our dataset by using this tool. In short, as a trainer places his/her hand in the interaction area of the LMC and performs a gesture, our tool stores the movements as a sequence of frames by using the LMC SDK.

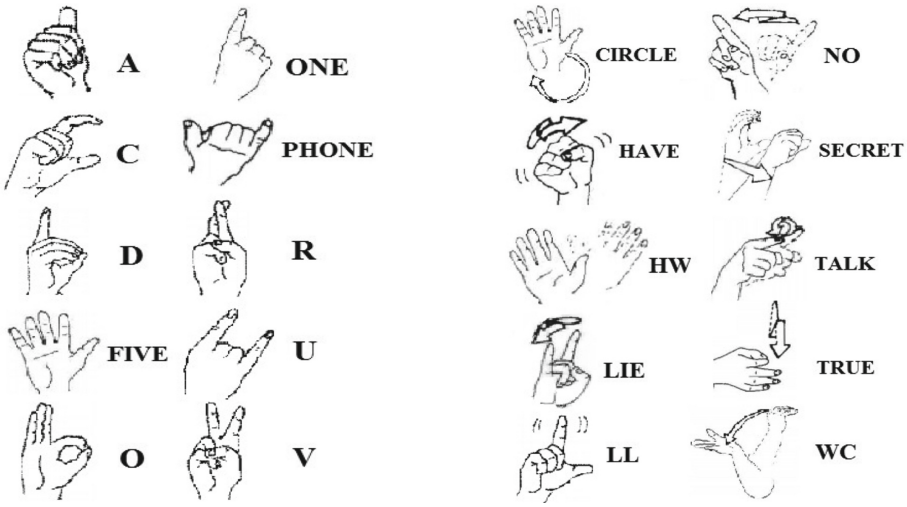
In order to evaluate the robustness of our approach, we collected samples for 10 static hand gestures (see Fig. 4a) and 10 dynamic hand gestures (see Fig. 4b). In particular, we recorded 30 samples for each gesture, thus, generating a dataset of 600 hand gesture samples. It should be noted that, both static and dynamic hand gestures vary in a range between 6 and 181 frames. This means that the samples do not have the same temporal sizes and they are not of a single frame as typically static gestures are described. As mentioned before, we decided to record the samples this way because in real-time scenarios, the hand gestures, even the poses, are not of a single frame, since it is not possible to maintain the hand completely immobile (considering that, for example, the LMC provides 30 frames per second). It is important to note that the error rate of the camera causes variations between frames too. Therefore, is really important to consider more than a single frame for a static hand gesture when working with real-time scenarios. The dataset is available online here[2].

For validating our approach we compared its recognition accuracy with some well-known algorithms of the hand gesture recognition field. Particularly, we analyzed the recognition accuracy of the following techniques: (a) our CNN+LSTM approach; (b) Support Vector Machines (SVM); (c) Dynamic Time Warping (DTW); and Human Activity Recognition CNN (HAR-CNN). For the experiments we performed a 10-fold cross-validation scheme.

4.2 Experimental Results and Discussion

Our approach reached the higher average recognition accuracy both for static gestures and for dynamic gestures. Hence, our CNN+LSTM approach obtained the higher overall recognition accuracy which value was 0.9883. These results demonstrated the robustness of our approach considering both types of gestures in real-time scenarios. Another technique that obtained a high average recognition accuracy was the HAR-CNN, which reached 0.98. Figure 5 shows

[2] http://si.isistan.unicen.edu.ar/bemyvoice/datasetdownload.html.



(a) Static hand gestures (b) Dynamic hand gestures

Fig. 4. List of gestures of our hybrid hand gesture recognition dataset.

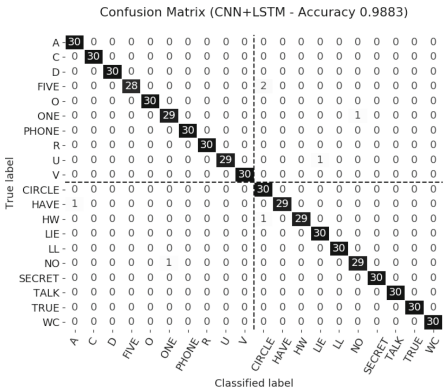
the confusion matrices generated after performing the cross-validation test on all the evaluated techniques.

By analyzing the confusion matrices, we identified that our approach confused 7 samples with very similar gestures. In particular, it confused: two samples of the 'FIVE' gesture and one sample of the 'HANDWAVING' gesture with the 'CIRCLE' gesture, which the three of them consist in moving the hand with the five fingers extended; one sample of the 'U' gesture with the 'LIE' gesture, which basically consist in shaking the hand with the same shape of the 'U' gesture; one sample of the 'NO' gesture with the 'ONE' gesture, in which the former gesture consists in shaking the hand with the shape of the latter; and one sample of the 'HAVE' gesture with the 'A' gesture whose hand shapes are very similar. It should be noted that, these samples were confused by the other techniques as well. The HAR-CNN approach confused 12 samples, among which we identified three non-similar classification mistakes: a sample of the 'SECRET' gesture with the 'HANDWAVING' gesture; a sample of the 'R' gesture with the 'HANDWAVING' gesture; and one sample of the 'TRUE' gesture with the 'HAVE' gesture. The DTW, which is a well-known technique of the dynamic hand gesture recognition field, reached an overall recognition accuracy of 0.8767. Regarding the SVM, which is a frequently used technique in the static hand gesture recognition field, reached an overall recognition accuracy of 0.8333, being the lowest value achieved of the evaluated techniques. It is important to note that, we evaluated the SVM technique varying the number of clusters but the best result was obtained by setting it with one cluster.

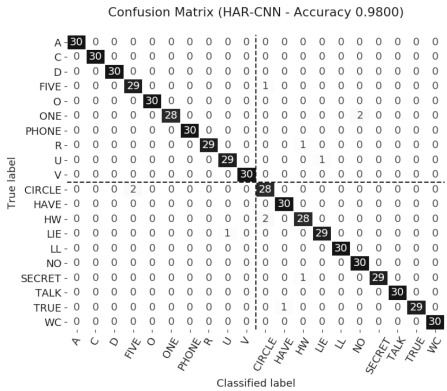
In order to be able to perform a deeper analysis, we reduced the confusion matrices to the misclassification rates of static and dynamic labels. In particular,

we desired to observe what percentage of the misclassified samples was confused with the same type of gesture (e.g. static gestures confused with other static gestures) and what percentage was confused with the other type of gesture (e.g. static gestures confused with dynamic gestures). Thus, we generated a 2×2 matrix that shows these percentages (see Fig. 6).

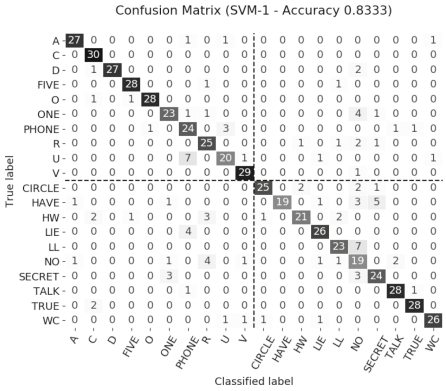
At this point, we identified that CNN+LSTM and HAR-CNN approaches did not confuse static gestures with other static gestures. In addition, we noted that the HAR-CNN recognized static gestures with a higher accuracy than dynamic gestures, while our CNN+LSTM approach obtained a higher accuracy recognizing dynamic gestures than static ones. Although both techniques achieved high precision values, this observation allows us to show the contribution made by the LSTM layer that, in short, analyzes the temporal evolution of the features extracted by the CNN, thus improving the analysis of dynamic gestures.



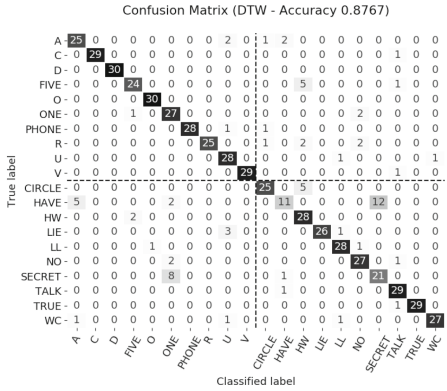
(a) Our CNN+LSTM



(b) HAR-CNN



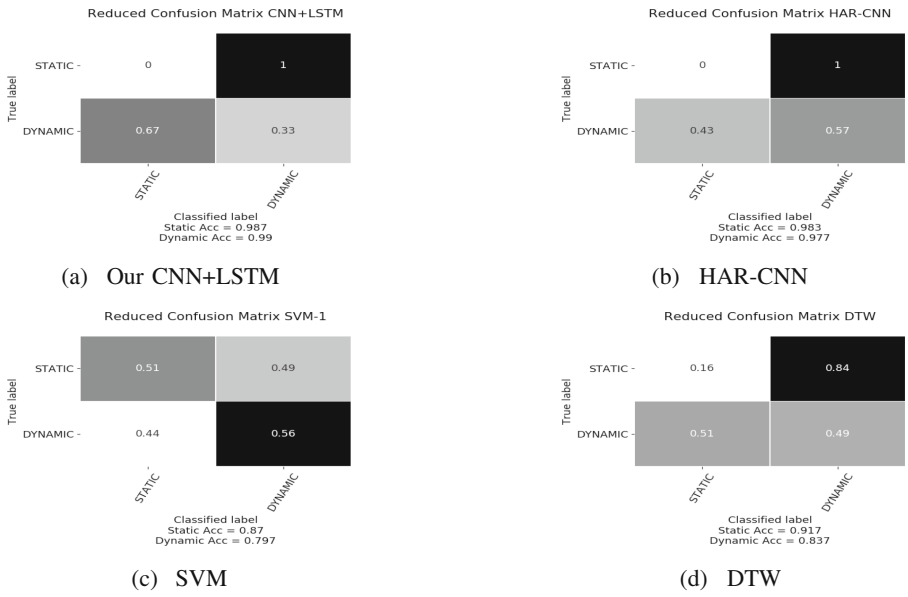
(c) SVM



(d) DTW

Fig. 5. Confusion matrices of all evaluated techniques.

The SVM technique reached a higher recognition accuracy when classifying static gestures than classifying dynamic ones. This decrease in the recognition accuracy can be understood due to the nature of the SVM technique, which looks for the hyperplane that best separates the classes. Therefore, the SVM benefits when the points of the trajectories are concentrated, which is the case of static gestures, and not when the points are dispersed in a non-linear or non-parametric way, which can be the case of dynamic gestures. In contrast to expectations, the DTW algorithm reached higher accuracy recognizing static gestures than dynamic ones. However, we identified similar behaviors to our CNN+LSTM approach regarding the distribution of misclassified samples.



(a) Our CNN+LSTM (b) HAR-CNN

(c) SVM (d) DTW

Fig. 6. Reduced confusion matrices of all evaluated techniques.

5 Conclusions and Future Work

In this paper, we have presented a deep learning approach for hybrid hand gesture recognition. The main contribution of this work is a robust approach based on a CNN+LSTM architecture that enables to simultaneously classify both static and dynamic gestures. Moreover, we proposed a data preprocessing step before feeding the neural network that allows us to compare gestures of varied temporal sizes, which is a valuable factor considering real-time scenarios. Furthermore, for the experiments, we generated a hybrid hand gesture dataset with the LMC. This dataset is easy to reproduce with the LMC SDK and we published it online

to contribute to this topic. Other contribution is the comparison of our approach, regarding the recognition accuracy, with other well-known techniques of the gesture recognition field such as SVM and DTW.

The encouraging results obtained suggest that static and dynamic hand gestures, even of different temporal sizes, can be recognized with the same approach. At this point, one of our future works is using this approach for addressing the continuous hand gesture recognition problem. In short, it consists in segmenting a stream containing several hand gestures and classifying them. Moreover, it would be interesting to analyze the performance of the approach in low processing capabilities scenarios such as mobile.

Acknowledgements. This work has been supported by CONICET PIP No. 112-201501-00030CO (2015–2017). In addition, we gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan GPU used for this research.

References

1. Chen, Y., Ding, Z., Chen, Y.L., Wu, X.: Rapid recognition of dynamic hand gestures using leap motion. In: 2015 IEEE International Conference on Information and Automation, pp. 1419–1424. IEEE (2015)
2. Cheng, H., Yang, L., Liu, Z.: Survey on 3D hand gesture recognition. IEEE Trans. Circuits Syst. Video Technol. **26**(9), 1659–1673 (2015)
3. Cheok, M.J., Omar, Z., Jaward, M.H.: A review of hand gesture and sign language recognition techniques. Int. J. Mach. Learn. Cybernet. **10**(1), 131–153 (2019)
4. Chuan, C.H., Regina, E., Guardino, C.: American sign language recognition using leap motion sensor. In: 2014 13th International Conference on Machine Learning and Applications (ICMLA), pp. 541–544. IEEE (2014)
5. Fok, K.Y., Ganganath, N., Cheng, C.T., Chi, K.T.: A real-time ASL recognition system using leap motion sensors. In: 2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), pp. 411–414. IEEE (2015)
6. Khelil, B., Amiri, H.: Hand gesture recognition using leap motion controller for recognition of Arabic sign language. In: 3rd International conference ACECS 2016 (2016)
7. López, G., Quesada, L., Guerrero, L.A.: A gesture-based interaction approach for manipulating augmented objects using leap motion. In: Cleland, I., Guerrero, L., Bravo, J. (eds.) IWAAL 2015. LNCS, vol. 9455, pp. 231–243. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-26410-3_22
8. Lu, W., Tong, Z., Chu, J.: Dynamic hand gesture recognition with leap motion controller. IEEE Signal Process. Lett. **23**(9), 1188–1192 (2016)
9. Marin, G., Dominio, F., Zanuttigh, P.: Hand gesture recognition with leap motion and kinect devices. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 1565–1569. IEEE (2014)
10. Marin, G., Dominio, F., Zanuttigh, P.: Hand gesture recognition with jointly calibrated leap motion and depth sensor. Multimed. Tools Appl. **75**(22), 14991–15015 (2016)
11. McCartney, R., Yuan, J., Bischof, H.P.: Gesture recognition with the leap motion controller. In: Proceedings of the International Conference on Image Processing,

Computer Vision, and Pattern Recognition (IPCV). p. 3. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp) (2015)
12. Mohandes, M., Aliyu, S., Deriche, M.: Arabic sign language recognition using the leap motion controller. In: 2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE), pp. 960–965. IEEE (2014)
13. Sainath, T.N., Vinyals, O., Senior, A., Sak, H.: Convolutional, long short-term memory, fully connected deep neural networks. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (2015)
14. Sarkar, A.R., Sanyal, G., Majumder, S.: Hand gesture recognition systems: a survey. Int. J. Comput. Appl. **71**(15), 25–37 (2013)
15. Wang, Q., Xu, Y., Chen, Y.L., Wang, Y., Wu, X.: Dynamic hand gesture early recognition based on hidden semi-markov models. In: 2014 IEEE International Conference on Robotics and Biomimetics. IEEE (2014)