HAND GESTURE RECOGNITION USING K-MEANS CLUSTERING AND CONVOLUTION NEURAL NETWORK

Gowtham A, Ahamed Afridi F, Nelson Louis, Priyanka M, Sam Anton

Gowtham A UG student, Electronics and Communication Engineering, SRMIST, Chennai.

Ahamed Afridi F UG student, Electronics and Communication Engineering, SRMIST, Chennai.

Nelson Louis UG student, Electronics and Communication Engineering, SRMIST, Chennai.

Priyanka M UG student, Electronics and Communication Engineering, SRMIST, Chennai.

Sam Anton UG student, Electronics and Communication Engineering, SRMIST, Chennai.

ABSTRACT

Hand gesture recognition can be defined as an ability of a computer to recognize meaningful expressions of form and motion by a human involving only the hands. These hand gesture recognitions have plenty of replications which can be applied for improving control, accessibility, communication and learning. There is a complexity issue with this hand gesture recognition feature extraction, for example the variation of the light and background. In this paper, the convolution neural network is applied for the recognition of gestures, and the characteristics of convolution neural network are used to overcome the issue of feature extraction process and reduce the number of parameters needs to be trained. The purpose of unsupervised learning, error back propagation algorithm can be finally achieved which is loaded into the convolution neural network algorithm, modify the threshold and weights of neural network to reduce the error of the model. The advantage of our method is that there is no need for feature extraction. Without explicitly segmenting foreground the proposed CNN learns to recognize the hand pose even in presence of complex, varying background or illumination.

Keyword:

1. *Digital Image processing.*
2. *K-means clustering.*
3. *Convolutional neural network.*

---

1. DIGITAL IMAGE PROCESSING

1.1CLASSIFICATION OF IMAGES:

The three types of images used in Digital Image Processing are:

- Binary Image
- Gray Scale Image
- Colour Image

1.1.1 BINARY IMAGE

It is a digital image which has only two possible values for each pixel. Black and White are the two constant colours that are supposed to be used in a binary image. Whereas, foreground colour is denoted using white and the background is denoted using black.

1.1.2 GRAY IMAGE

It is an image that denotes the absence of any chromatic variation which is why the gray scale images can also be called as Monochromatic. These images are different from one bit black and white images i.e., the computer imaging which are images with the true prominent colours, black and white (Bi-level or binary images).

1.1.3 COLOUR IMAGE

This colour image is a digital image that contains the colour information for each pixel. The specific value of each pixel helps in determining the pixel's appearing colour. The value can be determined by three numbers of three primary colours red, green and blue given the decomposition of the pixel colour. This way can be used to represent any colour that is visible to human eye. The decomposition of a colour in the three primary colours is quantified by a number between 0 and 255. For example, white will be coded as R = 255, G = 255, B = 255; black will be known as (R, G, B) = (0,0,0); and say, bright pink will be (255,0,255).

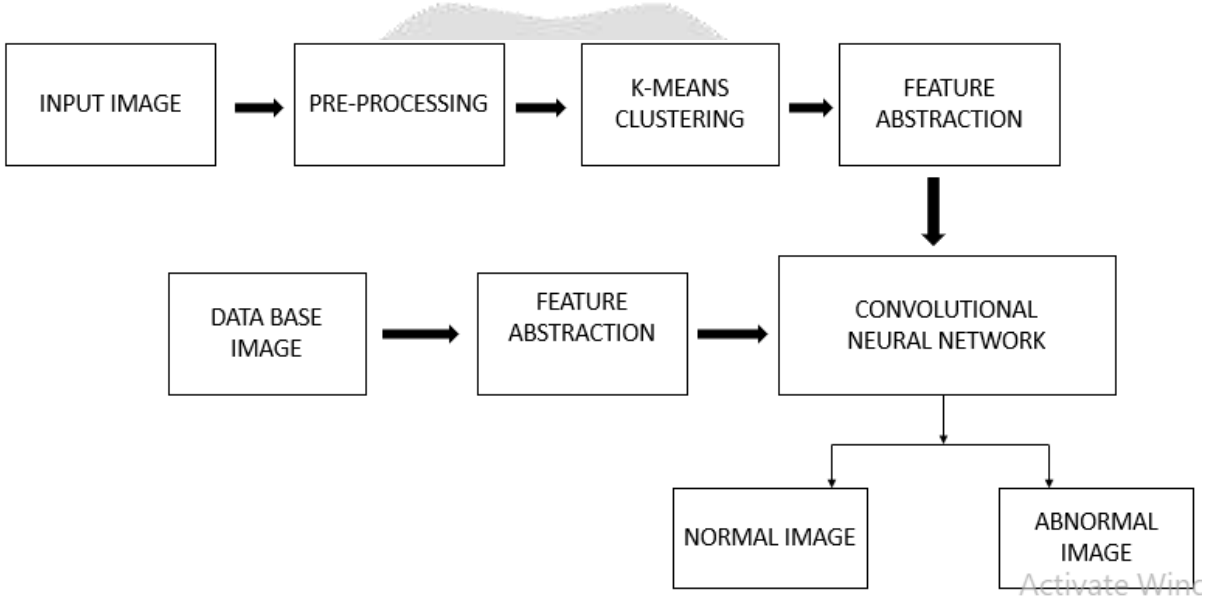

Fig-1 BLOCK DIAGRAM

2. K-MEANS CLUSTERING

This is a sort of unsupervised discovering that is utilized when you have unlabelled information to mark it i.e., when the information is without characterized classifications or gatherings where the information is characterized utilizing K-implies grouping. The essential objective of the K-implies bunching calculation is to discover bunches in the information, with the quantity of gatherings spoken to by the variable K. The calculation works iteratively to allocate every datum point to one of K bunches dependent on the highlights that are given. Information focuses are grouped dependent on highlight closeness. The results of the K-means clustering algorithm are:

- The centroids of the K clusters, which can be used to label new data
- Labels for the training data (each data point is assigned to a single cluster)

2.1 ALGORITHM

This algorithm uses iterative process to produce an improvised result. The number of K clusters and the data set are the inputs for the algorithm where data set is a collection of features for each data point. These algorithms start with initial estimates for the K centroids, which can either be randomly generated or randomly selected from the data set. The algorithm generally iterates between two steps:

- Data assignment step:

Each centroid denotes one of the clusters. In this first step, each data point is assigned to its nearest centroid, based on the squared Euclidean distance. More formally, if ci is the collection of centroids in set C, then each data point x is assigned to a cluster based on where distance (·) is the

standard (L2) Euclidean distance. Let the set of data point assignments for each I th cluster centroid be Si.

- Centroid update step:

In the second step of algorithm, the re-computation of centroids is done by taking the mean of all data points assigned to that centroid's cluster.

This iterates continues to and fro a stopping, criteria been met (i.e., no data points change clusters, the sum of the distances is minimized, or some maximum number of iterations is reached).
This algorithm is guaranteed to converge to a result. The result may be a local optimum (i.e. not necessarily the best possible outcome), meaning that assessing more than one run of the algorithm with randomized starting centroids may give a better outcome.

2.2 ADAPTIVE K-MEANS CLUSTERING

The strategy which is utilized in parcel of components in an informational index so that indistinguishable components fall under a similar bunch while components with various properties fall under various groups. The point of bunching is to perform huge pursuit of components in an informational collection. The grouping is generally utilized in multidimensional information that might be hard to bunch it in a successful way. K-implies grouping is one of the soonest bunching methods in the writing. In this grouping it relies upon distinguishing K components in the informational index that must be utilized in making an underlying portrayal of bunches. The development of the bunch seeds is finished by these K components. The components that are abandoned in the informational index is then doled out to one of these bunches. Evidently, this bunching technique might be exact, however the truth of the matter is it may not be most likely simple to distinguish the underlying K components or the seeds for the groups with lucidity.

The versatile K implies grouping calculation starts by choosing the K components from the info informational collection. The determination procedure of these K components is done arbitrarily. The properties of every component acquire their properties from the groups that are established by the component. This calculation relies on the capacity in registering the separation between a given component and a bunch. This capacity is additionally talented to process the separation between the two components. This capacity ought to have the capacity to consider the properties of the group that have been standardized so the separation isn't designated by one property or some property that isn't disregarded amid the separation calculation. In majority of cases the Euclidean separation that might be considered as effective. For example, considering the instance of the ghostly information given by 'n' measurements, the separation between the two information components E1 and E2 is processed.

3. CONVOLUTIONAL NEURAL NETWORK

Convolutional neural networks (CNN) is a recently emerging artificial neural network structure. Because of the usage of the CNN and the speech recognition for better result this algorithm is widely spread and applied. CNN is the most commonly used computer image recognition, but because of its constant innovation, this method has been applied to video analysis, drug analysis, natural language processing and other fields. The recent alpha-go on the network has been using CNN Algorithm which allows the computer to read based on the deep learning of the convolution neural net-work method has a local perception area, hierarchical structure, feature extraction and classification process and so on, can automatically learn the appropriate characteristics and classification, in the field of image recognition has been very Good practical effect. The characteristics of gesture images are embodied in the changes of different forms of gestures, but also in the local information of gestures.

3.1 STRUCTURE OF CNN

It is a multi-layer neural network that has two dimensional planes, each plane consisting of multiple independent neurons. The network consists of convolution layer, sampling layer and the whole connection layer, which assures that feature extraction and pattern classification are performed at the same time. The three ways to achieve the image information in CNN are displacement, scaling and distortion of the invariant, i.e., local experience, weight sharing and down sampling. Local sensation implies to the fact that each layer of neurons only to the local area of the upper layer. Through the local experience of the field, you can make each neuron to extract the initial structure of the image features such as corner, endpoint and direction line. Weight

sharing can reduce the need for training parameters. There is usually a descending profile behind the convolutions, which reduces the resolution of the image features. While the network has a certain degree of displacement, scaling and distortion of the invariance, can make the training out of the weight is more conducive to classification

- **Convolution layer:** Each image of a picture is processed by convolution, and the width and length of the image are com-pressed to obtain deeper image information.
- **Pooling layer:** Study found that every time when the convolution, convolution may be missing information, pool layer can solve this problem. No compression when the convolution aspect at the same time, to retain more information compressed work on completion of the pool floor.
- **Link layer:** The link layer is a layer the output neuron, in which each neuron connects with the input layer and will pass out information at the top of the consolidation, so that information can be delivered throughout the neurons. Normally two layers do not match the number of neurons, processed based on convolution kernel. And it helps to join the activation function to convert into nonlinear equations.
- **Classifier:** It is functioned to classify the clusters.

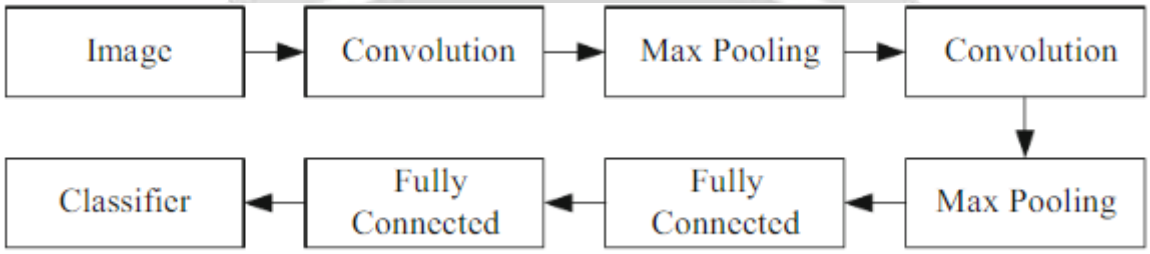

Fig-2 SIMPLE CNN STRUCTURE DIAGRAM

Fig-3 EIGHT HAND GESTURES (G1, G2, G3, G4, G5, G6, G7, G8)

3.2 ERROR BACK PROPAGATION

This error back propagation algorithm comes into action when there is a requirement to meet different application, with ability of automatic information process, fabulous capability of anti-jamming, robustness. The BP contains two signal propagation direction one being positive communication, signal is input to a neural network, and then through the handle to the output and the other one is back-propagation, according to the output values and com-paring the expected value on business value, then this error back-propagation to the input layer. The parameter of each layer is modified so that the output value error comes out with expected value. BP with widespread, mainstream is now deep learning algorithm using BP to modify the parameters and reduce the errors.

4. CONCLUSION

In this paper we proposed a gesture recognition method based on convolution neural networks to overcome the problem in gesture recognition. Convolution neural network in the processing of two-dimensional images, through the layer of convolution and down sampling operation, you can extract the image of the structural features. The artificial selection of features, at the same time with a certain degree of translation, scaling and rotation invariance to avoid the subjectivity. To reduce the difficulty, to improve the efficiency of the operation the local experience and weight

sharing technology is preferred compared to traditional methods, with a higher recognition rate and faster recognition speed. The robustness of the system is enhanced using depth image for segmentation of colour image, removing background noise of the outside world. Results showed that in semi-supervised case, the eight types of gesture recognition the recognition rate is up to 98.52%. Experiments proved that the method of using CNN is effective. But CNN may use different featured maps to capture the image multiple times. The future researches must focus on calibration of hand positions, as well as in blocks or background moving cases to track information and accurate collection of hand gestures.

5. REFERENCESC

[1] Y. Fang, K. Wang, J. Cheng and H. Lu. "A real-time hand gesture recognition method," in Proc. of IEEE Int. Conf. Multimedia and Expo, Beijing, China, July 2007, pp. 995-998.

[2] N. H. Dardas and D. G. Nicolas, "Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques," IEEE Trans. Instrumentation and Measurement, vol. 60, no. 11, pp. 3592-3607, Nov. 2011.

[3] S. Belongie, J. Malik and J. Puzicha. "Shape matching and object recognition using shape contexts," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 4, pp. 509-522, Apr. 2002.

[4]. Ijjina, E.P., Chalavadi, K.M.: Human action recognition usinggenetic algorithms and convolutional neural networks. PatternRecognit. 59(11), 199–212 (2016)

[5]. Chen, D., Li, G., Jiang, G., Fang, Y., Ju, Z., Liu, H.: Intelligentcomputational control of multi-fingered dexterous robotic hand. J.Comput. Theor. Nanosci. 12(12), 6126–6132 (2015)

[6]. Mahmoudi, M., Sapiro, G.: Sparse representations for range datarestoration. IEEE Trans. Image Process. 21(5), 2909–2915 (2012)

[7]. Nebti, S., Boukerram, A.: Handwritten characters recognitionbased on nature-inspired computing and neuro-evolution. Appl.Intell. 38(2), 146–159 (2013).

[8]. Li, G., Miao, W., Jiang, G., Fang, Y., Ju, Z., Liu, H.: Intelligentcontrol model and its simulation of flue temperature in coke ovenDiscrete. Contin. Dyn. Syst. Ser. S 8(6), 1223–1237 (2017)