

A SIGNATURE BASED DRAVIDIAN SIGN LANGUAGE RECOGNITION BY SPARSE REPRESENTATION

C. Anusuya¹, S. Md Mansoor Roomi², M. Senthilarasi³

¹PG Student, Dept. of ECE, Thiagarajar College of Engineering, Madurai.

²Assistant Professor, Dept. of ECE, Thiagarajar College of Engineering, Madurai.

³Ph.D Research Scholar, Dept. of ECE, Thiagarajar College of Engineering, Madurai.

ABSTRACT

Sign language is a visual-gestural language used by deaf-dumb people for communication. As normal people are unfamiliar of sign language, the hearing-impaired people find it difficult to communicate with them. The communication gap between the normal and the deaf-dumb people can be bridged by means of Human-Computer Interaction. The objective of this paper is to convert the Dravidian (Tamil) sign language into text. The proposed method recognizes 12 vowels, 18 consonants and a special character "Aytham" of Tamil language by a vision based approach. In this work, the static images of the hand signs are obtained a web/digital camera. The hand region is segmented by a threshold applied to the hue channel of the input image. Then the region of interest (i.e. from wrist to fingers) is segmented using the reversed horizontal projection profile and the Discrete Cosine transformed signature is extracted from the boundary of hand sign. These features are invariant to translation, scale and rotation. Sparse representation classifier is incorporated to recognize 31 hand signs. The proposed method has attained a maximum recognition accuracy of 71% in a uniform background.

KEYWORDS

Tamil sign language; Gesture recognition; Reversed Horizontal Projection Profile; Signature; Discrete Cosine Transform; Sparse Representation;

1. INTRODUCTION

Deaf-dumb people naturally communicate between their own community and normal people using sign language as a kind of gesture. In India, census 2011 reported that over 5 million people are deaf and 2 million people are mute [1]. Of these, 63% are said to be born deaf, the others losing their hearing in different accidents. There arises the need for sign language interpreters who can translate sign language to spoken language and vice versa. But, the availability of such interpreters is limited and expensive. This resulted in the development of sign language recognition (SLR) systems which could automatically translate the signs into the corresponding text or speech. Such systems can make deaf-hearing interaction easier in public functions such as courtroom, conventions and meetings also specific transactional domains like post office, bank etc. SLR, as one of the important research areas of human-computer interaction (HCI), has spawned more and more interest in HCI society. From a user's point of view, the most natural way to interact with a computer would be through a speech and gesture interface. Thus, the research of sign language and gesture recognition is likely to provide a shift paradigm from point-and-click user interface to a natural language dialogue-and-spoken command-based interface. Another application is a bandwidth conserving system allowing communication between signers where recognized signs are the input of the communication system at one end, can be translated to avatar based animations at the other end. Additional applications of SLR are an automated sign language teaching system and automated annotation system for the video databases of native signing.

Unlike general gestures, sign language is highly structured so that it provides an appealing test bed for new ideas and algorithms before they are applied to gesture recognition. Attempts to automatically recognize sign language began to appear in the literature in the 1990s. While significant progress has already been made in recognition of sign languages of other countries, a very limited work has been done in Indian Sign Language computerization [2]. SLR system can be categorized into two types, data glove based approach and vision based approach. Although the data glove based SLR achieves good performance, it involves complicated hardware and too expensive. The vision based approach does not involve complicated hardware usage and hence it is the most widely used method in SLR systems [3]. However, it is very difficult for the complex backgrounds and illuminations. Grobel and Assan [4] extracted two-dimensional (2-D) features from video recordings of signers wearing colored gloves. Vogler and Metaxas [5] used computer vision methods to extract the 3-D parameters of a signer's arm motions as the input of HMM, and recognized continuous American Sign Language sentences. Starner et al. [6] used a view-based approach for continuous American SLR, when the camera was mounted on the desk or in a user's cap in recognizing the sentences with 40 different signs. Subha Rajam and Balakrishnan [7], used binary number generated by the up and down position of fingers in a static gesture for recognizing 31 Tamil Sign alphabets in a black background where the signer is in need to wear a black band in the wrist to simplify the process of segmentation.

From the literature, the major challenges of SLR are developing an algorithm to solve signer-independent and large vocabulary continuous sign problems. Signer independence is highly desirable since it allows a system to recognize a new signer's sign language without retraining the models. In this work, a signer-independent, finger spelling recognition is developed using a contour descriptor namely discrete cosine transformed coefficients of normalized signature, which is invariant to translation, rotation and scale. The major contributions of this framework are the Region of Interest (ROI) segmentation using Reversed Horizontal Projection Profile (R-HPP) and the development of large vocabulary by combining a sequence of recognized gestures of sign alphabet.

The reminder of this paper is organized as follows. Section II proposes the region of interest (ROI) segmentation, feature extraction and recognition. In Section III, the experimental results are presented. The conclusion is given in the last section.

2. PROPOSED METHODOLOGY

The structure of the hand sign recognition algorithm is shown in Figure 1. The hand sign of Tamil alphabets are acquired using a digital camera. The still images of these hand signs are in RGB color space and it is converted to HSV color model for foreground extraction. The histogram of hue channel is derived and its valley point of the bimodal histogram provides the threshold value for segmentation. The segmented region has the palm and arm region. The region of interest (ROI) is only from the wrist portion of the hand to fingers. Hence, the ROI is cropped by reverse horizontal projection profiling. The signature of the ROI is extracted and normalized. Discrete cosine transform (DCT) is applied on the normalized signature values to make the features rotation invariant. In the training phase, the cosine transformed normalized signature of hand signs are stored in a database along with the corresponding labels. In the testing phase, a person shows the hand signs of tamil alphabets as an input and the features are extracted. The features of training hand signs set are extracted and stored in the database for template matching. Sparse representation classifier is used for the recognition of test signs using template hand signs. At the end, the system delivers a written language rendition of what the person signed during test phase.

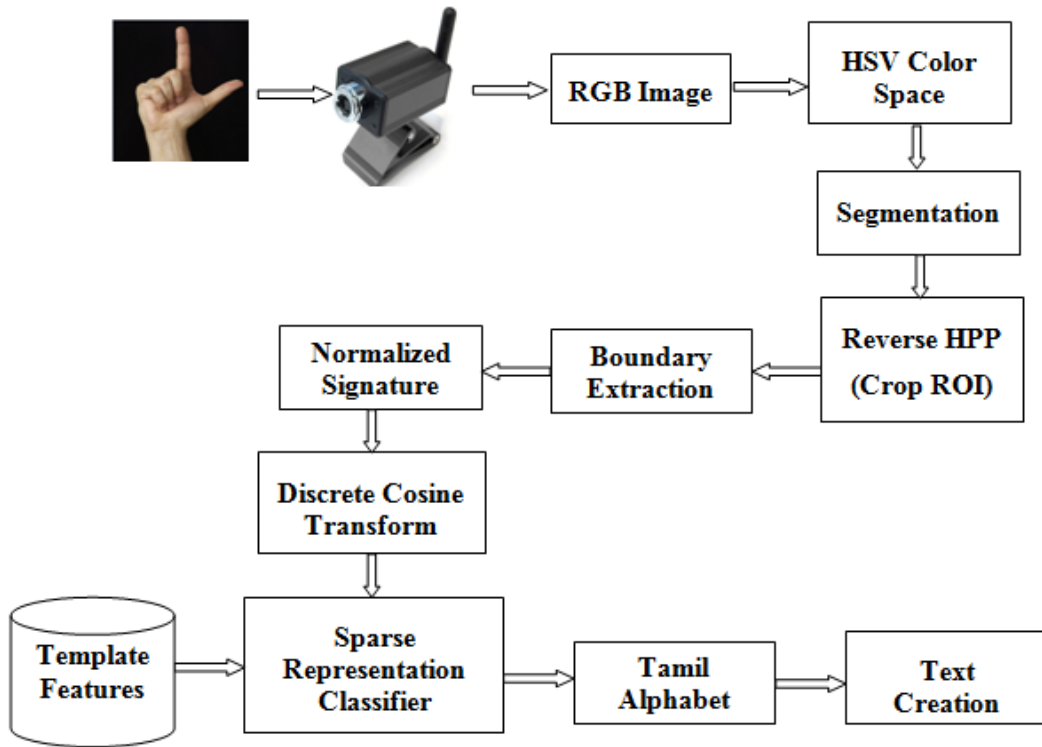


Figure 1. Structure of Tamil sign language recognition system

2.1. ROI segmentation

In this work, the gestures of 31 Tamil sign alphabets are acquired in a uniform green background. These images are in RGB color model and are stored in JPEG format. The advantage of image acquisition in the uniform green background is that skin region is highly discriminable in Hue channel rather than Saturation and Value channel of HSV color model. By selecting an adequate threshold value T , the hue image can be converted to binary image.

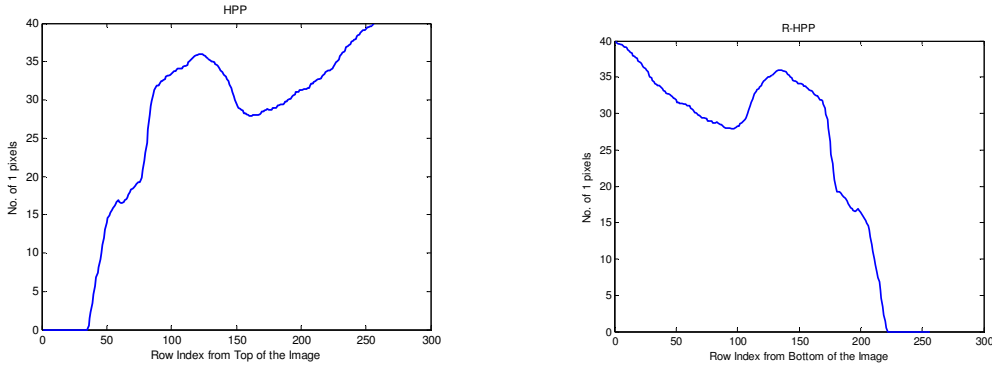
$$B(x, y) = \begin{cases} 1, & H(x, y) \leq T \\ 0, & H(x, y) > T \end{cases} \quad (1)$$

where ' B ' represents the binary image, ' H ' represents the hue image and ' T ' is the threshold selected from the valley of hue histogram. In Tamil SLR, the fingers are used to represent the Tamil alphabets, so the hand region up to the wrist is the Region of Interest (ROI), segmented based on the location of wrist. Wrist detection using Reversed Horizontal Projection Profile (R-HPP) is proposed in this paper. The projection of a binary image onto a line is obtained by partitioning the line into bins and finding the number of 1 pixels that are on lines perpendicular to each bin [8]. Horizontal projection profile can be obtained by finding the number of 1 pixels for each bin in the horizontal direction as follows.

$$HPP(i) = \sum_{j=0}^{n-1} B(i, j) \quad (2)$$

$$R - HPP(i) = HPP(m - 1 - i) \quad (3)$$

Where $i=0, 1, 2, 3, \dots, m-1$, 'B' is the binary image of size $m \times n$ and 'HPP' is the horizontal projection profile. The HPP and R-HPP for the binary image of gesture representing the Tamil alphabet is shown in Figure. 2 (a) and (b) respectively. In R-HPP, the magnitude of the profile starts reducing from the bottom of the hand until it reaches the wrist region, after that it is found to be increased. The corresponding row index of the first valley followed by a peak in R-HPP represents the location of wrist in the binary image. ROI is cropped up to the row corresponding to the first valley point of R-HPP. Segmentation can be avoided when no valley point is present before the peak in R-HPP, which implies that the gesture image has only palm region and no arm portion.



(a)

(b)

Figure 2. (a) HPP

(b) R-HPP

2.2. Feature Extraction

The contour of ROI is unique for every alphabets of TSL. This contour is modeled as signature which represents a two-dimensional object (ROI) by means of a one-dimensional function. Signature is a plot of distance from centroid to the boundary as a function of angle [9]. The centroid (x_c, y_c) of this boundary is calculated as follows.

$$x_c = \frac{1}{N} \sum_{i=1}^N x_i \tag{4}$$

$$y_c = \frac{1}{N} \sum_{i=1}^N y_i \tag{5}$$

where x_i and y_i are the x and y coordinate of i^{th} boundary pixels and N is the total number of pixels belongs to the boundary. Signature is invariant to translation, but it is rotation and scale dependent. The invariance to translation is achieved by shifting the origin of the coordinate system to the centroid of the object as in (6) and (7).

$$x_n = x_i - x_c \tag{6}$$

$$y_n = y_i - y_c \tag{7}$$

where ' x_n ' and ' y_n ' are the x and y coordinate of the boundary pixels after shifting. The distance (D) and angle (θ) from the centroid to every pixel on the boundary is calculated using (8) and (9)

respectively. The invariance to size is achieved after rescaling the signature by normalizing the largest sample to the unity as in (10).

$$D = \sqrt{x_n^2 + y_n^2} \tag{8}$$

$$\theta = \tan^{-1}\left(\frac{y_n}{x_n}\right) \tag{9}$$

$$D_{norm} = \frac{D}{\max(D)} \tag{10}$$

The invariance to rotation can be achieved by transforming the signature into frequency domain using discrete cosine transform (DCT). If the hand is rotated with an angle ϕ , this equals a translation of the circular signature in the phase spectrum with the same angle, whereas the magnitude spectrum remains same. Since the DCT is real valued transform, phase spectrum does not exist, there is only magnitude spectrum. Hence, Discrete Cosine Transformed circular signature of an object is an invariant descriptor with respect to the hand's translation, rotation and size, is considered as the feature vector of length 1×360 .

2.3. Hand Sign Recognition

Sparse representation (SR) of signals has gained importance in recent years. Given a matrix A of size $M \times N$ with over complete dictionary elements as its columns where many of its elements are zeros. Among all the atoms in an over complete dictionary, the sparse representation selects the subset of the atoms which most compactly expresses the input signal and rejects all other less compact representation. Therefore, the sparse representation of signal is naturally discriminative and hence this can be implemented for classification purpose. In [11], SRC has been proved to have better performance than the nearest neighbour and nearest subspace.

In SRC, the dictionary is constructed from the training samples of 31 hand signs as given below in equation 12 and the atoms are normalized to have unit l_2 norm.

$$D = [D_1, D_2, \dots, D_{31}] \in R^{m \times n} \tag{11}$$

Let $y \in R^m$ be the test sample, then the coefficient vector x is obtained by equation 12.

$$x = \min_x \|X\|_0 \quad \text{such that} \quad y = Ax \tag{12}$$

Where l_0 norm represented as $\|X\|_0$ gives the number of non-zero components in the vector x . This equation can be approximated with l_1 norm denoted as $\|X\|_1$ in the below given equation.

$$x = \min_x \|X\|_1 \quad \text{such that} \quad y = Ax \tag{13}$$

The residuals $r_i(y)$ is computed as

$$r_i(y) = \|y - D\zeta x\|_2 \tag{14}$$

Where ζ is the characteristic function. Then the class of query sign is determined by reducing the residual error. The Discrete cosine transformed Signature of the template hand sign images are

extracted and used for the recognition of hand signs using SRC. The features of test hand sign image is matched to the template features which has low residual error.

3. EXPERIMENTAL RESULTS

Sign recognition is evaluated on a set of 31 different signs of the Tamil Sign Language (TSL), each performed by 10 different persons in four distances. The images are captured at 3648×2736 pixels with uniform green background and constant illumination. These images are resized to 256×256 pixels for experimentation. The setup of the designed system consists of Intel Pentium CPU B950 @ 2.10 GHz processor, 2 GB RAM and Windows 7 Professional (32 bit) operating system. The right hand gestures of 12 vowels & Aytham and 18 consonants of Tamil language acquired at a distance of 15 cm from the signer are shown in Figure. 3 and Figure. 4 respectively. The samples of gestures acquired at various scales viz., 15 cm, 30 cm, 45 cm and 60 cm are shown Figure. 5. For the experimentation of proposed Tamil SLR system, 31 hand signs of 10 persons under four scales are captured with uniform background.

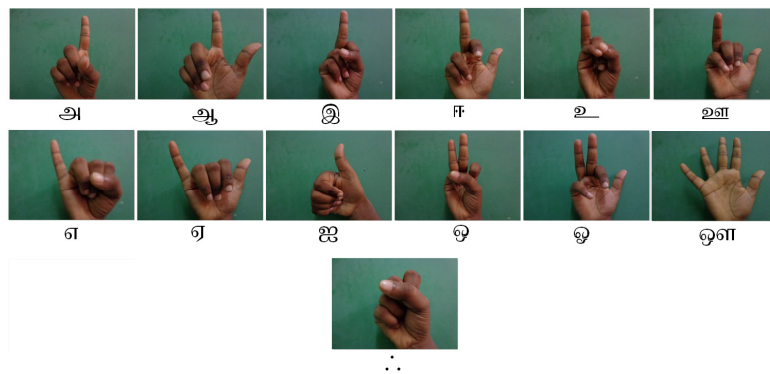


Figure. 3 Gestures of 12 vowels and Aytham in TSL

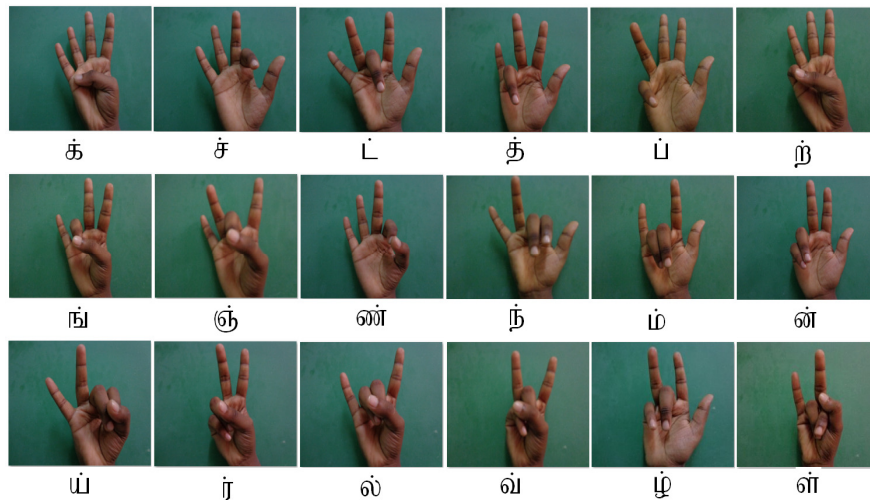


Figure. 4 Gestures of 18 consonants in TSL

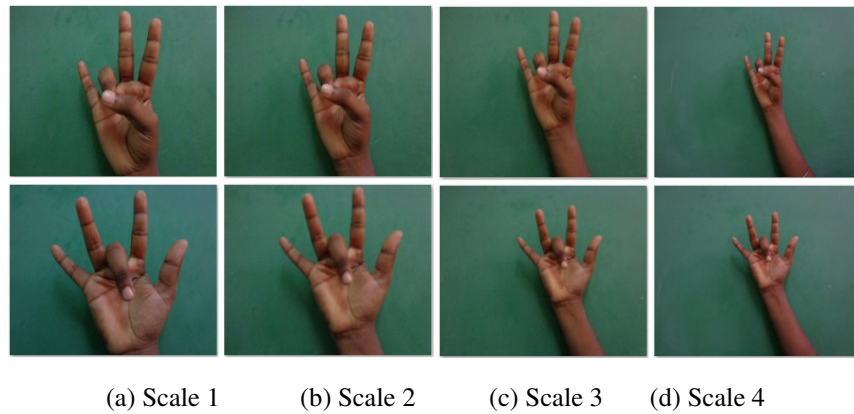


Figure. 5 Hand Signs in four different scales

In the proposed method, the acquired hand gesture image in RGB color model is converted into HSV color model for hand segmentation. As shown in Figure. 6, high contrast exists between the hand and background in the hue channel, which is helpful in segmenting the background effectively. Therefore the hue value corresponding to the valley of histogram ($T=0.2$) is used as a threshold to segment the hand region.

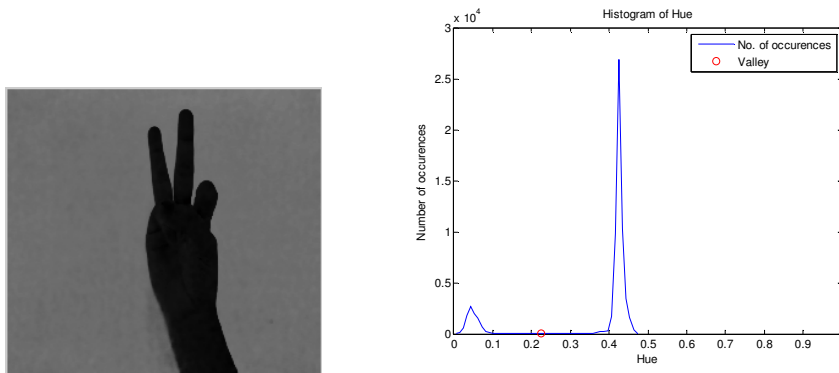


Figure. 6 Hue channel image and its histogram

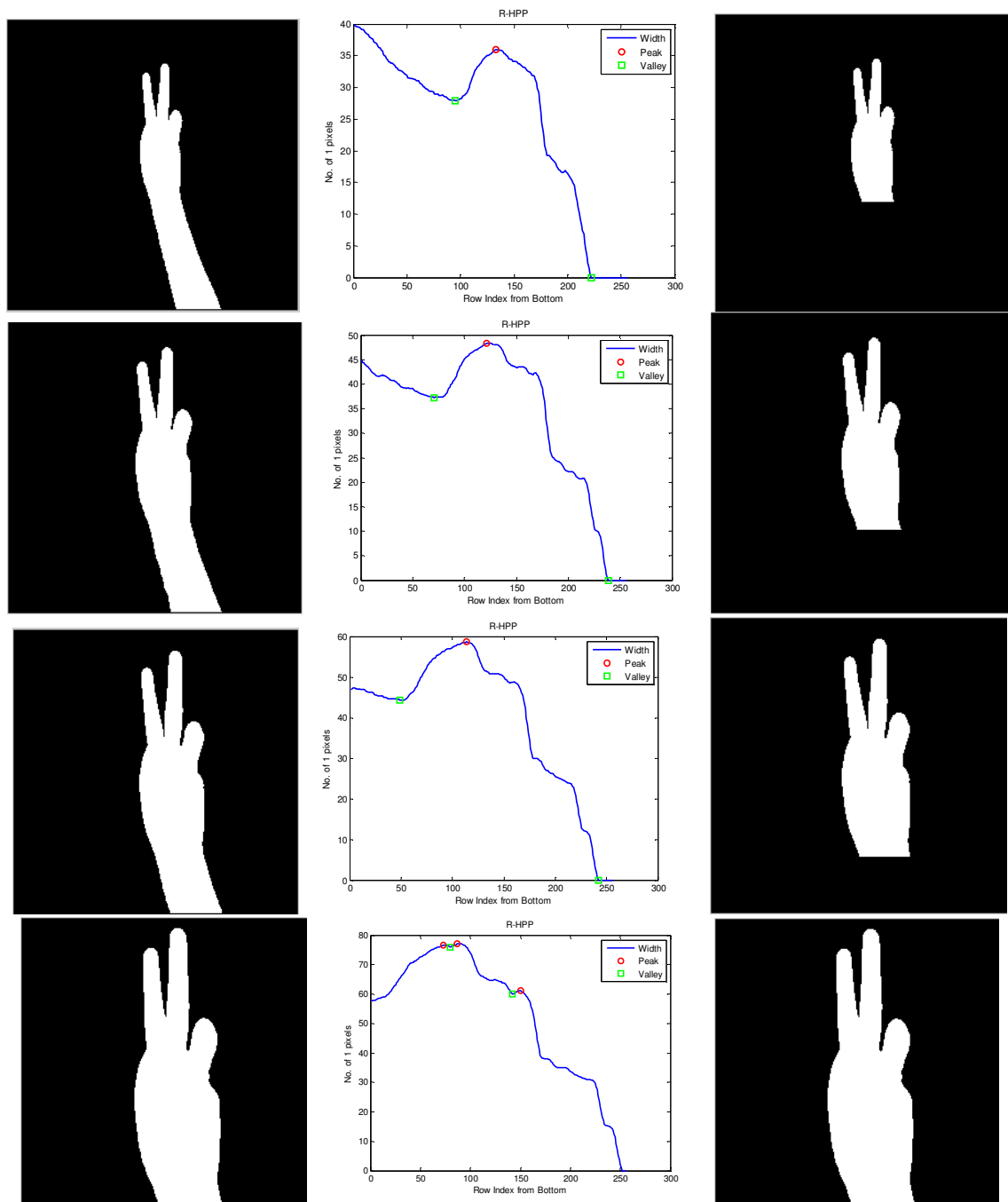


Figure. 7 Output of skin segmentation and ROI segmentation in different scales

The ROI is segmented from the R-HPP of the binary image. The R-HPP and the segmented ROI are shown in Figure. 7, where each row belongs to different scales. In the R-HPP curve, the red and green marker represents the peak and valley respectively. In Figure. 7, there is an arm in the binary image excluding last row. Hence there is a valley before the first peak in the R-HPP. Since, the valley point is not present before the peak in the R-HPP as shown in the last row of Figure. 7 the Cropping of palm region is not required. The contour of the ROI is represented by the signature. The normalized signature of hand gestures is unique for different alphabets of TSL as shown in Figure 8.

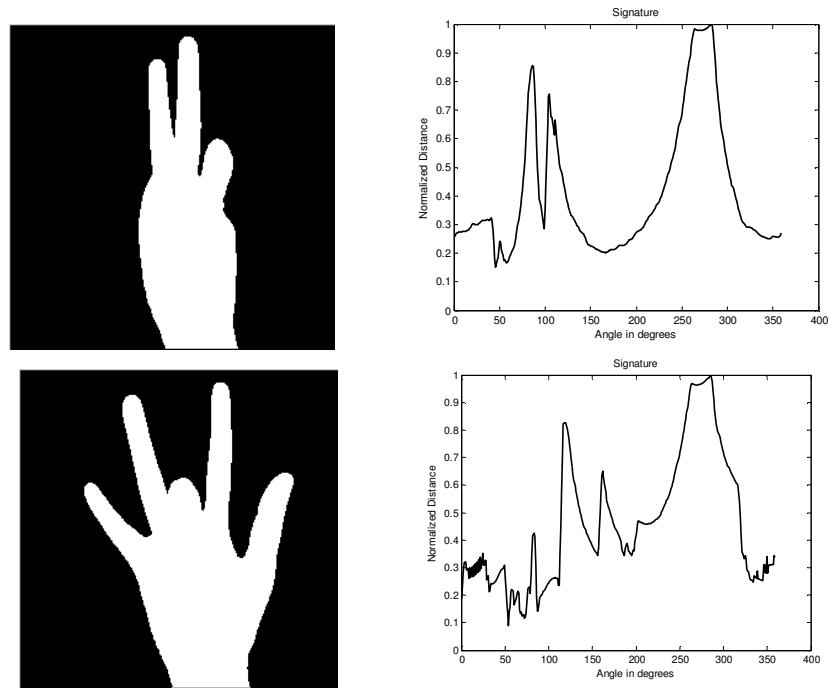






Figure. 8 Normalized signature for different hand gestures

In training phase, the features of 5 persons at 4 different scales are extracted and fed as an input to the SRC. The hand signs of remaining persons are used as a query image for testing. The recognition accuracy of the proposed system is provided in Table I. The accuracy varies due to the shape of the fingers and presence of wrist bands. The maximum accuracy attained is 71% and the average accuracy is 44%. The recognized gestures of alphabets are combined to form a word as shown in Table II. From the table it is observed that, the word 'அம்பலா' has three alphabets. The alphabets 'அ' and 'ம்' are represented by a single hand gesture, whereas the alphabet 'லா' is generated by using a consonant 'ம்' followed by vowel 'அ'. Similarly a large vocabulary of TSL has been created by using the proposed methodology.

Table 1. Recognition Accuracy of Proposed Method

Training Phase	Testing Phase				
Signer ID	Signer ID	Accuracy (%)			
		Scale1	Scale2	Scale3	Scale4
2-6	1	52	65	71	64
	7	45	48	29	29
	8	65	52	58	39
	9	29	39	07	23
	10	55	48	39	32

Table 2. Text(Tamil) from Sign Language

Gesture Input				
Recognized Alphabet	அ	ம்	ம்	ஆ
Word	அ+ம்+ (ம்+ஆ) = அம்மடா			

4. CONCLUSION

A method for the recognition of static signs of Tamil sign language alphabets for HCI is proposed. The static hand signs are captured using a web/digital camera. The normalized signature of the hand sign is extracted. The feature vector representing the hand sign is made scale, translation and rotation invariance by the cosine transform of normalized signature. Sparse representation classifier recognizes the test hand sign by matching with the template hand signs stored in training set. The output of the SLR is tagged with the Unicode of the tamil alphabets and the text corresponding to the sequence of static hand signs is generated and displayed as the recognition output. This approach can handle different types of alphabets in a uniform background and the maximum classification rate of 71% is achieved. In future the dynamic gestures can be considered for creating words or sentences.

REFERENCES

- [1] http://www.censusindia.gov.in/2011census/population_enumeration.html.
- [2] Anuja K., Suryapriya S. and Idicula S. M. (2009), "Design and Development of a Frame based MT System for English to ISL," *World Congress on Nature and Biologically Inspired Computing (NaBIC)*, pp. 1382-1387.
- [3] Divya S., Kiruthika S., Nivin Anton A. L. and Padmavathi S. (2014), "Segmentation, Tracking and Feature Extraction for Indian Sign Language Recognition," *International Journal on Computational Sciences & Applications*, vol. 4, no. 2, pp. 57-72.
- [4] Grobel K. and Assan M. (1997), "Isolated sign language recognition using hidden Markov models," *in Proc. Int. Conf. Systems, Man Cybernetics*, pp. 162-167.
- [5] Starner T., Weaver J. and Pentland A.(1998), "Real-time American sign language recognition using desk and wearable computer based video," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 1371-1375.
- [6] Vogler C. and Metaxas D. (1997), "Adapting hidden Markov models for ASL recognition by using three-dimensional computer vision methods," *in Proc. IEEE Int. Conf. Systems, Man Cybernetics*, pp. 156-161.
- [7] Subha Rajam P. and Balakrishnan G. (2011), "Recognition of Tamil Sign Language Alphabet using Image Processing to aid Deaf-Dumb People," *International Conference on Communication Technology and System Design*, vol. 20, pp. 861-868.
- [8] Ramesh Jain, Rangachar Kasturi and Brian G. Schunck (1995), "Machine Vision", McGraw-Hill, Inc., 1st Edition.
- [9] Rafael C. Gonzalez and Richard E. Woods (2002), "Digital image processing", 2nd Edition, Prentice Hall.
- [10] www.unicode.org/charts/PDF/U0B80.pdf.
- [11] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and M. Yi (2009), "Robust Face Recognition via Sparse Representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, pp. 210-227.

AUTHORS

C. Anusuya is a PG student in the Department of Electronics and Communication Engineering, Thiagarajar College of Engineering, Madurai. She has received her B.E. degree in ECE from University College of Engineering, Nagercoil. She has published 5 papers in an International Journal and Conferences.



S. Mohamed Mansoor Roomi is an Assistant Professor of ECE Dept. at Thiagarajar College of Engineering, Madurai. He has received his B.E. degree from Madurai Kamaraj University in 1990, M.E. degree in Power Systems and Communication Systems from Thiagarajar College of Engineering in 1992 and 1997 respectively. He has awarded Doctorate in Image Analysis from Madurai Kamaraj University in 2009. He has authored and co-authored more than 157 papers in various journals and conference proceedings and undertaken numerous technical and industrial projects.



M.Senthilarasi is a Research Scholar in the department of ECE, Thiagarajar College of Engineering, Madurai. She has received her B.E degree from Raja college of Engineering in 2005 and M.E from Thiagarajar College of Engineering in 2010. She has published and presented 12 papers in various conferences and journals.

