# Hand Orientation Redundancy Filter Applied to Hand-shapes Dataset


Marlon Oliveira
School of Computing,
Dublin City University
Dublin 9, Ireland
marlon.oliveira2@mail.dcu.ie

Houssem Chatbri
Insight Centre for Data Analytics,
Dublin City University
Dublin 9, Ireland
houssem.chatbri@dcu.ie

Naresh Yarlapati
Insight Centre for Data Analytics,
Dublin City University
Dublin 9, Ireland
naresh.yarlapati@insight-centre.org

Noel E. O'Connor
Insight Centre for Data Analytics,
Dublin City University
Dublin 9, Ireland
noel.oconnor@dcu.ie

Alistair Sutherland
School of Computing,
Dublin City University
Dublin 9, Ireland
alistair.sutherland@dcu.ie



## ABSTRACT

We have created a dataset of frames extracted from videos of Irish Sign Language (ISL) for sign language recognition. The dataset was collected by recording human subjects executing ISL hand-shapes and movements. Frames were extracted from the videos producing a total of 52,688 images for the 23 static common hand-shapes. Given that some of the frames were relativity similar we designed a new method for removing redundant frames based on labelling the hand images by using axis of least inertia - Hand Orientation Redundancy Filter (HORF) - and we compare the results with an iterative method - Iterative Redundancy Filter (IRF). This selection process method selects the most different images in order to keep the dataset diverse. The IRF dataset contains 50,000 images whereas the HORF consists of 27,683 images. Finally, we tested two classifiers over the HORF dataset and compared the results with the IRF dataset.



## CCS CONCEPTS

• **Computing methodologies → Machine learning approaches**; **Machine learning algorithms**; **Image processing**; **Machine learning algorithms**; **Image processing**;

## KEYWORDS

hand-shape recognition, axis of least inertia, hand-shapes dataset, hand orientation redundancy filter, machine learning

**ACM Reference Format:**

Marlon Oliveira, Houssem Chatbri, Naresh Yarlapati, Noel E. O'Connor, and Alistair Sutherland. 2019. Hand Orientation Redundancy Filter Applied to Hand-shapes Dataset. In *2nd International Conference on Applications of Intelligent Systems (APPIS 2019), January 7–9, 2019, Las Palmas de Gran Canaria, Spain,* Nicolai Petkov, Nicola Strisciuglio, and Carlos Travieso-Gonzalez (Eds.). ACM, New York, NY, USA, Article 4, 5 pages. https://doi.org/10.1145/3309772.3309794


## 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. It is also known by 50,000 non-deaf people [5].

ISL is not English or Irish, it is a language in its own right. However, the 26 letters/shapes are similar to the English alphabet. ISL contains more than 5,000 signs. Each sign consists of a hand-shape and a motion in 3D space. There are 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.

Despite the recent technological advances, we still do not have a system able to recognise hand gestures in a very efficient way. Such techniques could 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 system training and testing.

Previous works in this field have used smaller datasets. Farouk *et al.* proposed two datasets for Irish Sign Language of relatively limited size [1]. The first dataset contains 920 images generated by a computer, produced by the Poser software by SmithMicro. The second dataset contains 1620 real hand images. Both datasets proposed by Farouk *et al.* represent only 20 out of 26 ISL hand-shapes. The images show the hand and arm of a signer against a uniform black background. As for recognition, it has been achieved using handcrafted features as well as data-driven models.

A new ISL dataset with 58,114 images corresponding to the 23 ISL hand-shapes was introduced in [6]. Later on, a filter to reduce the frames similarity was proposed in [7], which we call the IRF dataset.

In this work we propose a new method for filtering redundant frames, using computation of the axis of least inertia, which we call the Hand Orientation Redundancy Filter (HORF) dataset. In addition to the HORF dataset being our main contribution in this paper, we




*APPIS 2019, January 7–9, 2019, Las Palmas de Gran Canaria, Spain*





ACM ISBN 978-1-4503-6085-2/19/01…$15.00
https://doi.org/10.1145/3309772.3309794

report recognition experiments using principal component analysis (PCA) and classifiers such as support vectors machine (SVM) and k-nearest-neighbour (k-NN).

Reducing this similarity between frames is an important step in pattern recognition. Otherwise, redundant frames may reduce the discriminative patterns and cause the model to overfit irrelevant information [10].

This paper tries to solve the questions: How should we make sure we select good hand-sign images for training and testing? How will the classifiers behave over this new method to select frames? How can it be compared to the previous way of frames selection? In addition, we show that our dataset can be used to successfully train classifiers.

The rest of the paper is organised as follows. Section 2 describes the dataset, Section 3 explains the iterative method for redundant frame filtering - IRF. Our pose angle computation by axis of inertia - HORF - is described in Section 4.1. Section 5 shows the classification stage and results. Finally, conclusions and future work in Section 6.

## 2 IRISH SIGN LANGUAGE DATASET

The dataset used in this work contains images of the Irish Sign Language (ISL). ISL is composed of 23 hand-shapes without motion and 3 motion gestures. Fig. 1 shows images for the 23 static gestures, note that images are cropped for illustration purposes.

In this section, we describe the data collection procedure.

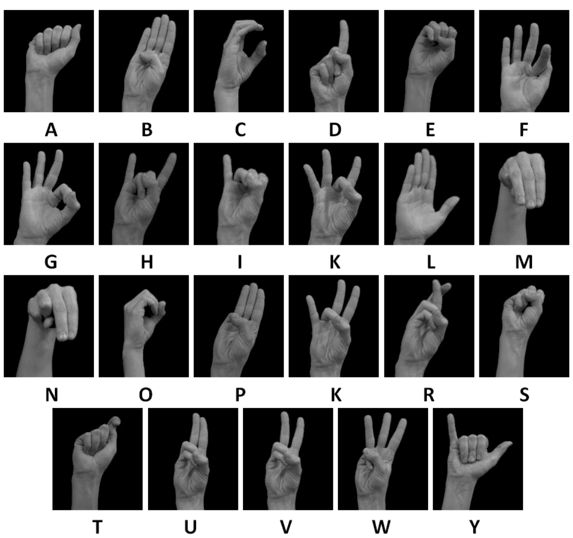


**Figure 1: Irish Sign Language alphabet with static gestures.**

The dataset was collected by recording short videos of 6 people performing the finger spelling ISL hand-shape in a dark background. Each gesture was recorded 3 times. Each hand-shape was performed by moving the arm in a 2D plane rotation in order to make the dataset robust to rotation.

The dataset for Irish Sign Language hand-shapes (ISL-HS) contains real hand images. The entire dataset is composed of 26 hand gestures corresponding to 23 static shapes and 3 dynamic ones (J, X and Z) [4]. Frames were extracted from videos and converted to greyscale in order to remove colour information. In addition, background was removed by using a pixel value threshold.

Each video delivered a different number of frames depending on the duration of the video. Videos were filmed at 30 frames per second and a standard definition (640×480) pixels. The total number of videos was 468 with a total of 52,688 frames for static gestures and 5,426 frames for the dynamic ones. The total number of images was 58,114. Some frames are naturally blurred because of the movement and some of them are very similar, given the fact of the start and stop of the rotation movement.

Each frame in the dataset has originally $640 \times 480$ pixels, we resized them to $160 \times 120$ pixels in order save memory and processing time.

**Algorithm 1** Diversity-based image selection

**Precondition:** Original dataset containing all $N$ frames
A $\leftarrow$ final dataset (empty in the beginning)
$I \leftarrow$ one frame selected randomly from the original dataset
$A \leftarrow I$
**for** $i \leftarrow 1$ to $N$ **do**
    $i_{max} \leftarrow \underset{0 \leq i \leq |D_{orig}|-1}{\text{argmax}} \ \gamma(I, A)$
    $I \leftarrow I_{i_{max}}$
**end for**

## 3 ITERATIVE REDUNDANCY FILTER (IRF)

In order to avoid redundancy in very similar frames, a method to filter frames with very low difference was proposed in [7]. This method is called the Iterative Redundancy Filter (IRF). The filtering method selected frames according to the following: Each image $I_u$ in the original dataset is represented by a feature vector $\overrightarrow{V_u}$. Thus, a *diversity score* is computed for the frame and used to make iterative frame selection from the original dataset in a way to increase the dataset diversity.

In order to create the feature vector $\overrightarrow{V_u}$ frames were split into a $K \times K$ grid and computing the number of pixels in each cell after detecting the edge and applying a threshold equal to 127. Thus, the feature vector $\overrightarrow{V_u}$ for a frame $I_u$ is a 2D histogram. Thus, calculating the dissimilarity between two frames $I_u$ and $I_v$ is done by adding the distances between the histogram bins of their feature vectors $\overrightarrow{V_u}$ and $\overrightarrow{V_v}$. The feature extraction mechanism, as well as the edge detector, is rather simple, since the purpose is to compute a fast distance between pairs of images. The image diversity score expresses how different an image is with respect to a set of images. For an image $I$ with feature vector $\overrightarrow{V_u}$ and a set $A$, the diversity score is calculated according to the following:

$$\gamma(I, A) = \frac{1}{|A|} \sum_{i=0}^{|A|-1} d(\overrightarrow{V}, \overrightarrow{V_i^A}) \tag{1}$$

where $\overrightarrow{V_i^A}$ is the feature vector of an image from the set $A$, and $|A|$ is the number of images in $A$.

Algorithm 1 shows the iterative process. The first step is to select randomly an image from the original dataset and add in the dataset $A$. Thus, the images that have the largest *diversity score* from the original dataset are selected by an iterative process. This approach was inspired by [2].

Because it is a computationally expensive process it was used on a subset of the original dataset with only 100 frames randomly selected instead of the entire original dataset.

Finally, the first 50,000 frames were taken as the final dataset for the IRF dataset. The dataset is divided into a training set and a testing set. Selection were made by random selection, then each of training and testing dataset consist of 25,000 images [7].

# 4 HAND ORIENTATION REDUNDANCY FILTER (HORF)

In this paper we propose another filter to reduce the similarity among the images in the ISL dataset. This new method is called Hand Orientation Redundancy Filter (HORF) and uses the computation of Axis of Least Inertia to identify the pose angle of the arm and the hand.

## 4.1 Computation of Axis of Least Inertia

The axis of least inertia describes the orientation of a shape curve. It is invariant to linear geometric transformations such as rotation, translation and scaling. The axis of least inertia of a shape is defined by the line for which the integral of the square of the distance to the boundary points of the object is minimum [9].

Let $B = \{b(x_i, y_i) | i = 1, 2, ..., N\}$ where $N$ is the number of boundary pixels belonging to the boundary of an object. Let $\alpha$ be the angle between the axis of least inertia and the x-axis. Then, the integral [3] of the square of the distances to boundary points (say $n$ points) on the object is computed as follows:

$$E = \frac{1}{2}(a + c) - \frac{1}{2}(a - c)\cos 2\alpha - \frac{1}{2} b \sin 2\alpha \tag{2}$$

where,

$$a = \sum_{1}^{n} (x_i - \bar{x})^2 \tag{3}$$

$$b = 2 \sum_{1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \tag{4}$$

$$c = \sum_{1}^{n} (y_i - \bar{y})^2 \tag{5}$$

$\bar{x}$ and $\bar{y}$ are the the mean values of the $x$ and $y$ co-ordinates of the boundary respectively. Differentiating the equation 2 with respect to $\alpha$, we get

$$\frac{\mathrm{d}E}{\mathrm{d}\alpha} = (a - c)\sin 2\alpha - b \cos 2\alpha \tag{6}$$

and the second order derivative for the above equation is as follows:

$$\frac{\mathrm{d}^2E}{\mathrm{d}\alpha} = 2(a - c)\cos 2\alpha + 2b \sin 2\alpha \tag{7}$$

Let $\frac{\mathrm{d}E}{\mathrm{d}\alpha} = 0$, then

$$\alpha = \frac{1}{2}\tan^{-1}\left(\frac{b}{a - c}\right), -\frac{\pi}{2} < \alpha < \frac{\pi}{2} \tag{8}$$

The slope angle $\theta$ is given by

$$\theta = \begin{cases} \alpha + \frac{\pi}{2} & if \frac{\mathrm{d}^2E}{\mathrm{d}\alpha} < 0 \\ \alpha & otherwise \end{cases} \tag{9}$$

Figure 2 shows an illustration of axis of least inertia belonging to the hand shape. The axis of least inertia passes through the centroid of the shape.

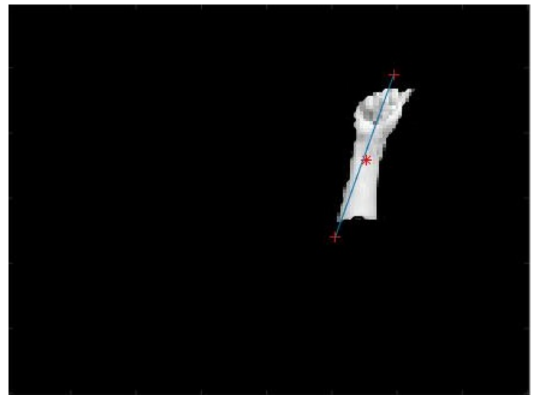

**Figure 2: Illustration of axis of least inertia of hand shape**

By using the axis of least inertia we could compute the pose angle of each hand/arm figure. Then we applied the computation explained in the Section 4.1 and we obtained a angle in radians (between 0 and $2\pi$) for each image.

As expected a certain number of images resulted in the same pose angle. Then we kept only one image with the same angle for each person (1 to 6), each shape (from A to Z) and each shot (1 to 3). We did not consider the shapes with movement such as 'J', 'X' and 'Z'.

Figure 3 show the occurrence by shape before filtering, using the filter proposed - IRF - in [7] and our new filter using pose angle - HORF - as well as the total of images proposed in the original ISL dataset [6].

# 5 CLASSIFICATION AND RESULTS

We split our HORF dataset into two sets: testing and training datasets. These sets are created by a random selection. Each dataset contains half of the HORF. Therefore, training set contains 13,841 images and the testing set contains 13,842 images.

The HORF dataset is available online[1].

The classifiers k-NN and SVM were tested for hand-shape recognition over our HORF dataset. In addition we use Principal Component Analysis (PCA) to extract features and reduce the dimensionality of the images. Classifiers are configured as follows: $k$ in k-NN is set to $k = 1$ and for SVM we used a polynomial kernel (degree 3).

Gaussian blurring with two-dimensions was applied over the images to check how the classifiers behave on blurred images. It was motivated by previous results by Farouk [1], which showed that such image filtering is beneficial for PCA accuracy. In this paper we use a kernel and a standard deviation computed according to

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

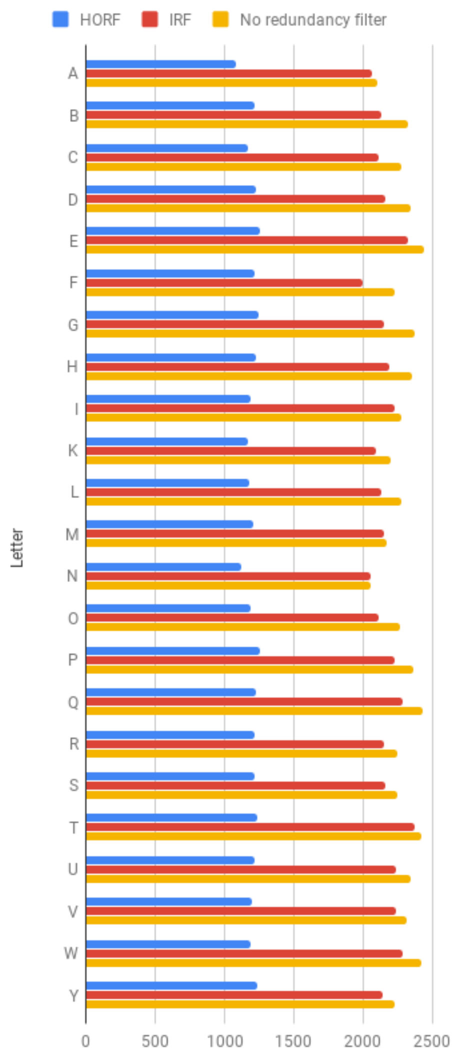


**Figure 3: Occurrence by shape**

$\sigma = 0.3 * ((ksize - 1) * 0.5 - 1) + 0.8$ with $ksize = 15$, because it has shown improved accuracy in [8].

Figure 4 shows the accuracy according to the number of eigenvectors for PCA plus k-NN as classifier. Note that in this plot we show the accuracy of the IRF dataset and the accuracy over our HORF dataset. In addition, the use of blurring is showed in this plot, as expected the blurring helped to improve the accuracy in all cases.

Table 1 shows the accuracy for the others classifiers, with and without use of blurring. Note that PCA+SVM showed the best accuracy for both cases, with and without use of blurring.

## 6 CONCLUSIONS AND FUTURE WORKS

In this study we introduced HORF, a dataset filtered by a method to reduce the redundant images using the computation of axis of

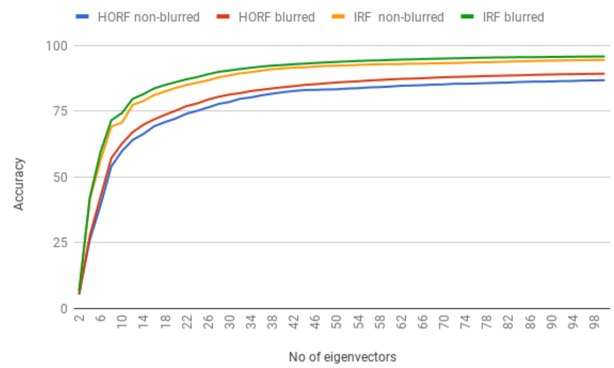


**Figure 4: Accuracy according to the number of eigenvectors using PCA+kNN to classify.**

**Table 1: Classifier performances for IRF and HORF. The best classifier is PCA+SVM**

| Model | IRF | | HORF | |
|---|---|---|---|---|
| | Non-blur. | Blurred | Non-blur. | Blurred |
| k-NN (k=1) | 95.50% | 96.91% | 87.73% | 91.26% |
| SVM | 97.86% | 99.80% | 92.73% | 99.14% |
| PCA+k-NN (k=1) | 94.56% | 95.90% | 86.85% | 89.32% |
| PCA+SVM | 99.61% | 99.87% | 97.77% | 99.42% |

least inertia and we compared with the previous proposed iterative image selection process, IRF.

The computation of axis of least inertia was applied to the Irish Sign Language dataset has shown an improved reduction in the number of redundant frames in the ISL dataset. A new and less redundant dataset for sign language is made available for the community.

Classifiers had shown accuracy slightly lower for HORF than for IRF, proving that the new dataset contains less similar frames.

As a future application, we are planning to create a classifier in two stages; the first will be using the computation of axis of least inertia and the second a machine learning common classifier. In addition, we want to use Convolutional Neural Network and XGBoost.

## ACKNOWLEDGEMENTS

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under grant number SFI/12/RC/2289.

## REFERENCES

[1] Mohamed Farouk. 2015. *Principal Component Pyramids using Image Blurring for Nonlinearity Reduction in Hand Shape Recognition.* Ph.D. Dissertation. Dublin City University, Ireland.
[2] Andreas Geiger, Philip Lenz, and Raquel Urtasun. 2012. Are we ready for autonomous driving? the KITTI vision benchmark suite. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 3354–3361.
[3] Berthold Horn, Berthold Klaus, and Paul Horn. 1986. *Robot vision.* MIT press.
[4] Learn Irish Sign Language. 2012. Learn Irish Sign Language. http://www.learnirishsignlanguage.ie/alphabet.html

[5] Lorraine Leeson and John I. Saeed. 2012. *Irish Sign Language : A Cognitive Linguistic Account.* Edinburgh University Press. https://doi.org/10.1016/B978-1-85617-745-0.00001-2

[6] Marlon Oliveira, Houssem Chatbri, Ylva Ferstl, Mohamed Farouk, Suzanne Little, Noel E. O'Connor, and Alistair Sutherland. 2017. A Dataset for Irish Sign Language Recognition. In *Irish Machine Vision & Image Processing Conference.* in press.

[7] Marlon Oliveira, Houssem Chatbri, Ylva Ferstl, Suzanne Little, Noel E. O'Connor, and Alistair Sutherland. in press. Irish Sign Language Recognition Using Principal Component Analysis and Convolutional Neural Networks. In *2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA).*

[8] Marlon Oliveira, Houssem Chatbri, Suzanne Little, Noel E. O'Connor, and Alistair Sutherland. 2017. A comparison between end-to-end approaches and feature extraction based approaches for Sign Language recognition. In *Image and Vision Computing New Zealand (IVCNZ), 2017 International Conference on (in press).* IEEE, 1–6.

[9] Du-Ming Tsai and Ming-fong Chen. 1995. Object recognition by a linear weight classifier. *Pattern recognition letters* 16, 6 (1995), 591–600.

[10] Chen Zhu, Xiao Tan, Feng Zhou, Xiao Liu, Kaiyu Yue, Errui Ding, and Yi Ma. 2018. Fine-grained Video Categorization with Redundancy Reduction Attention. *arXiv preprint arXiv:1810.11189* (2018).