Sign Language Video Processing for Text Detection in Hindi Language

https://doi.org/10.3991/ijes.v4i3.5973

Rashmi B. Hiremath, Ramesh M. Kagalkar
Dr. D. Y. Patil School of Engineering and Technology, Pune, Maharashtra, India

***Abstract*—Sign language is a way of expressing yourself with your body language, where every bit of ones expressions, goals, or sentiments are conveyed by physical practices, for example, outward appearances, body stance, motions, eye movements, touch and the utilization of space. Non-verbal communication exists in both creatures and people, yet this article concentrates on elucidations of human non-verbal or sign language interpretation into Hindi textual expression. The proposed method of implementation utilizes the image processing methods and synthetic intelligence strategies to get the goal of sign video recognition. To carry out the proposed task implementation it uses image processing methods such as frame analysing based tracking, edge detection, wavelet transform, erosion, dilation, blur elimination, noise elimination, on training videos. It also uses elliptical Fourier descriptors called SIFT for shape feature extraction and most important part analysis for feature set optimization and reduction. For result analysis, this paper uses different category videos such as sign of weeks, months, relations etc. Database of extracted outcomes are compared with the video fed to the system as a input of the signer by a trained unclear inference system.**

***Index Terms*—Gesture recognition, Image processing, Sign language, Video processing.**

I. INTRODUCTION

Gesture-based communication is a physical activity by utilizing hands and eye with which we can speak with stupid and hard of hearing individuals. Perceiving human activity from picture arrangements is a standout amongst the most difficult issues in PC vision with numerous vital applications, for example, clever video observation, content-based video recovery, human-robot cooperation, and brilliant home, and so forth. The errand is troublesome not just because of between class varieties, camera developments, foundation jumbling and fractional impediment, additionally to some between class covers and likenesses, for example, running versus running or strolling. Prior takes a shot at human activity acknowledgment in video regularly utilized worldwide representations. Dynamic signal acknowledgment applications require the gaining of a high information rate of hand stances generally gave utilizing movement following gloves that are prepared to do precisely recording finger joint movements through flex sensors in a firmly fitting glove. Hand signal gives a characteristic and natural correspondence methodology for human–computer cooperation. Effective human PC interfaces (HCIs) must be produced to permit PCs to outwardly perceive continuously hand motions. Be that as it may, vision-based hand following and motion acknowledgment is a testing issue because of the unpredictability of hand signals, which are rich in diversities because of high degrees of flexibility (DOF) required by the human hand. So as to effectively satisfy their part, the hand motion HCIs need to meet the prerequisites regarding ongoing execution, acknowledgment exactness, and robustness against changes and cluttered background.

Gesture-based communication is a basic specialized technique for the individuals who experience the ill effects of listening to imperfections. An essential part of communication through signing is hand signals. In this way, communication through signing can be viewed as a gathering of significant and easy to use signals, developments and stances. Hand motion acknowledgment is the most utilized methodology among other correspondence modalities for human-PC association. Dynamic hand motion correspondence is a more characteristic and humanoid method of correspondence with PCs, when contrasted with static hand signals. As a consequence of hands having the capacity to movement in any heading, and to twist to any point in all available directions, dynamic hand signaling is extremely adaptable. Similarly, static hand motions are restricted to altogether less stances.

Non-verbal communication is an imperative piece of correspondence for people. Hard of hearing individuals are individuals who cannot hear or see. A specialist has seen deafness as a physical imperfection. Hard of hearing individuals have their own dialect which is not oral but rather marked. Communication via gestures is a dialect that utilizations outwardly transmitted examples to pass on speaker's musings. It does not utilize sound examples. In the phase of hand signal investigation, hand stances and in addition movement examples are figured from the hand motion outline succession, and the hand motion model is made as needs be. The last stage is hand signal acknowledgment in which the yield of current motion model from the second stage is contrasted and every model consists of motion database where the most coordinated hand motion is chosen as definite acknowledgment result.

This paper proposes a system that used for recognition of hand gestures of sign language from input video stream of the signer and interprets into corresponding Hindi words and sentences. Section 2 illustrates the literature review on this area. Section 3 provides idea of existing systems and limitation s of existing system. Section 4 provides detail ideation of proposed framework. Section 5 shows the results from our proposed technique. And section 6 concludes the paper.

II. LITERATURE REVIEW

In the literature review, we are going to discuss topical methods over the Sign Language Recognition.

Muhammad Rizwan Abid [1] presented the state-of-the art dynamic sign language recognition (DSLR) framework for savvy home intuitive applications. In their model the

utilized the pack of-elements (BOFs) and a neighborhood part display approach for exposed hand dynamic motion acknowledgment from a video.

P. V. V. Kishore [2] proposed a framework to consequently perceive signals of communication through signing from a video stream of the underwriter. This framework changes over words and sentences of Indian communication via gestures into voice and content in English.

R. Kagalkar [3] displayed survey on an alternate techniques adopted to diminish boundary of communication by building up an assistive gadget for hard of hearing deaf-mutepersons. The headway in installed frameworks, gives a space to plan and build up a communication through signing interpreter framework to help the dump individuals, there exist various associate apparatuses.

V. Ekde and R. Kagalkar [4] described review on video content study into content depiction. Accordingly, this paper displayed three vital commitments to movement acknowledgment from video. Firstly, they presented a solitary system for consequently finding videos activity categories from natural-language descripters. Furthermore, a current movement recognition plan is enhanced abuse object setting alongside relationships amongst objects and activities. At long last, indicates language procedure is consequently extricating the imperative information about the relationship amongst objects and activities from a corpus of general content.

F. Shi et al. [5] proposed 3-D multi-scale parts model, which preserved the requests of events. The model has a coarse primitive level spatiotemporal (ST) highlight, and in addition word covering and event content measurements. This introduced model has higher determination covering parts that can consolidate temporal relations.

A. El-Sawah et al. [7] displayed a model for tracking of 3-D hand and detection of dynamic hand signal. Utilizing numerous cameras marginally upgrades the exactness however mainly enhances the coherence of the information by expanding the working region. The framework handles irregular occlusion however does not handle occlusion for amplified timeframes.

N. H. Dardas and N. D. Georganas [8] proposed a System for exposed hand recognition and tracking in existence of cluttered background utilizing method, for example, skin identification and calculation of hand stance contour correlation after subtraction of face from picture, for hand signals rearrangement through bag-of-features and multiclass support vector machine (SVM) and construct a linguistic use that used to produce gesture commands for an application control. The framework can accomplish attractive continuous execution paying little mind to the casing determination size and also high characterization exactness.

A. R. Varkonyi-Koczy and B. Tusor [9] described a hand position showing and movement showing with hand movement acknowledgment system. This acknowledgment framework can be utilized as an interface to make correspondence (if conceivable) with the keen environment by basic hand motions. Their framework can characterize hand motions that comprise of any blend of the beforehand characterized hand stances and in addition distinctive straightforward hand stances. The real usage of the Gesture Detector does not consider the position of the hand, just the state of it.

M. R. Abid et al. [11] extended video and voice acknowledgment structure for component correspondence by means of motions recordings in home instinctive applications. A nearby part demonstrate approach and Bag-of-Features for acknowledgment of fundamental component signal from the video use a thick looking at technique to focus whole part highlights neighborhood 3D multiscale and grasped three-dimensional histograms of an incline presentation (3D HOG) descriptor to address speak to highlights. They connected the k-means++ technique to bunch the elements. The framework does not join dynamic gesture based communication and voice acknowledgment into one application.

S. Shiravandi et al. [12] analyzed a strategy for acknowledgment of hand sign from video using dynamic Bayesian frameworks. The precision of a model increases because of a use of two systems which are equivalent with motion sorts. At the point when two comparable stances are shot in various directions, they have the distinctive histogram of heading.

T. Wenjun et al. [13] introduced hand development headings approach and conditions of hands philosophy from the key edges. Their introduced methodology is prepared to do adequately perceiving the dynamic hand signal. The shading based calculations confront the troublesome undertaking of recognizing protests, for example, the human arm and the face scene.

J. Bao et al. [14] proposed a Speeded-Up-Robust Features (SURF) following method for component hand flag or motion acknowledgment. To perceive a dynamic signal and to accelerate different estimations, the information stream bunching technique that bolstered connection examination is created. They accept that client ought to stop the video catching for some time before the important hand motion begins and after it closes. Furthermore, the movement direction course is utilized for the representation of motion.

Z. Yang et al. [15] presented a HMM-based technique for acknowledgment of complex single hand signals. For their investigations the catch motion pictures utilizing the web camera. Skin shading is utilized to fragment hand territory from the picture to shape a hand picture succession. Highlights utilized as a part of the framework contain hand position, speed, size, and shape. This framework not equipped for perceiving components to depict the development of fingers.

P. K. Pisharady and M. Saerbeck [16] showed solid hand signal area besides proposed acknowledgment count using dynamic mutilating of time and estimation of multiclass probability. The progressive thresholding of the signal likelihood has been utilized to recognize motions and separation of twisting. The multi-class likelihood estimation has been utilized for signal arrangement. The directional components are to be reached out to an arrangement of precise elements, measuring the edges between different human body links.

III. CHALLENGES OF SIGN LANGUAGE RECOGNITION SYSTEM

Sign language communication understanding comprises of semantic examination of hands following, hands shapes, hands introductions, sign verbalization furthermore with imperative etymological data spoke with head developments and outward appearances. Gesture-based

communication is from various perspectives diverse structure talked dialect, for example, facial and hand phrasing, references in virtual marking space, and syntactic contrasts as clarified.

The significant trouble in gesture-based communication acknowledgment contrasted with discourse acknowledgment is to perceive all the while diverse correspondence properties of an underwriter, for example, hands and head development, outward appearances and body posture. All these attribute must be considered all the while for a decent acknowledgment framework. The second significant issue confronted by gesture-based communication acknowledgment framework architects is following the signer in the jumble of other data accessible in the video. This is tended to by numerous analysts as marking space. A noteworthy test confronted by scientists to characterize a model for spatial data containing the elements made amid the communication through signing discourse.

Extra troubles emerge as background in which signer is found. A large portion of the strategies grew so far use basic foundations in controlled set-up, for example, straightforward foundations, uncommon equipment like information gloves, confined arrangements of activities, limited number of signers, coming about various issues in gesture based communication highlight extraction.

IV. PROPOSED SYSTEM ARCHITECTURE

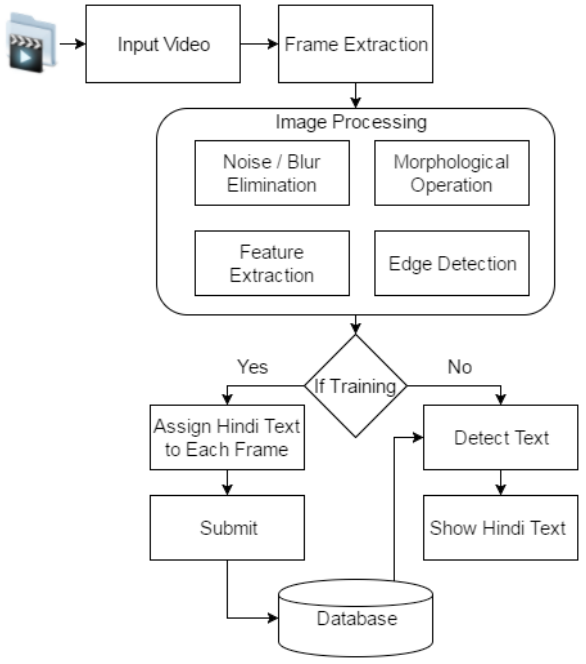


Figure 1. Proposed Architecture

The system design is based on four broad issues related as video preprocessing, image segmentation, feature extraction and pattern classification. This paper proposed a framework which is familiar with gestures of sign language from a video stream of the signer. The developed scheme converts words and sentences of Indian sign language into text in Hindi. Video is the frequent set of images therefor image processing is core part of our system. To achieve much better performance, various image processing techniques are used which can be listed as: frame differencing based tracking, edge detection, image fusion, image segmentation, dilation, erosion, techniques to section shapes in our videos. The proposed system consists of two major modules viz. Training and Testing is shown in Fig. 1. Before test system authority user must train the videos into database.

A. Training Section

The training section is used to train videos and stored on the database with its features which need for video testing. In this section user who is accountable for data training gives input sign video to system. Then the video is processed by means of frame extraction i.e. image generation processing. These frames are nothing but images since the video is a set of continuous images. This process in performed by extracting images with particular time interval such that frame obtained after every second etc.

$$V = \sum_{i=0}^{m} f_i(x, y)$$

Where, $V \rightarrow inputVideo$, f_i is a frame at i^{th} location, $f(x, y)$ denotes input image.

After that convert RBG (Red, Green, and Blue) color extracted frames in step 1 into Grayscale images/frames by eliminating the hue and saturation information and retaining the luminance.

Next, every is image processed out using edge detection, filtering, segmentation, and feature extraction. Image filtering process converts color image into gray by removing noise pixel values from it using Gaussian Filtering technique. It uses Gaussian function

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

Where, x is the distance from the origin in the horizontal axis, y s the distance from the origin in the vertical axis, σ is slandered deviation.

Filtered Image →

Edge detection is applied to identify shape existing in the image. Canny edge detection fused Wavelet based video object separation have turn out to be an essential part of achieving improved segmentation since it combines the essential qualities of the canny operator and two-dimensional wavelet transform.

Edge Image →

In segmenting videos dilation and erosion are used in combination to yield a binary gradient image previously applying discrete wavelet transform. Apply Image segmentation by performing edge detection algorithm is proposed based on morphology, the canny edge detector, and wavelet transform.

1. Image is smooth by convolution function as:
$$f_s(x,y) = G(x,y) * f(x,y)$$
2. Calculate the gradient magnitude as:
$$M(x,y) = \sqrt{g_x^2 + g_y^2}$$
3. Calculate the direction (angle) as:
$$\alpha(x,y) = \arctan(g_y / g_x)$$
Where, $g_x = \partial f_s / \partial x$ and $g_y = \partial f_s / \partial y$
4. Compute direction dx from $\alpha(x,y)$ and find points closest to $M(x,y)$ along direction
5. If $M(x,y)$ is less than at least one of the neighbor then compute suppression
$$g_N(x,y) = M(x,y) = 0$$
6. Else $g_N(x,y) = M(x,y)$
7. Reduce false edge points using Hysteresis thresholding as:
$$g_{NH}(x,y) = g_N(x,y) \geq T_H$$
$$g_{NL}(x,y) = g_N(x,y) \geq T_L$$
And
$$g_{NL}(x,y) = g_{NL}(x,y) - g_{NH}(x,y)$$
8. If all non-valid pixel in $g_{NH}(x,y)$ have been visited then set 0 to all pixel in $g_{NL}(x,y)$ that are never marked as valid.
9. For Morphological operation applies Dilation and Erosion.

- Dilation process enlarges (expands) the image. Rule for this process is if pixels beyond the image border are assigned the minimum value afforded by the data type.
$$A \oplus B = \bigcup_{b \in B} A_b$$
Where, A is any gray scale shape, B is symmetric structuring element. A_b is the translation of A by b.
Dilation is commutative operation therefore it also given by:
$$A \oplus B = B \oplus A = \bigcup_{a \in A} B_a$$
- Whereas Erosion process shrinks the image. Rule for this process is Pixels beyond the image border are assigned the maximum *value afforded by the data type.
$$A \circ B = (A \Theta B) \oplus B = \bigcap_{b \in B} A_{-b}$$

The feature extraction utilizing SIFT procedure will see shape of an image. During feature extraction process, search the invariance parameters so that the extraction method does not contrast as indicated by measured circumstances. In particular, methods utilized for extraction of feature ought to discover shapes dependable and powerfully irrespective of adjustments in illumination position, levels, size, orientation and description of the object in a video. The separated key points are scale invariant, introduction and decently invariant to brightening changes, and are amazingly particular of the picture. Thus, the SIFT is acknowledged in this paper for the recognition of sign hand motion.

SIFT Features of Image →

After all this process, images of video with corresponding extracted features and their assign Hindi text are stored into the database.

B. Testing Section

This module tests the video of sign learner and gets the result only if at slightest one video is trained. In this phase, all process on images is performed same as in training phases such as filtering, edge detection, segmentation and feature extraction. After recognition of testing video, it sends to further process. The last process of the system is a classification of different signs in the form of text description in Hindi language corresponding to the correctly classified sign. For this purpose of sign language recognition system, the proposed system has deployed a fuzzy inference mechanism by using if-else conditions to detect corresponding Hindi text of video. The if-then rule based fuzzy classification frameworks depends on the selection procedure of fuzzy divider.

V. RESULTS

In order to evaluate Hindi text extraction process from sign videos, 100 sign videos are used to train and store related Hindi text into database. In the presented scenario, the testing video is previously trained so that the presented tests already include some gestures with quite similar features to get the best result.

While training each frame is assign by tag or Hindi text as per sign in that frame. For result analysis, this paper considers the training section and testing section. Some results are predicted using the dependent fuzzy technique. For the dependent system, accurate results are 100% because database consists of the trained video that gives as testing. Video is processed as per given in section 4. For our experiment we divide videos into multiple categories viz. Months, Weeks, information, office, relations etc. as shown in table 1. These videos are trained before testing and stored into database with their corresponding features and Hindi text.

Some of the outputs detected from videos are shown in table II. In this input video text is the original text for that video and output text is the predicted text from implementation. Video frames are depends on size of input video and the frame extraction time interval.

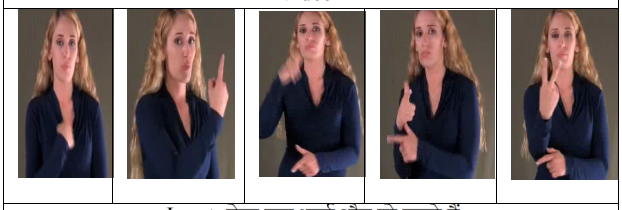
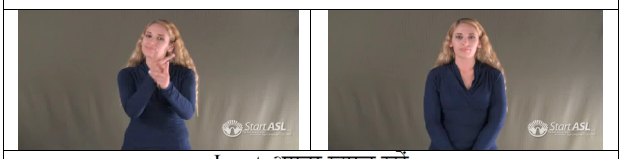
For videos processing, when frames are processed from image filtering and edge detection, then the time require to system is shown into following graph figure 2.

Features are extracted from each frame to recognize objects or and shapes from frame. Time require for this process is shown in figure 3.

TABLE I. DATASET DESCRIPTION

Sr. No.	Database Description	
	Category	*Number of Videos*
1	Week Days	7
2	Months	12
3	Relations	20
4	Office	14
5	Information	36
6	Other	50

TABLE II. OUTPUT PREDICTION

Video1

Input: मेरा एक भाई और दो बहने हैं Output: मेरा एक भाई और दो बहने हैं
Video2

Input: अपना ख्याल रखें Output: अपना ख्याल रखें
Video3

Input: सोमवार Output: सोमवार
Video4

Input: कार्यालय Output: कार्यालय

VI. CONCLUSION

This paper shows a procedure for recognition of dynamic sign language video into Hindi content dialect which might be words or sentences. This paper utilizes the systems of picture handling and engineered knowledge to get a precise result. For the execution of this system, it utilizes image processing procedures, for example, frame/ image extraction from videos as per time, erosion, dilation, edge detection, blur elimination, noise elimination, wavelet transform, image fusion techniques to area shapes in sign videos. It additionally utilizes SIFT highlight extraction method and most essential examination for the arrangement of elements decrease and in addition optimization. The features of video are tested base on fuzzy rule interface. In future, independent approach with highest accuracy will consider for study.

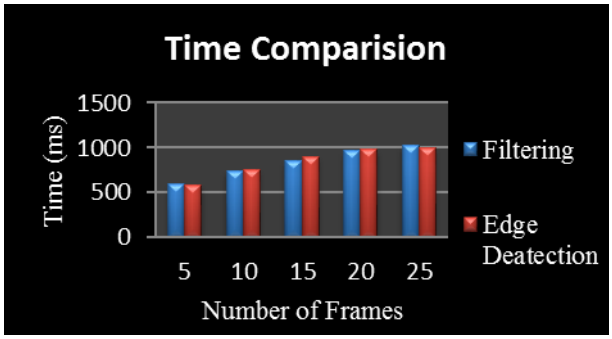


Figure 2. Time comparison for filtering and edge detection process

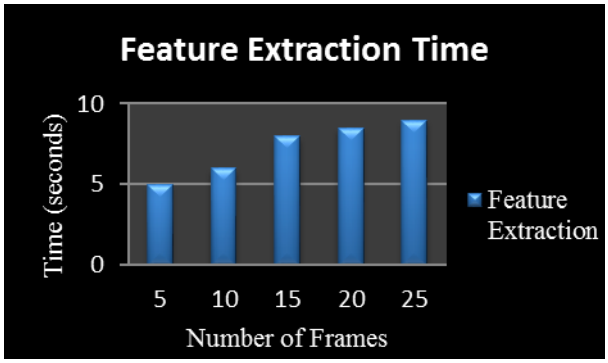


Figure 3. Time comparison for feature extraction process

REFERENCES

[1] M. Rizwan Abid, E. M. Petriu and E. Amjadian, "Dynamic Sign Language Recognition for Smart Home Interactive Application Using Stochastic Linear Formal Grammar", IEEE Transactions On Instrumentation And Measurement, vol 64, no. 3, Mar 2015

[2] P.V.V.Kishore, P.Rajesh Kumar, E.Kiran Kumar, and S.R.C.Kishore, "Video Audio Interface for Recognizing Gestures of Indian Sign Language", International Journal of Image Processing (IJIP), Volume (5) ,Issue (4), 2011.

[3] Rashmi B. Hiremath and Ramesh M. Kagalkar," Review Paper on Sign Language Recognition Techniques", International Journal of Computer Applications National Conference on Advances in Computing , 2015.

[4] Vandana D. Edke and Ramesh M. Kagalkar," Review Paper on Video Content Analysis into Text Description", International Journal of Computer Applications National Conference on Advances in Computing, 2015.

[5] F. Shi, E. M. Petriu, and A. Cordeiro, "Human action recognition from local part model," in Proc. IEEE Int. Workshop Haptic Audio Vis. Environ. Games (HAVE), Oct. 2011, pp. 35–38. http://dx.doi.org/10.1109/have.2011.6088408

[6] A. Chaudhary, J. L. Raheja, K. Das, and S. Raheja, "A survey on hand gesture recognition in context of soft computing," in Proc. CCSIT, vol. 133. Jan. 2011, pp. 46–55. http://dx.doi.org/10.1007/978-3-642-17881-8_5

[7] A. El-Sawah, N. D. Georganas, and E. M. Petriu, "A prototype for 3-D hand tracking and posture estimation," IEEE Trans. Instrum. Meas., vol. 57, no. 8, pp. 1627–1636, Aug. 2008. http://dx.doi.org/10.1109/TIM.2008.925725

[8] N. H. Dardas and N. D. Georganas, "Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques," IEEE Trans. Instrum. Meas., vol. 60, no. 11, pp. 3592–3607, Nov. 2011. http://dx.doi.org/10.1109/TIM.2011.2161140

[9] A. R. Varkonyi-Koczy and B. Tusor, "Human–computer interaction for smart environment applications using fuzzy hand posture and gesture models," IEEE Trans. Instrum. Meas., vol. 60, no. 5, pp. 1505–1514, May 2011. http://dx.doi.org/10.1109/TIM.2011.2108075

[10] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, pp. 1–27, 2011. http://dx.doi.org/10.1145/1961189.1961199

[11] M. R. Abid, L. B. S. Melo, and E. M. Petriu, "Dynamic sign language and voice recognition for smart home interactive application," in Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA), May 2013, pp. 139–144.

[12] S. Shiravandi, M. Rahmati, and F. Mahmoudi, "Hand gestures recognition using dynamic Bayesian networks," in Proc. 3rd Joint Conf. AI Robot. 5th RoboCup Iran Open Int. Symp. (RIOS), Apr. 2013, pp. 1–6. http://dx.doi.org/10.1109/rios.2013.6595318

[13] T. Wenjun, W. Chengdong, Z. Shuying, and J. Li, "Dynamic hand gesture recognition using motion trajectories and key frames," in Proc. 2nd Int. Conf. Adv. Comput. Control (ICACC), Mar. 2010, pp. 163–167.

[14] J. Bao, A. Song, Y. Guo, and H. Tang, "Dynamic hand gesture recognition based on SURF tracking," in Proc. Int. Conf. Elect. Inf. Control Eng. (ICEICE), Apr. 2011, pp. 338–341. http://dx.doi.org/10.3724/sp.j.1218.2011.00482

[15] Z. Yang, Y. Li, W. Chen, and Y. Zheng, "Dynamic hand gesture recognition using hidden Markov models," in Proc. 7th Int. Conf. Comput. Sci. Edu. (ICCSE), Jul. 2012, pp. 360–365. http://dx.doi.org/10.1109/iccse.2012.6295092

[16] P. K. Pisharady and M. Saerbeck, "Robust gesture detection and recognition using dynamic time warping and multi-class probability estimates," in Proc. IEEE Symp. Comput. Intell. Multimedia, Signal Vis. Process. (CIMSIVP), Apr. 2013, pp. 30–36.

AUTHORS

Rashmi B. Hiremath is M.E 2nd year student of Computer Engineering Department, Dr. D. Y. Patil School of Engineering and Technology, Pune,Maharashtra,India. Her main research interest includes Image processing and Gesture recognition. She can be contacted by email hiremathr709@gmail.com.

Ramesh M. Kagalkar was born on Jun 1st, 1979 in Kar- nataka, India and presently working as an Assistant. Professor, Department of Computer Engineering, Dr. D.Y.Patil School Of Engineering and Technology, Charoli, B.K.Via Lohegaon, Pune, Maharashtra, India. He has 14 years of teaching experience at various institutions. He is a Research Scholar in Visveswaraiah Technological University, Belgaum, He had obtained M.Tech (CSE) Degree in 2006 from VTU Belgaum and He received BE (CSE) Degree in 2001 from Gulbarga University, Gulbarga. He is the author of text book Advance Computer Architecture which cover the syllabus of final year computer science and engineering, Visveswaraiah Technological University, Belgaum. One of his research article A Novel Approach for Privacy Pre- serving has been consider as text in LAP LAMBERT Academic Publishing, Germany (Available in online). He is waiting for submission of two research articles for patent right. He has published more than 35 research papers in International Jour- nals and presented few of there in international conferences. His main research interest includes Image processing, Gesture recognition, speech processing, voice to sign language and CBIR. Under his guidance Ten ME students awarded degree in SPPU, Pune, five students at the edge of completion their ME final dissertation reports and two students started are started new research work and they have publish their research papers on International Journals and International conference. He can be contacted by email rameshvtu10@gmail.com.

Submitted 21 June 2016. Published as resubmitted by the authors 16 September 2016.