ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

Hand Gesture Recognition using DTW and Morphological Feature Extraction

Sindhu. K. M, Suresha.H.S

Dept. of E&C, Don Bosco Institute of Technology, Bangalore, India

Associate Professor, Dept. of E&C, Don Bosco Institute of Technology, Bangalore, India

ABSTRACT: This article proposes an automatic gesture recognition approach for sign language. Sign language uses both hands to represent each alphabet. The proposed approach that addresses the identification of local-global ambiguity, improves inter-class variability for each hand gesture. The hand region is segmented and detected by the reference of the YCbCr skin color model. The shape, texture and characteristics of the fingers of each hand are extracted using the algorithms Bag of contingency fragmentation and Morphological operation defects respectively for the process of recognition of the posture of the hand. In order to classify each hand position, nonlinear vector support machines of various kinds (SVM) are used, for which a recognition rate of 90% is achieved. The dynamic gestures are classified using Dynamic Time Warping (DTW) with the vector of trajectory characteristics with a recognition rate of 86.3%. The performance of the proposed approach is analyzed with classifiers well known as SVM. The experimental results are compared with conventional and existing algorithms to test the best efficiency of the proposed approach.

I. INTRODUCTION

India is diversified in culture, language and religion. Since there is a great diversity among Indian languages, the literature survey reports the non-existence of standard forms of Indigenous Sign Language (LSL) gestures. ISL alphabets are derived from British Sign Language (BSL) and French Sign Language (FSL). Because of these problems, the standard database for the ISL / gesture alphabet has not been developed so far. Few research works has been carried out on ISL recognition and interpretation through image processing / vision techniques. But these are only initial jobs proven with simple image processing techniques and are not treated with real-time data. The classification technique refers to the Euclidean distance metric. Subsequently we propose a system to translate the input speech to ISL that is shown with the help of a 3D virtual human avatar. The input to the system is the speech of the employee who is in English. The speech recognition module recognizes speech and performs a text output. This text is then passed to a parser module that tokenizes the string and labels the part of the voice using a sample file. The output of the analyzer is given to an eliminator module that performs a reduction task by removing unwanted elements and also the root form of the verbs are found using the stemmer module. The structural divergence of English and ISL is handled by a phrase reordering module using the ISL dictionary and the rule. This module generates ISL brightness strings that can be reproduced through virtual human 3D. A 3D animation module creates animation from motion-captured data. In this approach a lot of 3D model data is used which makes the system clumsy and bulk. Attempt to machine static translation as well as dynamic ISL gestures with image processing features such as skin tone detection space filter velocimetry and temporal tracking is developed. The representation of the power spectrum of each gesture is given as moving images. Edge detection, cropping and boundary tracking are used as characteristics for the recognition process. These methods work well for the static signs of ISL. They do not deal with the dynamic, global and local movements of ISL gestures. For example, the ISL signs of the letters A-B, M-N, U-V look similar. It is sometimes difficult for a human to correctly recognize the sign. When it comes to computers, the inter-class variability parameter must be considered.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

II. PROPOSEDMETHOD

The proposed method consists of the following steps: Segmentation and hand detection of each video frame using the skin color model, which is followed by extraction of the hand feature using the Morphological operation and the Fragmentation Bag contingency. At the same time, to extract the local movements in each the counting of fingers is taken into account. The extracted features become vectors of appropriate characteristics. For the classification of the gestures of each hand (Alphabets) we use nonlinear support vector machines of several kinds (SVM).

- Segmentation and hand detection from video

First, preprocessing and normalization are performed in the frames of video objects. Skin color segmentation is performed in the YCbCr color space, as it reduces the effect of uneven illumination on an image. YCbCr is a nonlinear RGB signal encoded with simple transformation; The explicit separation of luminance and chrominance components makes this color space attractive for modeling skin color. The RGB color frames I (m, n, p) (where m, n and pare the number of rows, number of columns and number of color planes) are converted into YCbCr images using

$$Y = 0.299R + 0.587G + 0.114B$$

A parametric method called Single Gaussian Model for skin color uses mean and covariance of chrominant color with a bivariate Gaussian distribution with c as color vector representing random measured values of chrominance (x,y) of a pixel with coordinates (i,j) in an image. Ws is the class describing the skin. Iis is the mean vector and the covariance matrix for skinchrominance.
The input image frame color vector is compared with the stored skin color model data and the pixels representing minimum distance is considered as skin. Those skin pixels are converted back to RGB space. By considering the skin region threshold value hand is detected in each frame. Now the detected hand image is given by D (i,j).

- DTW

In the time series analysis, the dynamic time deformation (DTW) is an algorithm to measure the similarity between two temporal sequences that can vary in speed. For example, similarities in walking could be detected using DTW, even if one person walked faster than the other, or if there were accelerations and decelerations during the course of an observation. DTW has been applied to temporal sequences of video, audio and graphic data - in fact, any data that can be converted into a linear sequence can be analyzed with DTW. One well-known application has been automatic speech recognition, to cope with different speech speeds. Other applications include speaker recognition and online signature recognition. It is also seen that it can be used in the partial application of the form of adaptation.
In general, DTW is a method that calculates an optimal match between two given sequences (for example, time series) with certain constraints. The sequences are non-linearly "deformed" in the time dimension to determine a measure of their similarity independent of certain non-linear variations in the time dimension. This method of sequence alignment is often used in the classification of time series. Although DTW measures an amount similar to the distance between two given sequences, it does not guarantee that the inequality of the triangle is maintained.In addition to a measure of similarity between the two sequences, a so-called "deformation trajectory" occurs, by deforming according to this trajectory the two signals may be aligned over time. The signal with an original set of X (original), Y (original) points becomes X (deformed), Y (original). This finds applications in genetic sequence and audio synchronization. In a related art, variable speed sequences can be averaged using this technique, see the mean sequence section.

- Morphological

Mathematical morphology (MM) is a theory and technique for the analysis and processing of geometric structures, based on set theory, lattice theory, topology and random functions. MM is most commonly applied to digital images, but it can also be used in graphics, surface meshes, solids, and many other spatial structures. The concepts of continuous topological and geometric space such as size, shape, convexity, connectivity and geodesic distance were introduced by MM in continuous and discrete spaces. MM is also the basis of morphological image processing, which consists of a set of operators that transform images according to the previous characterizations. The basic morphological operators are erosion, dilation, opening and closing. MM was originally developed for binary images, and was later extended to grayscale functions and images. The subsequent generalization to complete lattices is

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

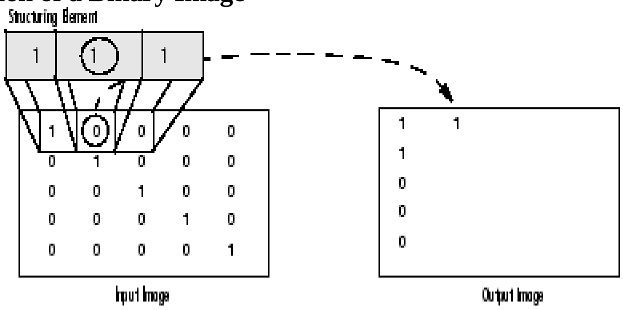
(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

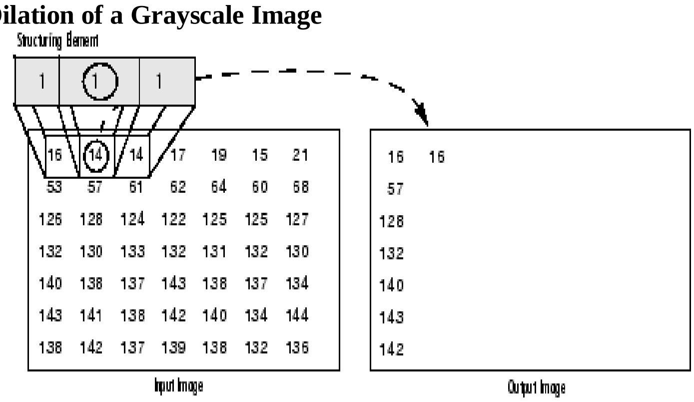
widely accepted today as the theoretical foundation of MM. Morphology is a comprehensive set of image processing operations that process images based on shapes. Morphological operations apply a structuring element to an input image, creating an output image of the same size. In a morphological operation, the value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbors. By choosing the size and shape of the neighborhood, you can build a morphological operation that is sensitive to specific shapes in the input image. The most basic morphological operations are dilation and erosion. Dilation adds pixels to the boundaries of objects in an image, while erosion removes pixels from the boundaries of objects. The number of pixels added or removed from objects in an image depends on the size and shape of the structuring element used to process the image. In morphological dilation and erosion operations, the state of any pixel given in the output image is determined by applying a rule to the corresponding pixel and its neighbors in the input image. The rule used to process the pixels defines the operation as a dilation or erosion. This table lists the rules for expansion and erosion.
The following figure illustrates the dilation of a binary image. Notice how the structuring element defines the surrounding pixel environment of interest. The dilation function applies the appropriate rule to the neighborhood pixels and assigns a value to the corresponding pixel in the output image. In the figure, the morphological dilation function sets the output pixel value to 1 because one of the elements in the neighborhood defined by the structuring element is activated.

- **Morphological Dilation of a Binary Image**



The following figure illustrates this processing for a grayscale image. The figure shows the processing of a particular pixel in the input image. Note how the function applies the rule to the input pixel's neighborhood and uses the highest value of all the pixels in the neighborhood as the value of the corresponding pixel in the output image.

- **Morphological Dilation of a Grayscale Image**



Processing Pixels at Image Borders (Padding Behavior)
The morphological functions position the origin of the structuring element, its central element, on the pixel of interest in the input image. For pixels at the edge of an image, the parts of the neighborhood defined by the structuring element may extend beyond the edge of the image. To process border pixels, the morphological functions assign a value to these undefined pixels, as if the functions would have padded the image with additional rows and columns. The value of these

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

fill pixels varies for dilatation and erosion operations. The following table describes the filler rules for expansion and erosion for binary and grayscale images.

SVM
In automatic learning, vector support machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze the data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or other of the two categories, an SVM training algorithm constructs a model that assigns new examples to one category or another, converting it into a linear non-probabilistic binary classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are mapped in that same space and are predicted to belong to a category based on which side of the gap they fall into.
In addition to performing linear classification, SVMs can perform a nonlinear classification efficiently using what is called a kernel trick, implicitly implying their entries into high-dimensional feature spaces.
When data is not labeled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find the natural grouping of data to groups, and then map new data to these formed groups. The clustering algorithm that provides an enhancement to support vector machines is called vector support clustering [2] and is often used in industrial applications, when data is not tagged or when only some data is labeled as a preprocessing for a Classification pass.

Data classification is a common task in automatic learning. Suppose each of the data points belongs to one of the two classes and the objective is to decide in which class a new data point will be found. In the case of support vector machines, a data point is seen as a dimensional vector (a list of numbers), and we want to know if we can separate such points with a -dimensional hyper-plane. This is called a linear classifier. There are many hyper-planes that can sort the data. A reasonable choice as the best hyper-plane is the one that represents the greatest separation, or margin, between the two classes. So we choose the hyper-plane so that the distance from it to the nearest data point on each side is maximized. If such a hyper-plane exists, it is known as the maximum margin hyper-plane, and the linear classifier it defines is known as the maximum margin classifier; Or, equivalently, the perceptron of optimal stability.

III. RESULTS

Fig 1: Hand gesture

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 5, May 2017

Fig 2: Recognition of hand

IV. CONCLUSION

This paper proposed an approach to address the inter-class ambiguity issue in ISL alphabet recognition. With the help of local-global finger movement information and shape-texture features, accurate recognition of each ISL signs have been achieved for both static & dynamic gestures. However, due to the less stable nature of morphological operations and bag of contingence features, the accuracy slips down slightly in case of dynamic gestures of ISL. Our future prospective work concentrates on the analysis of dynamic nature of gestures under different conditions.

REFERENCES

[1] Giulio Marin, Fabio Dominio and Pietro Zanuttigh,“Hand Gesture Recognition With Leap Motion And Kinect Devices”, IEEE, 2014.
[2] J. Rekbai, J. Bhattacharya and s. Majumder “Shape, Texture and Local Movement Hand Gesture Features for Indian Sign Language Recognition”, IEEE, 2011.
[3] Siddharth S. Rautaray and Anupam Agrawal,“Vision based hand gesture recognition for human computer interaction: a survey”, Spinger, 2015.
[4] Chao Xu, Parth H. Pathak and Prasant Mohapatra,“Finger-writing with Smartwatch: A Case for Finger and Hand Gesture Recognition using Smartwatch”, International Workshop, 2015.
[5] Pavlo Molchanov, Shalini Gupta, Kihwan Kim, and Jan Kautz, “Hand Gesture Recognition with 3D Convolutional Neural Networks”, IEEE, 2015
[6] Molchanov P, Gupta S, Kim K, and Pulli K,“Multi-sensor system for driver's hand-gesture recognition”, Vol. 1, pp. 1-8, IEEE, 2015.
[7] Zhang Y and Harrison C, “Tomo: Wearable, low-cost electrical impedance tomography for hand gesture recognition”, pp. 167-173, IEEE, 2015.
[8] Zhang C and Tian Y, “Histogram of 3d facets: A depth descriptor for human action and hand gesture recognition”, Elsevier, 2015.
[9] Pisharady P K, andSaerbeck M, “Recent methods and databases in vision-based hand gesture recognition: A review”, IEEE, 2015.
[10] Kılıboz N C and Gudukbay U, “A hand gesture recognition technique for human–computer interaction”, Elseveir, 2015.