Machine Learning and Data Mining Techniques for Sign Language Recognition and Retrieval System

R. Madana Mohana[1] Dr. A. Rama Mohan Reddy[2]

[1]Associate Professor, Department of Computer Science and Engineering, Bharat Institute of Engineering and Technology, Ibrahimpatnam - 501 510, Hyderabad
[2] Professor of CSE, Sri Venkateswara University College of Engineering, S. V. University, Tirupati - 501 510, Andhra Pradesh

Abstract: - This paper presents a technique to acknowledge 32 American Sign Language distinctive letters and numbers from image signs, independence of signer and environment of image capture. Input pictures are mapped to the YC_bC_r color space, binarized and resized to 70×70 px. Principal Component Analysis is then performed on these binary images exploitation their pixels as features. This technique recognized signer-dependent signs with an accuracy of 100% and signer-independent signs with an accuracy of 62.37% that will increase to 78.49% if dissimilar signs only used.

Keywords- SLR Systems; Image Processing; Principal Component Analysis, Support Vector Machine.

1 Introduction

Sign Language Recognition (SLR) systems are technological contributions that enhance the lives of the hearing impaired. An ideal SLR system can enable its user to communicate with other users, computers and the Internet in their natural environment, while minimizing user constraints and bandwidth usage. They can also serve as tutors, providing immediate and accurate feedback for students of sign language.

Systems that focus on a specific mode of sign language communication called 'fingerspelling', in which words or sentences are spelt out, are particularly useful in this regard. This representation uses only hands, and the letters are signed using signs from the sign language manual alphabet. In many ways, fingerspelling serves as a bridge between the sign language and the oral language that surrounds it and thus, can also be used for representing words of the corresponding oral language that have no sign language equivalent.

One of the earliest works in SLR Systems was by [1], in which they used Hidden Markov Model to recognize 40 American Sign Language (ASL) signs. Accuracy of 99% was achieved when the user wore coloured gloves, and 92% was obtained without coloured gloves. Since then, several other studies have been performed on sign and gesture recognition. A survey of such methods employed in various sign language recognition systems has been performed by Ong and Ranganath [2]. These studies can broadly be classified as follows:

According to type of sign recognized - An SLR system can either aim to recognize static hand postures or the sign language alphabet from single-gesture images [3], letters with local movement using sequential feature vectors and dynamic information [4] [5], or sign language words and sentences, which includes local and path movement of hands, using segmentation and tracking of the hands, which are captured as real-time continuous signed videos [1]. Fingerspelling videos are a subset of such continuous signing videos, in which individual letter signs spell out words and sentences. Gesture frames need to be isolated from such videos before recognition, either by annotation [6], or through automatic segmentation methods [7].

According to capture method - Initial attempts at sign language recognition relied largely on sophisticated hardware to capture user input. These included data or cyber gloves [8], motion sensing input devices such as Kinect [9], etc. A later development was to use coloured gloves, which helped in isolating the high coloured hand area from the rest of the image [10] [11], as well as depth sensors [12]. However, due to the intense computational needs, high cost and/or unrealistic constraints that these methods placed on the user, most work in SLR systems now focuses on using commonly available hardware such as webcams, mobile cameras, etc. to capture signed videos or images and improving their quality for use in recognition systems via explicit image processing techniques [7]. Capturing or videotaping using such devices is largely non-intrusive, can be used anywhere, and the data collected can be stored and retrieved at any time with minimum hassle. An alternate to videotaping is an optical motion-tracking system [13], in which a set of digital cameras are used to record the position and movement of infrared light-emitting diodes placed on the subject's body, but this method requires that the diodes always face the cameras.

According to user constraints - Ideal SLR systems must recognize signs accurately independent of signer style and skin colour, background, lighting and scaling conditions. However, systems that are successful and yet impose no constraints on its users are difficult to design, owing to differences in the style of signing among various users, environment of image capture, etc. Hence, current SLR systems either use huge datasets, or require images from system testers to be a part of its training set [3], or impose minor constraints to improve recognition accuracy, such as those with respect to:

(i) background, which is usually required to be plain, non-reflective and in sharp contrast to skin colour [14],

(ii) lighting conditions, which are required to be bright and uniform [1] ,

(iii) signing style of the tester, which is required to be the same and/or consistent with that of system trainers [3], minimally varied and sometimes employing the use of coloured gloves and,

(iv) images captured, which require that a majority of the captured image is covered by the hands and/or face performing the sign, while avoiding occlusion [15].

In this paper, we propose a novel method to achieve a signer-independent sign language recognition system, by using a combination of certain techniques of image processing before employing the mathematical procedure of Principal Component Analysis. We have attempted to achieve signer-independence while not having a very large dataset or sign image corpus, by reducing any input image to a standard form comprising only of the hand involved in signing the gesture. We have been able to successfully combine our preprocessing techniques with an eigen-vector based strictly signer-dependent method such as PCA as well as with a learning method such as SVM.

This paper is organizes as follows: Section 2 gives a brief introduction of PCA, Section 3 explains our implementation of the proposed scheme, Section 4 presents the experimental results, Section 5 explains an alternate approach using Multiclass-SVM and Section 6

2 Principal Component Analysis

Principal Component Analysis is a method of obtaining dimensionality reduction, by extracting the most relevant information from high-dimensional data. A PCA vision system has two phases - an offline phase for training, and an online phase for recognition. For the set of images in the training dataset, it finds eigenvectors and eigenvalues from its covariance matrix. The eigenvectors are ordered by their eigenvalues, and only the most useful vectors (principal components) are selected as features. These give 'eigenimages' for images in the training set, which are a reduced-dimensional form of the original images, obtained by multiplying them with the chosen set of eigenvectors. New unknown images (also converted to eigenimages) are tested against these training eigenimages for classification and assigned the same class as the most similar of the training images. Most similar is defined as the least Euclidean distance in the k-dimensional space spanned by the features.

PCA was initially implemented as a technique for face recognition by [16] and [17]. However, one of the pioneering works on the application of PCA to SLR systems was done by [7], in which they achieved 98.4% offline recognition rate, for recognizing signer-dependent sign images from real-time video. Most research on PCA SLR systems, like in [3], [18], etc., has been restricted to signer-dependent sign recognition. We propose a scheme to achieve signer independence using Principal Component Analysis, achieved by employing certain techniques of image preprocessing before the implementation of PCA.

3 Implementation

Our system follows the following sequence of steps:

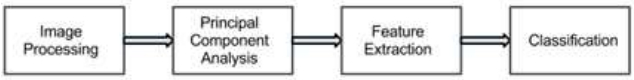

FIGURE I: SYSTEM OVERVIEW

A. Image Processing

A procedure to preprocess the images of signs is necessary to better extract image features in the case of a signer-independent system, which should be invariant to background data, translation, scale, and lighting conditions. Such a procedure would also enhance image quality for feature extraction and minimize noise.

The image is transformed from RGB to YCbCr colour space. YCbCr / Y'CbCr is a way of encoding RGB information, where Y' is the luma component and Cb and Cr are the blue-difference and red-difference chroma components. Y′ (with prime) is distinguished from Y which is luminance, meaning that light intensity is nonlinearly encoded based on gamma corrected RGB primaries. The threshold value and effectiveness value is computed for the three channels of the image individually using Otsu's thresholding algorithm [19], which helps identify the foreground (signed hand) from the noisy background. This method also returns an effectiveness metric, a value from 0 to 1 which determines how well an image can been separated into the two classes while

minimizing intra-class variance. Using a combination of the YCbCr colour space, which is better for highlighting lighting and colour differences in an image, and the Otsu's algorithm, we are able to retrieve the best parameters for binarizing any image. The channel and threshold value corresponding to the highest effectiveness is selected as the form of the image to be binarized. With this method, a binarized image has pixel value of 1 in the hand region and 0 elsewhere.

Holes in the hand region are sometimes caused by poor lighting, shadow, occlusions or other noise. These are filled using a small repeating disc-shaped structural element that minimally connects disconnected portions of the signing hand. This is done so that the majority portion of the binary image, which is the hand, is a connected component that can be segmented for later use. Then, a bounding box of the hand in the image is obtained, a method which requires isolated white pixels in the image to be cleaned so that the bounding box covers the hand alone. The image is cropped according to the boundaries of this box, which makes the image comprise solely of the hand. A further step is to remove the wrist from the image, which may otherwise interfere with the shape of the sign. The wrist-segmented image is then resized to a pixel size of 70x70. This resolution was chosen empirically, and it was found to give the most distinguishing set of features when PCA is implemented. Increasing the resolution further did not improve the rate of recognition significantly. However, capturing the images to test this system such that the signing hand was against flat and non-reflective surfaces is found to maximize the accuracy of recognition.

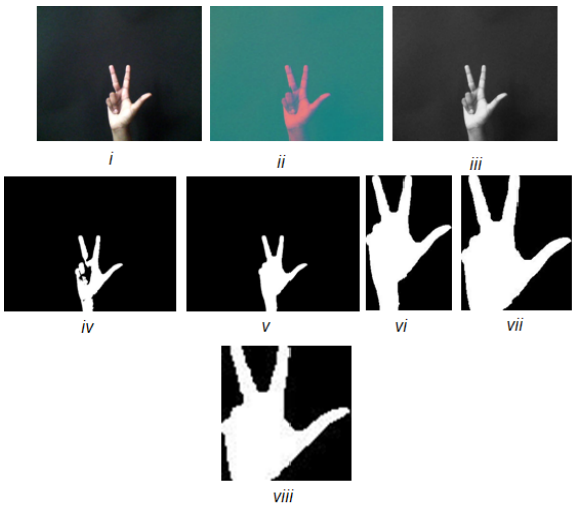

FIGURE II: IMAGE PROCESSING STEPS - (i) Original RGB Sign Image (ii) YCbCr Image (iii) Isolated Channel Representation (iv) Binarized (v) Filled (vi) Bounding Box Around Hand (vii) Wrist Segmented (viii) Resized to 70x70

B. Principal Component Analysis and Feature Extraction

PCA is implemented on these preprocessed images to obtain eigenvectors. They were ordered by their eigenvalues, and it was found that the first 20 eigenvectors were sufficient to optimally preserve the image information of each sign image instance in their representation as an eigenimage, and thus were selected for use as features. The PCA algorithm used for training and testing is as follows:

Algorithm: SLR System
Input : Training image set, Testing image set.
Output: The label of the sign class that the test image is recognized as.

1. Preprocess all training images, and convert all of them to image column vectors. Append these columns to create training image matrix A.
2. Get mean column vector M of the data matrix A. Subtract mean M from each of the columns of A to result in mean centered matrix A.
3. Compute the covariance matrix C of A as C = AA'.
4. Obtain the eigenvectors matrix E and eigenvalues vector V of C.
5. Rearrange the eigenvector columns in E as the corresponding eigenvalues in V are sorted in descending order. Select the first 20 eigenvectors (columns) of E, to create F.
6. Project the centered matrix A onto F to get the feature matrix P = F'A.
7. Obtain test image I. Preprocess and transform the image I into a column vector J. Subtract the mean vector M from the image vector J, J = J – M.
8. Project the image vector J onto the eigenmatrix F to get the feature vector Z = F'J.
9. Compute the Euclidian distance d between the feature vector Z and all the column vectors in the feature matrix P to identify the column having the minimum d.
10. Obtain the label corresponding to the column having the minimum d.

SLR system ends.

C. Classification

The eigenimage of a test image is compared with those of the training images using the Euclidean Distance metric:

$$d = \sqrt{\sum_{i=1}^{n} (X_i^2 - Y_i^2)} \quad (1)$$

where d is the Euclidean Distance measure between the column vectors of the training and testing eigenimages. The test image is recognized as the sign represented in the training instance with which it has the least d value, and is assigned to the sign-class to which this training instance belongs.

4 Experimental Results

A. Dataset

The [20] dataset contains sets of ASL signs signed by 5 volunteers, out of which we have used the 1st volunteer's set for training the model and testing signer-dependent recognition. This dataset consists of 900 images, with 25 samples each of the 26 letters of the alphabet and 10 numerals. The images captured in this dataset do not use any special gloves on the signer's hands, and are also wrist-segmented. The two dynamic ASL gestures, for the letters "J" and "Z", have been presented as static signs, with a rotated gesture to differentiate them from others that are similar.

Out of the 25 instances for each letter/numeral in the above-mentioned dataset, 22 were used as training images (a total of 792 images) and 3 instances for signer-dependent testing in the recognition phase (a total of 108 images).

The dataset for testing signer-independent recognition was created by us, by capturing images from a webcam of 2 users signing 2 instances of 26 alphabets and 10 numerals each (a total of 144 images).

B. Similarity between Signs

In principle, there are 36 classes in the training dataset. However, there are gestures that are difficult to classify due to their similarities.

Some gestures are identical – "0" and "O", "1" and "D", "2" and "V", and "6" and "W", so we consider them as the same class.

The difference between "M", "N", "A", "O", "S" and "T" are the slight variation in orientation of hand and thumb placement with respect to a closed fist. Similarly, "2"/"V" and "K" only differ by slight variation in thumb placement. "Z", "G" and "1"/"D" vary by the difference in the angle of rotation of hand. "R" and "U" also differ only by a minor variation in the position of the first two fingers. These differences are so small that they can be virtually indiscernible in a binary image. Consequently, it is difficult to distinguish among these signs accurately in our proposed algorithm, as the distinguishing features can get lost during the binarizing process.

Therefore, the result of the testing has been presented in two ways – one where we consider 32 unique classes and one where we consider 23 classes after we merge similar signs into a joint class.

C. Recognition Rate

Recognition rate is defined as the ratio of the number of successfully recognized test images to the number of samples used for testing in the online recognition phase.

The results of our SLR system are given in the following table:

Table 1
RECOGNITION RATE FOR PCA

PCA	Number of Classes	
	32 classes	***23 classes***
Signer-Dependent	100%	100%
Signer-Independent	62.37%	78.49%

5 Alternate Approach

Multiclass-SVM, a machine learning algorithm, was also run on the above preprocessed images to compare the efficiency of our signer-independent implementation of PCA. The approach selected was one-versus-all, as it was found to be faster and more accurate than the one-versus-one approach. The following is the SVM algorithm used:

Algorithm: Multiclass-SVM SLR System
Input : Training image set, Testing image set.
Output: The label of the sign class that the test image is recognized as.

1. Convert the preprocessed images to image row vectors. Append all training images to create image matrix Data (mxn), and all test images to create image matrix Data1 (m1xn).
2. Create a row-vector flag (mx1), which contains the class labels for each instance of a class ordered according to the training set.
3. Train the multiclass-SVM on the training data.
4. Obtain test image I.
5. To classify the test image in the jth iteration, divide the training data such that one class is the jth class and the other is all classes not j. That is, one class is the jth class and the other is all the classes from (1 to j-1) and (j+1 to m). Run the SVM classifier to place the test image in one of these two classes. The image is classified when it is put in the jth class. Continue this step while the image has not yet been classified and more than two classes are still left to train the SVM.
6. Obtain the label corresponding to the class to which test image I belongs.

Multiclass-SVM SLR System ends.
The recognition rates given by Multiclass SVM are given below:

Table 2
RECOGNITION RATE FOR MULTICLASS-SVM

Multiclass SVM	Number of Classes	
	32 classes	*23 classes*
Signer-Dependent	100%	100%
Signer-Independent	39.78%	55.91%

6 Conclusion and Future Work

A Sign Language Recognition system was implemented using PCA as well as SVM and tested on signer dependent and independent fingerspelling images. A novel approach to achieve signer independence was implemented, which involved processing the images to achieve a standard binarized representation, thus removing the effects of differing backgrounds, lighting conditions, signing styles or skin colour.

With help of this method, an accuracy of ~80% was achieved by using PCA for signer-independent data, which is an excellent result considering the signer dependent nature of PCA. PCA was also seen to outperform SVM using this method.

Compared to the onus on the recognition of manual signs in current SLR systems, the focus on detection of signs that include not just hands, but also non-manual components like facial expressions and body language is relatively minimal. This is also a computationally intensive task and thus has great scope for future work in the field of sign language recognition. Yet another possible design alternate to improve speed and accuracy of SLR systems, particularly in the detection of same/similar signs, could be the use of custom features, which is an area that needs to be explored in further research.

References

[1] Thad Starner and Alex Pentland. "Real-time american sign language recognition from video using hidden markov models." In *Motion-Based Recognition*, pp. 227-243. Springer Netherlands, 1997.

[2] Sylvie C.W. Ong and Surendra Ranganath. "Automatic sign language analysis: A survey and the future beyond lexical meaning." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27, no. 6 (2005): 873-891.

[3] M. G. Suraj and D. S. Guru. "Appearance based recognition methodology for recognising fingerspelling alphabets." In *International Joint Conference on Artificial Intelligence*, pp. 605-610. 2007.

[4] Shu-Fai Wong and Roberto Cipolla. "Real-time interpretation of hand motions using a sparse bayesian classifier on motion gradient orientation images." In *Proceedings of the British Machine Vision Conference*, vol. 1, pp. 379-388. 2005.

[5] Jose L. Hernandez-Rebollar, Robert W. Lindeman, and Nicholas Kyriakopoulos. "A multi-class pattern recognition system for practical finger spelling translation." In *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces*, p. 185. IEEE Computer Society, 2002.

[6] Vassilis Athitsos et al. "Large lexicon project: American Sign Language video corpus and sign language indexing/retrieval algorithms." In *Proc. Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*. 2010.

[7] Henrik Birk, Thomas B. Moeslund and Claus B. Madsen. Real-Time Recognition of Hand Alphabet Gestures Using Principal Component Analysis, In *Proceedings of the 10th Scandinavian Conference on Image Analysis (SCIA'97)*, Lappeenranta, Finland, 1997.

[8] Wen Gao, Jiyong Ma, Jiangqin Wu, and Chunli Wang. "Sign language recognition based on HMM/ANN/DP." In *International Journal of Pattern Recognition and Artificial Intelligence* 14, no. 05 (2000): 587-602.

[9] Nicolas Pugeault, and Richard Bowden. "Spelling it out: Real-time asl fingerspelling recognition." In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pp. 1114-1119. IEEE,2011.

[10] Suat Akyol and Ulrich Canzler. "An information terminal using vision based sign language recognition." In *ITEA Workshop on Virtual Home Environments, VHE Middleware Consortium*, vol. 12, pp. 61-68. 2002.

[11] Ila Agarwal, Swati Johar, and Jayashree Santhosh. "A Tutor for the hearing impaired (developed using Automatic Gesture Recognition)." In *International Journal of Computer Science, Engineering and Applications (IJCSEA)* Vol 1.

[12] Dominique Uebersax, Juergen Gall, Michael Van den Bergh, and Luc Van Gool. "Real-time sign language letter and word recognition from depth data." In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pp. 383-390. IEEE, 2011.

[13] László Havasi, and Helga M. Szabó. "A motion capture system for sign language synthesis: overview and related issues." In *Computer as a Tool, INR 144,266.17OCON 2005. The International Conference on*, vol. 1, pp. 445-448. IEEE, 2005.

[14] Rogerio Feris, Matthew Turk, Ramesh Raskar, Karhan Tan, and Gosuke Ohashi. "Exploiting depth discontinuities for vision-based fingerspelling recognition." In *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on*, pp. 155-155. IEEE, 2004.

[15] Ming-Hsuan Yang, Narendra Ahuja, and Mark Tabb. "Extraction of 2d motion trajectories and its application to hand gesture recognition." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24, no. 8 (2002): 1061-1074.

[16] Michael Kirby, and Lawrence Sirovich. "Application of the Karhunen-Loeve procedure for the characterization of human faces." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 12, no. 1 (1990): 103-108.

[17] Matthew Turk, and Alex Pentland. "Eigenfaces for recognition." *Journal of cognitive neuroscience* 3, no. 1 (1991): 71-86.

[18] Jiang-Wen Deng, and Hung-Tat Tsui. "A novel two-layer PCA/MDA scheme for hand posture recognition." In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 1, pp. 283-286. IEEE, 2002.

[19] Otsu, N., "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 9, No. 1, 1979, pp. 62-66.

[20] A. L. C. Barczak, N. H. Reyes, M. Abastillas, A. Piccio, and T. Susnjak. "A new 2D static hand gesture colour image dataset for asl gestures." *Res Lett Inf Math Sci* 15 (2011): 12-20.