A Dataset for Irish Sign Language Recognition

Marlon Oliveira*, Houssem Chatbri†, Ylva Ferstl ‡, Mohamed Farouk⋆, Suzanne Little†, Noel E. O'Connor† and Alistair Sutherland*

*School of Computing, Dublin City University, Ireland
†Insight Centre for Data Analytics, Dublin City University, Ireland
‡ADAPT Centre, School of Computer Science and Statistics, Trinity College Dublin, Ireland
⋆College of Computing & Information Technology, Arab Academy for Science & Technology, Egypt

Abstract

We introduce a new image dataset for Irish Sign Language (ISL) recognition. We filmed human subjects performing ISL hand-shapes and movements, resulting in 468 videos. Then, we extracted frames from the videos. This resulted in a total of 58,114 images for the 23 common hand-shapes from the ISL language. This dataset is a part of our ongoing work on ISL recognition using pattern recognition methods. In addition to the dataset, we report experiments using Principal Component Analysis (PCA) where we reached 95% recognition accuracy.

Keywords: Irish Sign Language, Pattern Recognition, Image Dataset

1 Introduction

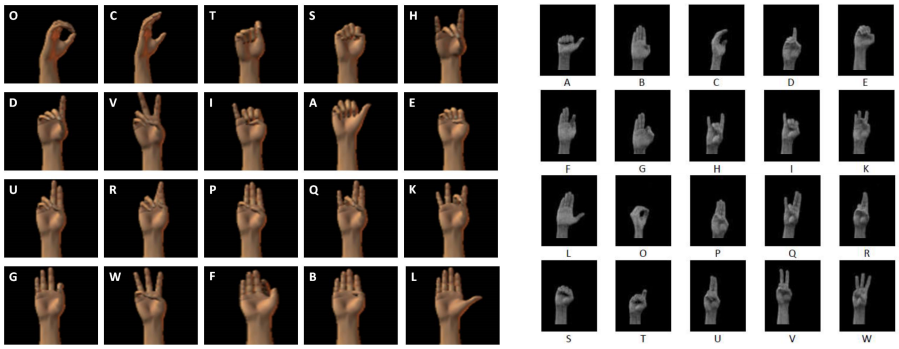
Irish Sign Language (ISL) is an indigenous language that is used by around 5,000 Deaf people in the Republic of Ireland and 1,500 in Northern Ireland. In addition, it is known by 50,000 non-Deaf people [Leeson and Saeed, 2012]. ISL is not based on English or Irish, it is a language in its own right.

ISL contains more than 5000 signs. Each sign consists of a hand-shape and a motion in 3D space. There are around 23 basic, common hand-shapes in ISL and each hand-shape is labelled with a different letter of the alphabet. These hand-shapes can be seen in a wide range of possible angles in 3D space. The remaining three letters of the alphabet, 'J', 'X' and 'Z' are used to label gestures involving motion and actually use one of the 23 hand-shapes.

Computer vision provides the technology to assist people who use ISL with tools such as automatic transcript, human-machine interaction, machine translation, etc. In order to design such tools, large amounts of data are necessary for training and testing the system. In this paper, we introduce a new image dataset for ISL recognition. The dataset contains 58,114 images for the 23 ISL hand-shapes. In addition to the dataset being our main contribution in this paper, we also report recognition experiments using Principal Component Analysis (PCA).

Earlier works in this area have used rather smaller datasets. For instance, Farouk et al. proposed two ISL datasets [Farouk, 2015]. The first dataset is composed of computer generated images, produced by a the Poser software by SmithMicro; the total number of images is 920. The second dataset is composed of real hands, and has a total of 1620 images. Both datasets represent only 20 ISL hand-shapes as illustrated in Figure 1 (excluding 'm', 'n' and 'y and the dynamic shapes 'J', 'X' and 'Z'). The images show the hand and arm of a signer against a uniform black background.

Compared to previous works on ISL, our dataset is larger and contains all hand-shapes. It is then fit to train and test classifiers for ISL recognition. The rest of the paper is organised as follows: Sec. 2 details our data collection procedure and the final dataset. Sec. 3 reports a recognition experiment using PCA. We end the paper in Sec. 4 with concluding remarks and future work.



(a) Computer generated (b) Real hands

Figure 1: Datasets for ISL created by [Farouk, 2015]

2 The Irish Sign Language hand-shape (ISL-HS) dataset

The ISL-HS dataset contains real hand images, unlike synthetic images used in previous works. ISL-HS is composed of 23 hand-shapes combined with different motions.

To build the dataset, we recorded short videos. We asked 6 people (3 males and 3 females) to perform the finger spelling ISL hand-shapes. Each shape was recorded 3 times.

Each of the 23 hand-shaped was performed by moving the arm in an arc from the vertical to the horizontal position. This was performed to simulate rotated hand-shapes that can occur in real word conversations. For the 3 motion gestures 'J', 'X' and 'Z' there was no rotation, only the motion indicated in Figure 2. All the hand-shapes in our dataset, apart from the 3 with motion, are rotated in a plane.

The videos were converted into frames. Frames were converted to grayscale and the background was removed from the frame using a pixel-value threshold. This produced frames contain only the arm and the hand.

The number of frames for each video depends on the time taken by the human subject to perform the gesture. Videos were recorded at 30 frames per second (fps) and a resolution of 640×480 pixels. The device used to record the videos was an Apple iPhone 7. The videos were saved with *.mov* extention. The video format is RGB24.

The illumination sources were a combination of natural and artificial, as the videos were recorded in our laboratory of post-graduate computing students. Illumination was different for each person, because they were recorded at different times of day and on different days.

In total, 468 videos were recorded. From these videos we obtained a total of 58,114 frames, consisting of 52,688 frames for the rotated shapes and 5,426 for the 'J', 'X' and 'Z'. Figure 2 shows cropped images of our ISL-HS dataset, and Figure 3 shows the class distribution across the image dataset. The variation observed in Figure 3 is due to the speed variation among the subjects when performing the ISL hand-shapes and rotating them. Note that the letter '*X*' has the lowest number of frames because this is a dynamic festure with a short motion.

We are releasing the dataset online[1] and providing both videos and images.

3 Principal Component Analysis (PCA)

PCA is an efficient method for dimensionality reduction [Han and Liu, 2014]. It uses the covariance matrix of the data to create a space known as an eigenspace. Each dimension in the space is represented by an eigenvector of the covariance matrix. The number of eigenvectors required to represent the full data is considerably lower than the dimensionality of the original data.

[1] https://github.com/marlondcu/ISL

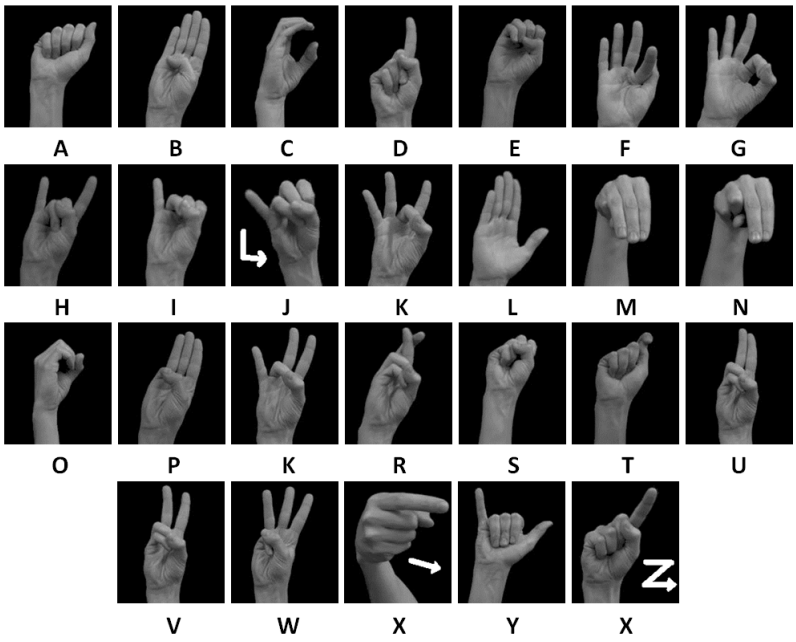


Figure 2: Irish Sign Language hand-shapes

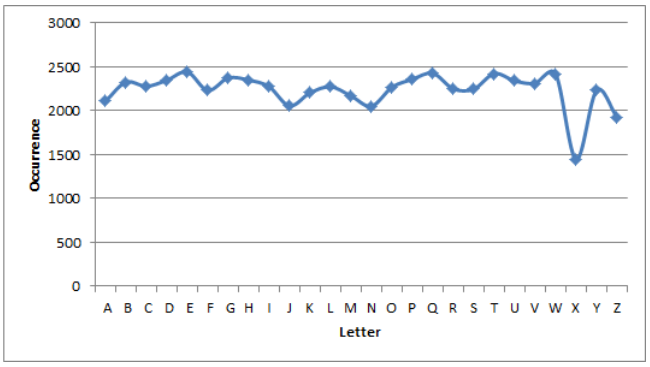


Figure 3: Frequency of the different hand-shapes in the dataset

In order to apply PCA over our training dataset we combine all the images into the same array and then compute PCA. Since each image has 640×480, we re-sized them to 160×120 pixels. When vectorised this becomes 19,200 pixels in a row array, for each image.

In this experiment we considered only the 23 common hand-shapes with rotation. Then images corresponding to the letters 'J', 'X' and 'Z' were not used. The dataset used contains 52.688 in total. This dataset was divided into a training set and a testing set, by iterating through the images and taking one image for training and the next for testing, and so on. Thus, both our training and testing datasets contain 26,344 images.

By projecting the images from the training set into the most significant D_i eigenvectors, we obtain a D_i-dimensional space containing N_{im} points for each pose angle. Each point represents an image. In this work we tested different numbers of eigenvectors and measured how it affects the accuracy.

In order to classify the correct hand-shape we used the k-Nearest Neighbour (k-NN) algorithm, with $k = 1$ and Euclidean distance. We projected each testing image into the training dataset eigenspace and classified according to the nearest point (shortest Euclidean distance).

The accuracy in recognising the correct hand-shape strongly depends on the number of the eigenvectors (dimensions) considered. For example, assuming $D_i = 15$, we obtained 88% of recognition accuracy, using more eigenvectors the accuracy increases as well. e.g. for $D_i = 29$ we obtained 95%. Figure 4 shows the

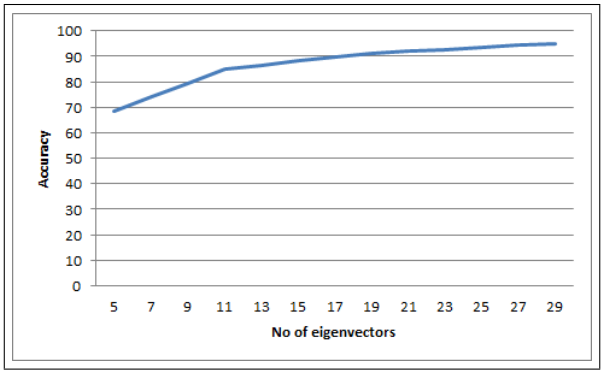


Figure 4: Accuracy according to the number of eigenvectors

accuracy according to the number of eigenvectors.

In this work blurring was applied over the images. A Gaussian kernel of size 36×36 pixels was used with standard deviation equal to 60. Using blurring was motivated by earlier results by Farouk [Farouk et al., 2013], which showed that such image filtering is beneficial for PCA accuracy.

4 Conclusions

In this work, we proposed an Irish Sign Language hand sign dataset (ISL-HS). Compared to previous works, our dataset is larger, more complete and contains rotation variation. In addition, we reported a recognition experiment using PCA, and we were able to reach 95% of recognition accuracy.

In the future, we are planning to try different classification methods in addition to PCA (e.g. Convolutional Neural Networks), and apply recognition to videos in addition to images to leverage the dynamic aspect of some of the ISL hand-shapes.

Acknowledgments

This research was funded by CAPES/Brazilian Science without Borders, process no.: 9064-13-3. The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund. This research also emanated from a grant in part from the IRC under Grant no. GOIPD/2016/61, in part from the EU H2020 Programme under grant agreement no. 688099 (Cloud-LSVA), and in part from SFI under Grant no. SFI/12/RC/2289 (Insight). The authors would like to thank Dr. Robert Smith from the Institute of Technology Blanchardstown for the feedback he gave about Irish Sign Language.

References

[Farouk, 2015] Farouk, M. (2015). *Principal Component Pyramids using Image Blurring for Nonlinearity Reduction in Hand Shape Recognition*. PhD thesis, Dublin City University, Ireland.

[Farouk et al., 2013] Farouk, M., Sutherland, A., and Shokry, A. (2013). Nonlinearity Reduction of Manifolds using Gaussian Blur for Handshape Recognition based on Multi-Dimensional Grids. *ICPRAM*.

[Han and Liu, 2014] Han, F. and Liu, H. (2014). Scale-invariant sparse PCA on high-dimensional meta-elliptical data. *Journal of the American Statistical Association*, 109(505):275–287.

[Leeson and Saeed, 2012] Leeson, L. and Saeed, J. I. (2012). *Irish Sign Language : A Cognitive Linguistic Account*. Edinburgh University Press.