Computers in Industry 99 (2018) 313–323

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

ABC algorithm based optimization of 1-D hidden Markov model for hand gesture recognition applications

K. Martin Sagayam, D. Jude Hemanth*

Department of ECE, Karunya University, Coimbatore, India

ARTICLE INFO

Keywords:
HCI
Virtual reality
1-D HMM
Baum-Welch algorithm
Viterbi algorithm
ABC algorithm
Cambridge hand gesture

ABSTRACT

Hand gestures are extensively used to communicate based on non-verbal interaction with computers. This mode of communication is made possible by implementing machine learning algorithms for pattern recognition. A stochastic mathematical approach is used to interpret the hand gesture pattern for classification. In this work, a predominant method is used by 1-D hidden Markov model (1-D HMM) for classifying the patterns and to measure its performance. During training phase, 1-D HMM is used to predict its next state sequence of hand gestures using dynamic programming methods such as Baum-Welch algorithm and Viterbi algorithm. However, dynamic programming based prediction methodologies are complex. To enhance the performance of 1-D HMM model, its parameter and observation state sequence must be optimized using bio-inspired heuristic approaches. In this work, Artificial Bee Colony (ABC) algorithm is used for optimization. A hybrid 1-D HMM model with ABC optimization has been proposed which has yielded a better performance metrics like recognition rate and error rate for Cambridge hand gesture dataset.

1. Introduction

Over the past decade, hand gesture and posture recognition have turned out to be a popular research in the field of pattern recognition, signal processing, computer vision and machine learning. Integrating gestures with machines is mostly preferred for human-computer interaction (HCI). The recent computer vision based technologies have made the system more realistic using gesture patterns. In numerous applications, hand gestures are the commonly used tool for interfacing field of virtual games, medical, physically impaired people, education, teleconferencing, etc. The hand gesture pattern supports the system in a user-friendly manner for handling non-verbal communication with the system. The various methods used to develop the virtual system based on hand gesture and posture recognition have been surveyed by K. M. Sagayam and D. J. Hemanth [1]. E. Stergiopoulou [2] and Hongo et al. [3] have discussed the skin color detection using the region of interest (ROI) method for hand gesture recognition. S. Bilal et al. [4] have presented a detailed survey report on hand gesture recognition using hidden Markov model (HMM). This report has adapted direct trajectory shape or feature matching for sign language (SL) recognition. The basic feature points of gesture pattern are derived using linear discriminant analysis (LDA). There are three major steps involved in the system, according to Manresa et al. [5] (i) Segmentation of hand contour using skin-color detection, (ii) Position determination using pixel-based tracking method, and (iii) State estimation of hand gesture model based on the number of classes defined. The extension of the above work has been done by Yoon et al. [6]. It comprised of three modules: (i) tracking, (ii) localization, and (iii) spotting. The tracking of hand gesture is based on the shape based descriptor for determining the geometrical features like centroid, line, circle, triangle and velocity. Similarly, the localization algorithm of gesture recognition has helped to determine position with the skin-color detection techniques. Such values can be determined by K-mean clustering algorithm and the resultant features are tested using hidden Markov model (HMM). The above literature has not focused on the optimization techniques of the gesture patterns.

1.1. Related works

Hand gesture recognition is the most predominant research in various applications. Peter Sykora et al. [7] have discussed the depth of hand gesture data with 1500 sets for training and testing using support vector machine (SVM) algorithm. The feature point can be extracted using scale-invariant feature transform (SIFT) and speeded-up robust transform (SURF) features. The validation of hand gesture data is done by dividing 100 frames per class for training and 50 frames per class for testing process. A real-time classified output using SIFT and SURF is 81.2% and 82.8% respectively. An idea of reducing the computational

* Corresponding author.
E-mail addresses: martinsagayam.k@gmail.com (K.M. Sagayam), judehemanth@karunya.edu (D.J. Hemanth).

https://doi.org/10.1016/j.compind.2018.03.035
Received 21 June 2017; Received in revised form 22 March 2018; Accepted 27 March 2018
0166-3615/

complexity of HMM algorithm was proposed by M. Mohandes et al. [8]. It produced 95% accuracy rate for sign dependent and independent condition for Arabic sign language (ArSL) recognition system. The incremental learning hidden Markov model (ILHMM) for dynamic hand gesture recognition system has been proposed by Meng Hu et al. [9]. An Arabic number from 0 to 9 of 10 different classes are used as an input video stream of 30 frames per second. Each image frame has dimension of 480 × 360. Arabic numbers are recognized by unique hand gesture of the system with depth camera data and inferred sensor data. An experimental result shows the comparison of recent methods in terms of accuracy, timeliness consumption, and recognition rate. The system has achieved 95.1% of accuracy rate for hand gesture recognition using incremental learning method. The depth camera is used for handwriting recognition suggested by Necati Cihan Camgoz et al. [10]. It is difficult to acquire hand trajectories of different background conditions from the depth camera. This system has been adopted with feature augmented method for evaluation of target class performance measure. Yona Falinie A. Gaus et al. [11] have emphasized new method based on feature points such as angular position, distance and velocity of hand gesture. It has been implemented using stochastic mathematical approach called hidden Markov model (HMM). The static hand gestures have constant sequence in its state machine. The dynamic gestures has variable form in HMM model. The accuracy rate of left and right hand gesture recognition is 83.1% respectively. A new method has been proposed by Radu-Laurentiu Vieriu et al. [12] to overcome the problem in static gesture recognition. Subsequently, another related work was demonstrated by Jijun Rao et al. [13] using 3-pivot MEMS accelerometer with less expenditure for training and testing using HMM approach. The experimental results yield higher recognition rate for possible implementation in real-time applications. Qingqing Huang et al. [14] have explained the structure of HMM model in which the output of state sequence is obtained for all possible state variables.

Table 1 shows the recognition accuracy (RA) of real-time gesture data of other stochastic methods from the year 2008 to 2013. The 1-D HMM illustrates low quality results than the higher versions since it consumes more execution time for training and testing phase as shown in Janusz Bobulski [17] and K.M. Sagayam et al. [22]. The hand posture and gesture recognition using HMM approach were presented by A. Sachin et al. [18]. It produces 82.5% recognition rate for 118 hand gesture samples with a uniform background. The proposed system is used to enhance the performance measures of 1-D HMM.

2. Proposed system

The proposed system comprises of five major modules: hand gesture data set, image preprocessing, feature extraction, optimization and classification as shown in Fig. 1.

The Cambridge hand gesture data of nine different classes have to be pre-processed with the suitable edge detection technique. The feature points based on Zernike moment (ZM) are extracted from these images. The feature vector points of various classes must be labeled. It has to be trained and tested with 1-D HMM. The recursiveness is observed in both iterative approaches due to the self-looping. This leads increased computational time. Thus, a meta-heuristic approach has been proposed in this work called artificial bee colony (ABC) optimization to overcome the complexity of 1-D HMM. The subsections describe the mechanism of all block in detail.

3. Hand dataset and pre-processing

In this work, publicly available datasets from Cambridge hand gesture data are used. This dataset consist of images with following varieties: flat to left, flat to right, flat to contract, spread-left, spread-right, spread-contract, V-shape left, V-shape right and V-shape contract [2]. Each class contains 100 frames of nine hand gesture, which includes three characteristic hand shapes (“flat”, “spread”, and“V-shape”) and three characteristic motions (left, right and contract). There are 900 gestures with 320 × 240 dimensions available in the dataset. The sample hand gesture data from Cambridge hand gesture data is shown in Fig. 2.

Fig. 3 shows the hand gesture movement from flat to left position at 0th, 15th, 30th, 45th and 60th frame. These gesture patterns show the presence or absence of shadows which may lead to the wrong prediction of edge points.

3.1. Image pre-processing

The phenomenon of transferring original data into another format without mixing of unnecessary frequency components is termed as pre-processing. This process is carried out by removal of noise, background subtraction, and edge detection techniques. The edge detection technique is predominantly used to locate the contour of an object in an image. This was fixed by the point of interest by changing the image brightness in the curved line portion called as edges. It has been categorized into two types: (i) first-order Gradient-based and (ii) second-order Laplacian-based approach. Gradient-based approach is further classified into three types namely, Sobel, Prewitt, and Robert operator. Similarly, Laplacian-based approach is further classified into two types namely Laplacian of Gradient (LoG) and Canny edge operator [36,37]. Some pros and cons of edge detectors are shown in Table 2.

The main objective of detecting the edges of an image is to get the original content without loss of information [5]. The change in intensity level over an edge gives the gradient value of an image. If the gradient value is very high, the edges of an image are clear. If the gradient value is very low, then the edges do not occur clearly. The estimated filtered value with magnitude and direction of the gradient are located in the horizontal (x-axis) and vertical (y-axis). Hand gestures are clearly emphasized by Sobel-Feldman operator or Sobel operator. The true value

Table 1
Survey on recognition accuracy of real-time gesture data for other stochastic methods.

Method	Year	Dataset	RA rate (%)
H. Ragheb et al., [24]	2008	ViHASi	72.00
C.-C.Chen & J. Aggarwal, [23]	2009	UT-Tower	90.43
M.S. Ryoo & J.K.Aggarwal, [25]	2009	UT-interaction	91.67
J.C. Niebles et al., [26]	2010	Olympic sports	91.10
A. Patron-Perez et al., [27]	2010	TV human interaction	46.00
H. Kuehne et al., [28]	2011	HMDB51	57.20
G. Denina et al., [29]	2011	Video web	72.00
K. Soomro et al., [30]	2012	UCF-101	83.50
K.K. Reddy & M. Shah, [31]	2013	UCF-50	91.20
M.M. Gharasuie, H. Seyedarabi, [32]	2014	Numbers (0–9)	93.84
H. Kim & I. Kim, [33]	2015	Kyonggi dataset-240 gestures	86–95 (4–15% Improved)
Archana Ghotkar et al., [34]	2016	ISL (20 dynamic signs)	89.25
Pablo Barros et al., [35]	2017	RPPDI dynamic hand gesture + Cambridge hand data	71.33–93.33