Data Article

ArASL: Arabic Alphabets Sign Language Dataset

Ghazanfar Latif a, b, *, Nazeeruddin Mohammad a, Jaafar Alghazo a, Roaa AlKhalaf a, Rawan AlKhalaf a

a College of Computer Engineering and Sciences, Prince Mohammad Bin Fahd University, Al Khobar, Saudi Arabia
b Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Kota Samarahan, Malaysia

Article info

Article history:
Received 19 November 2018
Accepted 14 February 2019
Available online 23 February 2019

Abstract

A fully-labelled dataset of Arabic Sign Language (ArSL) images is developed for research related to sign language recognition. The dataset will provide researcher the opportunity to investigate and develop automated systems for the deaf and hard of hearing people using machine learning, computer vision and deep learning algorithms. The contribution is a large fully-labelled dataset for Arabic Sign Language (ArSL) which is made publically available and free for all researchers. The dataset which is named ArSL2018 consists of 54,049 images for the 32 Arabic sign language sign and alphabets collected from 40 participants in different age groups. Different dimensions and different variations were present in images which can be cleared using pre-processing techniques to remove noise, center the image, etc. The dataset is made available publicly at https://data.mendeley.com/datasets/y7pckrw6z2/1.

© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Data

The ArSL2018 is a new comprehensive fully labelled dataset of Arabic Sign Language images launched in Prince Mohammad Bin Fahd University, Al Khobar, Saudi Arabia to be made available for

* Corresponding author. College of Computer Engineering and Sciences, Prince Mohammad Bin Fahd University, Al Khobar, Saudi Arabia.
E-mail address: glatif@pmu.edu.sa (G. Latif).
researchers in the field of Machine Learning and Deep Learning. It is useful for application and device development in the assistive technology field for the benefit of the deaf and hard of hearing individuals. Examples of related datasets can be found in Refs.[3–5]. The ArSL2018 dataset is unique in the sense that it is the first large comprehensive dataset for Arabic Sign Language according to the author(s) knowledge. There is a large potential for this dataset to be used by researchers to both increase accuracies of classification and recognition and for development of prototypes useful for the deaf community.

The ArSL2018 dataset is compiled of 54,049 images in gray scale with 64 × 64 dimension. Variations of images were introduced with different lighting and different background. Fig. 1 shows a sample of the pictures of the Arabic Sign Language signs and alphabets in the dataset. In order to assist researchers to access the ArSL2018 dataset for classification and recognition, we have collected, labelled, generated and published the ArSL2018 dataset [1]. Table 1 shows the classification of the Arabic Alphabet signs, with labels and number of images. The dataset has been identified to be sufficient for both training and classification, and has been tested as such. The dataset can be used as is and maybe increased with more variations in the second version of the dataset.

There are still some limitations to the ArSL2018 dataset which include, 1) dataset was collected in one location, 2) not enough lighting and noise variations were introduced, 3) the number of participants providing samples were only 40 participants. The limitations are minor and could be addressed in the second version of the dataset.
| Arabic Letter | Representation |
|--------------|----------------|
| ألف          | "Alif"        |
| باء          | "Baa"         |
| تاء          | "Taa"         |
| ذال          | "Dal"         |
| راء          | "Ra"          |
| زاي          | "Zay"         |
| سين          | "Sin"         |
| شين          | "Shin"        |
| ضناء         | "Sad"         |
| طاء          | "Taa"         |
| زا           | "Zaa"         |
| أين          | "Ayn"         |
| غين          | "Ghain"       |
| فا            | "Faa"         |
| كاف          | "Kaf"         |
| لام           | "Lam"         |
| ميم           | "Mim"         |
| نون          | "Nun"         |
| هاء          | "Haa"         |
| وو           | "Waw"         |
| ياء          | "Yaa"         |
| دال          | "Dal"         |
| خاء          | "Haa"         |
| خاء          | "Haa"         |
| خاء          | "Haa"         |

*Fig. 1. Representation of the Arabic Sign Language for Arabic Alphabets.*
2. Experimental design, materials, and methods

The ArSL2018 dataset images were taken at Prince Mohammad Bin Fahd University and in the Khobar Area, Kingdom of Saudi Arabia from volunteers of different age groups. A smart Camera attached to tripod was used to capture the images. Volunteers were made to stand around 1 m away from the camera. Variations of images were introduced with different lighting, angles, timings and different background. The total number of images per alphabet varies, however, the total number of images compiled for the dataset were 54,049 images. The images were taken in RGB format with a range of pixel values 0 to 255. Converted to grayscale images, with a fixed dimension 64×64 and converted to grayscale images, with a range of pixel values 0 to 255.

Acknowledgments

We would like to thank Prince Mohammad Bin Fahd (PMU) for the support both in logistics and financial that allowed for the creation of this dataset. We would also like to thank the students who participated in taking the pictures as well as the participants who took the time to participate in this study and posing for the pictures.

Transparency document

Transparency document associated with this article can be found in the online version at https://doi.org/10.1016/j.dib.2019.103777.

References

[1] Ghazanfar Latif, Jaafer Alghazo, Nazeeruddin Mohammad, Roaa AlKhalaf, Rawan AlKhalaf, “Arabic Alphabets Sign Language Dataset (ArSL)”, Mendeley Data, v1, 2018. https://doi.org/10.17632/y7pckrw6z2.1.
[2] Roaa AlKhalaf, Rawan AlKhalaf, Ghazanfar Latif, Nazeeruddin Mohammad, Jaafer Alghazo, Arabic Sign Language (ArSL) recognition system using convolutional neural network (CNN), Int. J. Pure Appl. Math. (2018) in press.
[3] Z. Zafrauli, H. Brasejar, T. Starner, H. Hamilton, P. Presti, American sign language recognition with the kinect, in: Proceedings of the 13th International Conference on Multimodal Interfaces, ACM, 2011, November, pp. 279−286.
[4] V. Abhinos, C. Neidle, S. Scarloff, J. Nash, A. Stefan, Q. Yuan, A. Thangali, The American sign language lexicon video dataset, in: Computer Vision and Pattern Recognition Workshops, 2008. CVPRW’08. IEEE Computer Society Conference on (pp. 1-8), IEEE, 2008, June.
[5] Sign Language MNIST, Kaggle, 2017. Source: https://www.kaggle.com/datamunge/sign-language-mnist/.