A Machine Learning Based Approach for Hand Gesture Recognition using Distinctive Feature Extraction.

Himadri Nath Saha
Department of Electronics and Communication Engineering
Institute of Engineering and Management, Kolkata, India
himadri@iemcal.com

Shinjini Ray
Department of Computer Science and Engineering
Institute of Engineering and Management, Kolkata, India
shinjini.ray@gmail.com

Sudipta Saha
Department of Computer Science and Engineering
Institute of Engineering and Management, Kolkata, India
subho040995@gmail.com

Sayan Tapadar
Department of Computer Science and Engineering
Institute of Engineering and Management, Kolkata, India
saytap.tapadar5@gmail.com

Suhrid Krishna Chatterjee
Department of Computer Science and Engineering
Institute of Engineering and Management, Kolkata, India
c.suhrid@gmail.com

***Abstract*—In a world of almost 7 billion people more than 500 million suffer from some physical, sensory or mental disability. Their lives are often impeded by such deformities which bars them from full participation in society and the enjoyment of equal rights and opportunities. Sign language is common for the deaf and the dumb. Sign language is an efficient alternative to talking, where the former is replaced by hand gestures. Hand gestures are combination of hand shapes, orientations and movement of the hands, alignments of the fingers and positioning of the palm which are used to express fluidly a conveyer's thoughts. Signs are used to communicate words and sentences to audience.**

The objective of this paper is to optimize an algorithm for recognition of hand gestures with reasonable accuracy, where the input to the pattern recognition system will be given from the hand.

Possible reference models are already available such as ASL or American Sign Language. Image is collected from a webcam followed by preprocessing. Further segmentation of the figure is done through polygon approximation and approximate convex decomposition. Feature extraction is done by recording the unique feature among the various convex segments of the hand. The resultant singularities are then used as extracted feature vectors. This involves training with the obtained features which are approximately unique for different hand gestures. Thus, we will be able to identify sign languages and successively make disabled individuals socially acceptable.

Keyword: Hand gesture recognition, feature extraction, polygon approximation, convex decomposition, machine learning.

I. INTRODUCTION

The feature-extraction based technique of hand gesture recognition is an emerging section of human-computer interaction (HCI). Gone are the days when keyboard and mouse had a significant stance in human-computer interaction. Burgeoning technology has set the ball rolling in evolution of human interaction with the machine and its incessant possibilities. Human for long has successfully managed to utilize and exploit the very best of what a machine has to offer. With our recent endeavors in image processing techniques and machine learning methods, we have been able to identify relations in serpentine patterns and solutions to convoluted puzzles. This has promoted massive attention towards gesture detection possibilities.

Every individual is entitled to the right to freedom of expression. But some are unfortunate given physical disabilities and mental frailties. Medical innovation has been ceaseless in their contribution to the betterment of such individuals. But, cost and inefficient machines has been major setbacks given the drawbacks of unavailability of adequate technology and pitiable economic conditions in major parts of the world. The advent of sign languages opened up possibilities to communication and seemed to be a potential solution to the problem of inexpressibility. However owing to the miniscule number of deformed individuals, people with knowledge of sign language seems tantamount to none. A machine hosts innumerable possibilities and prospects that are being explored every day. In addition machines are ubiquitous. Hence training a machine to recognize gestures can help those distinguished individuals to interact and be socially accepted.

Gesture in laymen's terms is representation of physical behavior or emotional expression. Two further elaborate, they can be classified into two types, namely - body gesture and hand gesture. Further, they are distinguished on basis of static gesture[] and dynamic gesture[]. The more commonly used hand gesture denotes a sign or a symbol. Gesture or so called sign languages are used as the tool of communication between human to human interactions. Such communications are bound to a specific set of protocols or symbols - the Indian Sign Language (ISL)[] or the American Sign Language (ASL)[]. The aim of the research is to illustrate the extraction of unique features from each symbol or gesture in reference to an existing data set (ISL or ASL) and then train the machine with the obtained feature vectors using standard classification models.

Extraction of unique features and successive training eases the classifiers of the mammoth number of data required to generate patterns. Further it ensures better accuracy and efficiency in output due to the waning of ambiguity in the generated data. Our approach initially follows certain pre-processing steps. We do make use of white gloves to eliminate unwanted wrinkles on the palm and fingers. From the threshold image of the figure we obtain its minimum fitting polygon through what we call polygon approximation technique[]. This is followed by convex decomposition that segregates each of the convex parts of the hand. This is an essential step to identity the crucial aspects of the gesture and is the final step before we proceed to extracting the required unique features. The wrist point is the mid-point of the lowermost polygon. In addition we find a reference line that is parallel to the bottom boundary of the frame and passes through the wrist point. The angle of the mid-points of each segment and there corresponding length in reference with the wrist point is then calculated. We then generate a feature vector of dimension 36 in accordance of partition 0 degree to 180 degree in degrees of 5. In a clockwise fashion we check if a partition contains a line that joins the wrist point and mid-point of a segment. If so we store its calculated length in the n^{th} partition. This forms our feature vector for each gesture before proceeding to training the machine with this generated unique

data. Thus we are able to produce an efficient and effective method for hand gesture recognition.
The novelty of the proposed method is listed as follows.

(i) Distinctive feature extraction for each and every gesture enables significant differentiation and further efficient classification of the obtained data.
(ii) Acknowledging the demand and importance of real time application, the proposed technique is equipped for large scale deployment.
(iii) The obtained results from trained data looks promising with tolerable acceptance rate.

The rest of the paper is organized as follows. The materials and methods section elaborately explains the procedure used to identify the gestures. Sub-section B and C highlights methods of generating required figure partitions while D and E generates the raw data for filling the feature vector. Sub-section F delineates the production of the unique features for different gestures and successive training with obtained data. The fourth section presents the conclusion and a brief insight into future works that encompasses improvements and enhancement possibilities.

II. RELATED WORKS

The American Sign language (ASL) [1] provides a comprehensive list of hand gestures, that can be used for non-verbal communication. Based on these specified gestures, there have been extensive research on trying to recognize and comprehend them. [2] presents a real time Markov model-based systems which uses a single camera for recognizing sentence-level continuous American Sign Language (ASL). This is a majorly used technique to recognize gestures. However this implementation requires polygon approximation followed by convex decomposition.

A convex decomposition method is the alternating sum of volumes (ASV) method. [3] uses convex hulls and set-difference operations. A convergent convex decomposition is proposed which uses a combination of ASV decomposition and remedial partitioning for the non-convergence. Their paper uses ASVP decomposition for feature extraction. Hierarchical relationships between the recognized form features are also obtained. [4] proposes a strategy to decompose a polygon, containing zero or more holes, into approximately convex pieces. They use a mechanism of approximate convex decomposition (ACD),that focuses on key structural features and ignores less significant details. They proposed a simple algorithm that computes an ACD of a polygon. This is done by iteratively resolving the most significant non-convex feature. As a by-product, it also produces an elegant hierarchical representation that provides a series of decompositions which are increasingly convex.

[5] presents an algorithm to compute a convex decomposition of a non-convex polyhedron of arbitrary genus (handles) and shells (internal voids). Their algorithm is based on the simple cut and split paradigm of Chazelle. With the help of zone theorems on arrangements they show that this cut and split method is quite efficient. The algorithm is extended to work for a certain class of nonmanifoldpolyhedra. Also presented is an algorithm for the same problem that uses clever heuristics to overcome the numerical inaccuracies under finite precision arithmetic.

[6] presents an original approach for 3D mesh approximate convex decomposition. Their proposed algorithm computes a hierarchical segmentation of the mesh triangles. This is done by applying a set of topological decimation operations to its dual graph. The decimation strategy is guided by a cost function describing the concavity and the shape of the detected clusters. The derived segmentation is finally exploited, to form an approximation of the original mesh, using a set of convex surfaces. This new representation is particularly adapted for collision detection. Their experimental evaluation conducted shows that the proposed technique efficiently decomposes a concave 3D mesh into a small set (with respect to the number of its facets) of nearly convex surfaces. Furthermore, it automatically detects the anatomical structure of the analyzed 3D models, which makes it an ideal candidate for skeleton extraction and patterns recognition applications.

[7] proposes an automatic gesture recognition on Indian sign language, that uses both the hands to represent an alphabet. They address inter-class variability enhancement and local-global ambiguity identification for each hand gesture. Hand region is segmented and detected by YCbCr skin color model reference. Texture, shape and finger features of the hands are extracted using Wavelet Packet Decomposition (WPD-2), Principle Curvature Based Region (PCBR) detector, and complexity defects algorithms respectively for hand posture recognition process. Furthermore, to classify each hand posture, Support Vector Machines (SVM) is used, that achieves a recognition rate of 91.3. To classify dynamic gestures, Dynamic Time Warping (DTW) is used with the trajectory feature vector, and it achieves 86.3% recognition rate.

The Kinect sensor, have provided new opportunities for human-computer interaction. But when compared to the entire human body, the hand is a smaller object with more complex articulations and is more easily affected by segmentation errors. It this faces challenges in recognizing hand gestures. [8] focuses on building a part-based, robust hand gesture recognition system using a Kinect sensor. To handle the noise from the Kinect sensor, they propose a novel distance metric, which is the Finger-Earth Mover's Distance (FEMD). It measures the dissimilarity between hand shapes. They achieve 93.2% mean accuracy on a 10-gesture dataset and are robust to hand articulations, distortions and orientation or scale changes, and can work in uncontrolled environments (cluttered backgrounds and lighting conditions). They have further demonstrated their superiority in two real-life HCI applications.

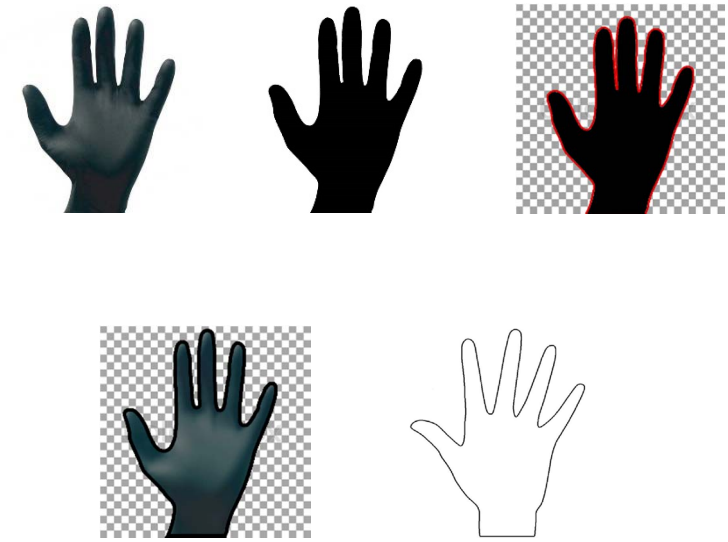
[9] mentions the key aspects of sign-language recognition (SLR). Data available and their relative merits are analyzed to find features that can be extracted. Methods for combining the sign classification results into full SLR are given showing the progression towards speech recognition techniques and the further adaptations required for the sign specific case. They show the task of continuous sign recognition, and the work towards true signer independence, and the effective combination of the different modalities of sign, making use of the current linguistic research and adapting to larger more noisy data sets.

[10] presents a large vocabulary sign language interpreter with real-time continuous gesture recognition of sign language using a data glove. They perform a statistical analysis according to four parameters in a gesture: position, posture, orientation, and motion. They implement a prototype system with a lexicon of 250 vocabularies and collected 196 training sentences in Taiwanese Sign Language (TWL), using hidden Markov models (HMMs) for 51 fundamental postures, 6 orientations, and 8 motion primitives. Their average recognition rate is 80.4%.

With oral languages, lexical borrowing from one manual language into another is accompanied by lexical restructuring in accordance with the formational and morphological principles of the borrowing language. [11] presents a study that examines some English words that are fingerspelled by signers in a systematic and predictable manner. The focus of the study is on the formational aspects of signing. An analysis of loan signs and the English influence that prompts their borrowing also depends on the social world of signers, which is discussed in terms of those aspects of social interaction that create ASL-English bilinguals.

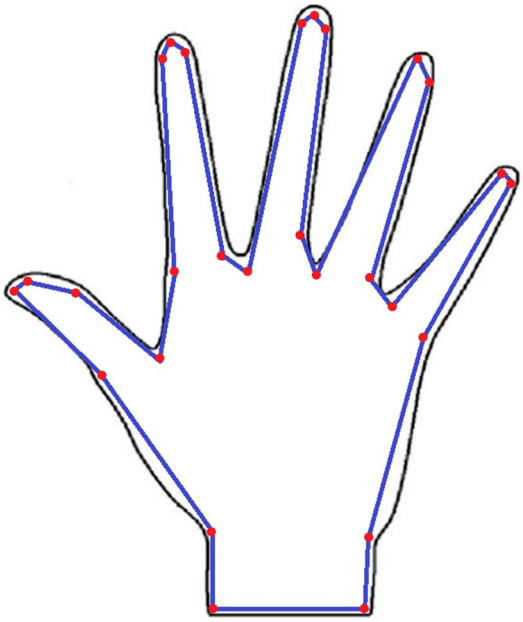
III. MATERIALS AND METHODS

A. Preprocessing



In the above figure, the vertices colored red are of the polygon which is sketched out from the curved outline of the image. These vertices when joined as shown by the blue lines, constitute a simple polygon P.

B. Polygon Approximation



The approximation of arbitrary two-dimensional curves by polygonal figures is an imperative technique in digital image processing. For numerous applications, the widely used procedure for this purpose is to represent lines and boundaries by means of polygons that have a minimum or a small number of vertices satisfying a given fit condition.

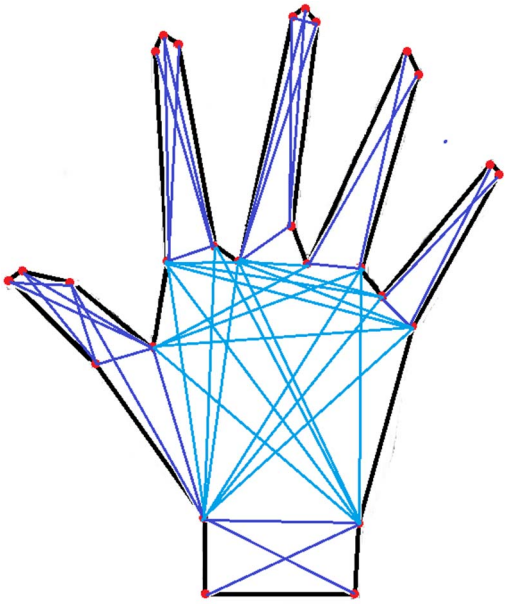
The number of line segments used in the process of creation of a polygon determines the accuracy of the approximation algorithm. For an algorithm to be effective and accurate, it must not exceed the minimum number of required sides necessary to preserve the actual shape of the curve. A polygon thus created with only the minimum requisite number of line segments is often named as a Minimum Perimeter Polygon. A higher number of edges in an approximated polygonal figure adds to the source of noise to the model.

Polygon approximation removes the non-essential curves and bends and provides us with a discrete figure that is bounded by line segments. The resultant image consists of a simple polygonal region in the plane bounded by a non-self-intersecting, closed, polygonal path. The polygonal path itself is part of the polygon; it is usually called the boundary.

C. Convex Decomposition

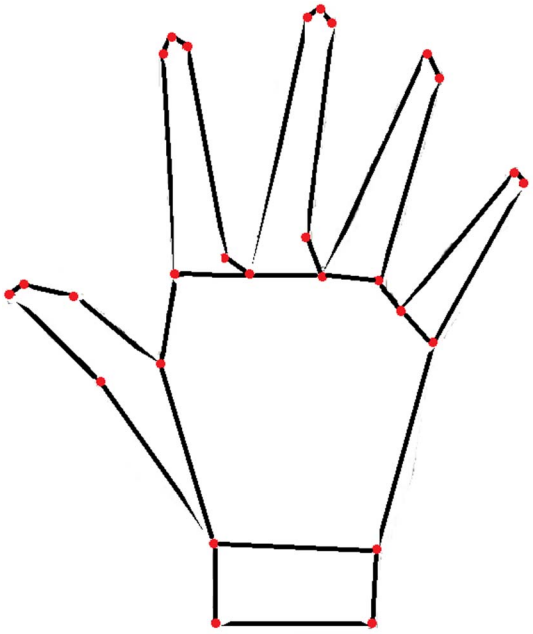
One common approach for dealing with complex models or many-sided figure representation is to partition them into pieces that are easier to handle. Numerous difficulties can be solved more proficiently when the input is convex. Decomposition into convex components results in pieces that are easy to process and additional approximation simples and eases computation, ensuring quicker and efficient output with manageable margins of inaccuracy.

Convex decomposition is of two types: exact and approximate. We do not use exact decomposition because we want to minimalize the number of clusters in the model. Approximate convex hull decomposition gives us the minimum number of clusters ensuring that each cluster has a concavity lower that a predefined threshold.

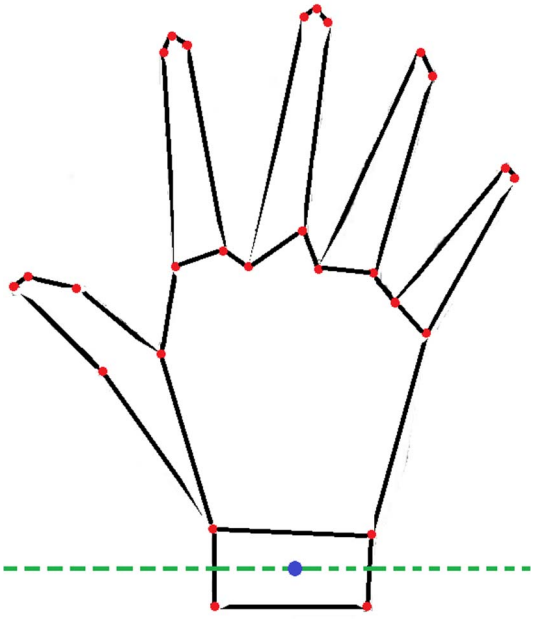


The figure given above shows the result of convex decomposition on the approximated polygonal segments obtained in the previous section.

In graph theory, a clique is defined as a complete subset of vertices of a graph such that every two distinct vertices in the clique are adjacent. Subsequently, a maximum clique is one that has the largest possible number of vertices among other cliques. The maximum clique in the aforementioned visibility graph G is a complete subgraph of G having the maximum number of vertices. We compute the maximum clique of the simple polygon P of G in time O(n2e), where n and e are number of vertices and edges of G respectively.

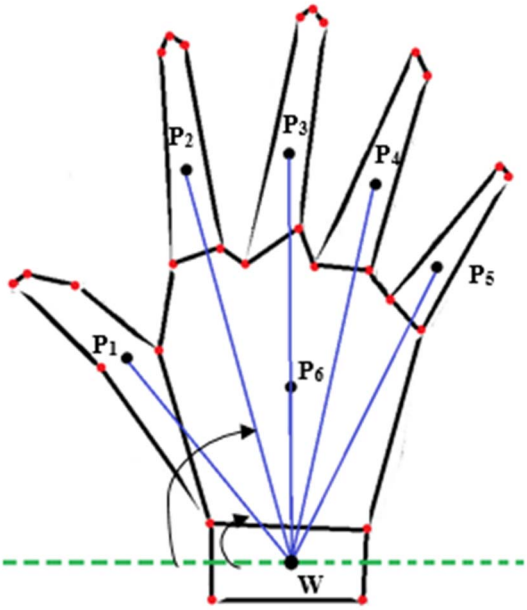


D. Finding Wrist Point and Reference Line



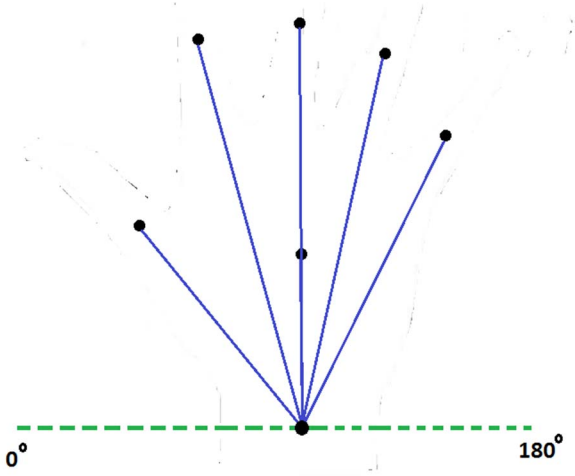
E. Calculating Angle and Distance with respect to Wrist Point.

We now have the line of reference as well as the wrist point which are used to realize the different polygon segments in a given figure. We achieve this by observing the incrementing angles between the line of reference and each line joining the mid points of the polygonal segments. Our main objective here is to find unique vector to represent different hand gestures.



With reference to the (), we begin by finding the mid-points of the segments obtained by the convex decomposition process. These points, marked as P_1 through to P_6, are then extended to join the wrist point marked as W. What is important to note here is that with each segment mid-point, the angle increases with the minimum being a possible 0 degree and the maximum being a possible 180 degrees.This gradual expanding gap between the wrist point and each subsequent line signifies the change in finger positions and helps us to differentiate between them. Another aspect of the line joining the wrist point and the mid points of the segments is its length.

The property which helps us to decide whether a finger is in a folded or unfolded position is the distance between the wrist point and the mid-point of the corresponding polygonal segment of that finger. While the finger is in a folded position, the centroid of the segment gets located in a vertically lower level inside the polygon than it would if the finger was in an unfolded position. As a result, it becomes fairly easily distinguishable to recognize different signs, especially the ones that only differ by the positioning of a single finger.



Outcome of the above stated procedure produces the angle and distance of each mid-point of the different generated convex segments with respect to the reference line and wrist point. It shall be noted that the duality (angle and distance) needs to be converted to a singularity for feature extraction and forming a unique one-dimensional array representation of each distinguishable hand gesture.

F. Generating and Training Feature Vector.

We create a feature vector of dimension 36 for each gesture. Diving the scale from 0 degree to 180 degree in equivalent divisions of 5 degrees we are able to make 36 partitions. In a given interval we find the lines joining the mid-point of a segment and wrist point. Then we store the length of that line for the particular or n^{th} interval. In situations where we have more than one line in a given partition we add the lengths of the generated line and save them to the feature vector. Here length is calculated using Euclidean distance between the midpoint and wrist point.

We now create a matrix of features by combining the feature vector of all the gestures present in our training set. We then train the machine with our matrix of features using SVM, Artificial Neural Network, Naive Bayes and K-NN classifier and compare the accuracy of each on a common test set.

IV. Conclusion and future works

V. References

[1] Stokoe, William C. "Sign language structure." (1978).

[2] Starner, Thad, Joshua Weaver, and Alex Pentland. "Real-time american sign language recognition using desk and wearable computer based video." IEEE Transactions on Pattern Analysis and Machine Intelligence 20.12 (1998): 1371-1375.

[3] Kim, Yong Se. "Recognition of form features using convex decomposition." Computer-Aided Design 24.9 (1992): 461-476.

[4] Lien, Jyh-Ming, and Nancy M. Amato. "Approximate convex decomposition of polygons." Proceedings of the twentieth annual symposium on Computational geometry. ACM, 2004.

[5] Bajaj, Chanderjit L., and Tamal K. Dey. "Convex decomposition of polyhedra and robustness." SIAM Journal on Computing 21.2 (1992): 339-364.

[6] Mamou, Khaled, and Faouzi Ghorbel. "A simple and efficient approach for 3D mesh approximate convex decomposition." Image Processing (ICIP), 2009 16th IEEE International Conference on. IEEE, 2009.

[7] Rekha, J., J. Bhattacharya, and S. Majumder. "Shape, texture and local movement hand gesture features for indian sign language recognition." Trendz in Information Sciences and Computing (TISC), 2011 3rd International Conference on. IEEE, 2011.

[8] Ren, Zhou, et al. "Robust part-based hand gesture recognition using kinect sensor." IEEE transactions on multimedia 15.5 (2013): 1110-1120.

[9] Cooper, Helen, Brian Holt, and Richard Bowden. "Sign language recognition." Visual Analysis of Humans. Springer London, 2011. 539-562.

[10] Liang, Rung-Huei, and Ming Ouhyoung. "A real-time continuous gesture recognition system for sign language." Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on. IEEE, 1998.

[11] Battison, Robbin. "Lexical borrowing in American sign language." (1978).