IV. Review based on experimental results

1. Dataset of chosen method
2. Accuracy and Parameter
3. Result with numerical estimation
4. Dataset with accuracy

This section provides a comparative description of chosen research work respected to dataset and experimental results of their dataset. No research can be verified without experimental results or real-life implementation. Different researcher used different procedure to recognize hand gesture at their most efficient level. As a consequence, they used different dataset to experiment their method and they got different accuracy. The dataset was either previously used authenticated dataset or self-made dataset. self-made dataset was made by the help of volunteer of different age, gender and skin tone. in respect to accuracy, some accuracy is high and some accuracy was low. the factor that improvise the accuracy was form of hand, gesture variance and lighting condition. Some experiment was done by simple tools and some needed a high-tech system to recognize the hand gesture. So, the experimental parameter was also different from each other. So, considering all these aspects like experimental parameter, dataset preferences and acquired accuracy, the whole section is organized and they are described in different subsection under the experimental results.

i. Dataset of chosen method

Table 4. Dataset and parameter with reference

|  |  |
| --- | --- |
| Unnamed Dataset [16] | 11.Relationship between geometric forms  2.Construction of geometric forms using CSG operators  3.Transformation of geometric forms using extrusion technology |
| Self-made dataset [15] | Linux OS, Pentium iii 500 MHz, Hitachi iP5000 image processing board, Nikon laired-s270 |
| Self-made dataset [7] | Acer 5750G notebook, Intel corei5-2450M, 2.5 GHZ, 4GB ram, Win & DEV C++ and C, OpenCV 2.1 |
| Self-made dataset [5] | Corei7(4720),2.6 GHz, 8GB ram |
| Hand Gesture Dataset LSP [22] | Taken under various characteristics such as scale, rotation, translation, lighting and noise. |
| Confusion matrix of the classification [17] | A total of 10 gestures that match the pinch, scale, buckle, hold, grab, rotate, crawl, OK, cut, shoot. |
| NUS Hand Posture Dataset-Ii [21] | The accuracy of recognition of manual gestures is tested on subsets |

ii. Accuracy and Parameter

[Table no. 5 and 6]

Unnamed dataset was used in extrusion technique and the parameter used is the relationship between the geometric forms. Fingertips method used their own made dataset and it used Linux OS, Pentium iii 500 MHz, Hitachi iP5000 image processing board, Nikon laired-s270 [16]. becoming a random dataset, it prevails accuracy 80% and the main fact is that it deals with real time video to recognize hand. In trajectory base method self-made dataset were used and the parameter used to perform the experiment were Acer 5750G notebook, Intel corei5-2450M, 2.5 GHZ, 4GB ram, Win & DEV C++ and C, OpenCV 2.1 [15]. The team used a self-made dataset but the noticeable thing about the dataset is that it works ion infrared. Neural Network based method also used their own dataset and their parameter were Corei7(4720),2.6 GHz, 8GB ram [7].It takes a lot of system to be assembled to test the project. it can recognize within 55ms.Baseline Convolutional Neural Network Architecture select Hand gesture dataset of LSP as their dataset and the parameter was total of 10 gestures that match the pinch, scale, buckle, hold, grab, rotate, crawl, OK, cut, shoot [22]. Auto encoded model choose confusion matrix of the classification as their dataset and various scale, pinch, rotation was taken as parameter [17].The error rate was mainly due to 15 percent of the buckle gesture error evaluated into the wrong gesture, whereas 24% of the shooting error was recognized as the wrong gesture. NSU hand posture dataset was taken by combination of Histogram of Oriented Gradients (image processing) and Local Binary Pattern method and the parameter was the accuracy of recognition of manual gestures is tested on subsets [21]. The accuracy was tested on manual gesture and the dataset got the highest accuracy point over all the dataset.

The AR application took 53.8 seconds to finish the assignment and GeoGebra took 71.95 seconds. Conversely, Cabri3D took 47.67 seconds since this is an extremely straightforward application [16]. The disarray network of arrangement demonstrates that the blunder rate was fundamentally due to a 15% of the signal clasp mistake partitioned into the wrong motion cut, while 24% of motion shoot mistake isolated into the wrong motion squeeze, from the first database pictures can be seen, the principle reason is that the high closeness edge highlight of these pictures. It drove along these lines to the ultimate result of disarray [17]. Execution process utilizing K-Means takes longer than utilizing the SVM strategy, however the K-Means precision rate is lower than the SVM technique. [18]. 50 video samples for training and testing by 4 people each test 8 moves and 1 moves 50 times. The average rate of recognition is 96.25%. [20]. Experiments with a set of five people with different skin tones led to the following results. [19]. Understudies needed to work in one of the instructional meetings: Students need to construct a progressive surface by turning a B-Spline bend (cubic, 5 - 6 control focuses) around the pivot on the hub. [18]. There have been significant positive correlations between DAT: SR and PSVT: R on the one hand and non - verbal reasoning on the other[19].

Azad el at. [2] Make three types of their input video in long, medium and short. Feature extracted using VLAD. Classification were done through combinedly SLNF and ELM. The experiment was applied to four data set and they showed 98.05, 97.31, 95.24, and 68.66% respectively. Taking EMG signals as input CHEN et al. [3] extracts feature using convolutional neural network and then classify them using SVM classifier. There are totaling 11 layer and the experiment was done on Nina pro database and it results in 65.4 and 68.2 % on CNN and SVM respectively. Leap motion is used as tracking technology by Jung et at. [4] And virtually hand is managed in two mode. It also provides a system to reset the.Self-made dataset left GAS 4.7 by accuracy. Munasinghe [5] applied his experiment in self-made database with Gaussian mixture-based subtraction, median blur and thresholding. He tested in low light and good light results in 71.3 and 85% respectively. With a convolution layer size containing 300 layer and dropout 0.4% Alam et al. [6] applied their procedure Khushboo el data set and the accuracy rate are at highest peak 98.88%. Liu el at. [7] used 2 cross and 10 cross validation to train and applied it on self-made database and he got 92.3 and 94.7% accuracy. Using IP camera hand contour was extracted using convex hull detection and applied the experiment on a self-made database and the outcome in 10 different gesture. Average accuracy was 90% in Rani et al. [8] self-made database. Hand posture pipeline and pose estimation was applied with CTE value 290. Mentionable steps in Chiang et al. [9] framework Connected component learning, contour smoothing, convex hull and blending hand posture. The experiment was performed on a self-made database. For index finger the result 92.12% and for two fingers 98.87%. Kitimat et al. [10] used two types of method for his classification. First one is multilayer perceptron and second one is k-means classifier. Average accuracy found 93.Two type of wrapping DWT and elastic matching was used as matching technique in Liu [11] experiment. After particle filtering hidden markov model is used for classification on ten gesture of Microsoft MSR dataset and the acquired accuracy is 93%.

Liu et al. [12] was aimed to find out the best classifier. So, he fed into the input in HMM and then classify the input using gyroscope, accelerometer and Kinect. The experiment is tested on $1Unistroke recognition application and the accuracy rate is 91%. Extracting gesture from multiple sequence of gesture in Gudavalli et al. [13] experiment. Than extraction of motion component and location component is performed using PCA and PWDTW. Additionally, shape component is extracted using STIP. The experiment is tested on CHEALEARN gesture dataset and from the dataset, every time the cascading method take least time than other traditional method. Detecting multiple figure is an important feature of Oka et al. [14] experiment. After measuring fingertips trajectories gesture is recognized with thumb detection and symbolic gesture recognition. The experiment was performed on a self-made dataset and there are two types of accuracy. For sing le finger accuracy rate is higher counting 99.2% than double finger accuracy rate 97.5% though their difference is not that much noticeable.

Optical flow is extracted for each two consecutive frames and a 8-bin histogram is created based on the flow direction. [38] [19]. Ten pictures of ten people were taken. The segments of the skin used to calibrate our Gaussian skin model were taken randomly. Accuracy by masking 96 percent and 98 percent by gradient approach[39].The experiments are carried out on the NUS data set-II hand posture. Chaikhumpha Tanatcha [20] proposed system offers good results in recognizing gestures in real time from color image sequences through the single-hand movement trajectory and multiple HMM usage. The HOG highlight has turned out to be outstanding amongst other highlights for securing data on the edge and neighborhood shape and has been effectively connected in fields, for example, target discovery and recognizable proof. The LBP include is an extremely successful surface component descriptor that is strong for changing the dim scale and changing the picture pivot [39].

iii. Result with numerical estimation

Table 5. Previous Research based on Experiments

|  |  |  |
| --- | --- | --- |
| Experimental Steps | Parameter Used | Numerical Estimation |
| Subjects perform tasks based on the scenarios for Cabri3D, GeoGebra, AR application. [16] | 11.Relationship between geometric forms  2.Construction of geometric forms using CSG operators  3.Transformation of geometric forms using extrusion technology | 53.8 seconds to complete the task and 71.95 seconds to complete GeoGebra. In contrast, Cabri3D took 47.67 seconds because this is a very simple application. |
| Choosingimages. [17] | A total of 10 gestures that match the pinch, scale, buckle, hold, grab, rotate, crawl, OK, cut, shoot. | The error rate was mainly due to 15 percent of the buckle gesture error evaluated into the wrong gesture, whereas 24% of the shooting error was recognizedas the wrong gesture, |
| Four - movement K-Means SVM error rates (forward, right, left and stop) [18] | K-Means and SVM testing era distance from 2 m to 4 m, and body slope position is 450,00, -450.or four - movement rates (forward, right, left and stop) | Implementation process using K-Means takes longer than using the SVM method, but the K-Means accuracy rate is lower than the SVM method. |
| Hand gestures from the default database [19] | Classified by set of if-then rules. | Accuracy of (c / t) \* 100. |
| Confusion matrix representing the accuracy of recognition for every class of hand gesture. [21] | The accuracy of recognition of manual gestures is tested on subsets | For subset A, 97.8 percent and for subset B, 95.07 percent. |
| The dataset contains 3750 images of 25 persons. The dataset consists of six unique images. [22] | Taken under various characteristics such as scale, rotation, translation, lighting and noise. | |
| 275 pictures, 25 samples per 11 movement. For 11 classes or gestures with 25 samples in each class, a multi - class SVM is available to solve the objective. [23] | 25 samples in each of the class | 100% accuracy  Accuracy reduced in complex background |
| 249 gestures of 21 different persons. To assess the performance of the proposed method, only depth maps have been used. [25] | RGB-D depth sequences | |
| EMMIE 's a meeting scenario. Two users tracked can see headword displays. [31] | All clients can see a projection divider - measured showcase. The triangular hotspot for one tracker is mounted on the left half of the table; further roof mounted trackers can't be seen here. | |
| A total of 5 training sessions lasting 6 hours. [35] | Can work to structure their structure in 3D layers. Transparencies, as used in technical drawings, permit to see objects inside | |
| Some student participated in an experiment to follow some steps using a prototype device [32] | Four Colored Markers | 92% of students find that the prototype system is faster than using conventional methods. |
| Normalization of hand gesture  MTS-Three sequence of video, Intra class +, Inter class –  Projected on three Cartesian plans, WDMM-Compute 1 image from a depth sequence,  HOG+LBP-Feature extraction using VLAD  Classification-SLNF+ELM [2] | Video sequence as an input,  Key frame extraction,  Extracted feature to classify | Long =Original video  Middle=50%  Short=30%  Size-ofWDMM=8X8. HOG size=9  No of HOG vector=36 |
| Feature extraction  CNN (based on deep learning method)  Classification-SVM [3] | EMG signals | 11layers(4(convolutional)+6 connected layer+1 output layer) |
| Leap motion  -Track hand, Trigger  Virtual hand management  -Normal mode, Boost mode  Reset hand [4] | Any kind of general gesture | Classroom range=6m  Alpha=3  Beta=5  Hand movement range (0.5-0.8) m |
| Background subtraction  -Gaussian mixture-based segmentation  Apply blur to image  -Median blur  Apply thresholding  Outs binarization [5] | Dynamic gesture, image with categorization | 29 frames per second(1280X720) video quality  Motion history image size (100X56) |
| Background removal  Connected component labeling  Hand contour  Contour smoothing  Contour feature and convex hull  Bending posture  Hand posture recognition [5] | Corei7(4720),2.6 GHz, 8GB ram | Processing speed 15frame/s  Range 16-50 cm |
| Training Phase  -Training hand gesture, Fingertips 3D positions, Gesture feature derivations, Modeling and identification  Recognition phase  -Gesture recognition  Error estimation [6] | Kinect sensor V2 | 2 cross fold and 10 cross fold validation |
| Web camera  -Capture RGB image  Hand detection  -RGB translation  -Skin detection and contouring (Morphological processing)  -Delete the face and wrest  Gesture recognition  -Angel measure  -Convex hull  Fingertips [7] | Acer 5750G notebook  Intel corei5-2450M  2.5 GHZ  4GB ram  Win &  DEV C++ and C  OpenCV 2.1 | Image size=640X480  Cr=131-183  Cb=102-154  Tc=80 |
| Input  Batch normalization  Convolution  -ReLU  Max pooling  Dense activation  Normalization  Global average pooling  Dropout  Dense activation  -SoftMax  Output [8] | Intel corei7 3.5 GHz processor  16GB RAM  NVIDIA GeForce GTX 1080 GPU | Convolutional layer size=300  Dense layer=128 unit  Dropout=0.4 |
| Data preprocessing  Method1 and 2  -Smoothing, Segmentation Normalization  Feature extraction (Method 1)  -Wrapper subset  Classification  Method 1:  -Multilayer perceptron  Method 2:  -K means classifier [9] | MTX sensor  3D accelerometer  3D gyroscope  3D magnetometer | Sampling rate=100 HZ  No of feature=32  Window=6.40/s |
| Video stream  Image thresholding  -OTSU’S method  Hand contour extraction  -Convex hull detection  -Convex defects detection  Hand gesture extraction  Overlay and virtual object in AR [11] | OpenCV  Python  Simple webcam  IP camera | |
| Basic propositions  -Palm pose, Fingers  Development  -A C# interfaces, XAML text interface, A visual gesture builder  Hand pose estimation  -Convolution, Histogram, CTE training  Hand pose Pipeline  -Segmentation, Centering, Global pose classification  -Global pose refinement  Finger regression [12] | Corei7-372QM  2.6GHZ processes  550us single thread | 209 CTE predictors  21 CTE activator  Classifier  M=20 ferns  C=10 class  K=12 bits  Aggregation area=64X64 |
| Input signals  -Fed into HMM  Hmm classification  -Gyro, Acc, Kinect  Pooling [13] | Quad core  1.7 GHz  4GB RAM | 8HMM state  9-dimensional state |
| Matching technique  -DWT (Dynamic time wrapping)  -Elastic matching  Statistical model  -Particle filtering  Gesture recognition  HMM [14] | 1.7 GHz  4GB RAM | 9-dimensional state  30 Hz Kinect  200 Hz  Window size=9-19  Confidence =95%  Distance=1.2-3.5 m |
| Detecting multiple fingers  -Extracting hand gesture, Image binarization, finding fingertips, Finding palms center  Measuring fingertip trajectories  -Determining, Predicting  Gesture Recognition  -Detection of thumb  -Symbolic gesture recognition [15] | Linux OS  Pentium iii 500 MHz  Hitachi iP5000 image processing board  Nikon laired-s270 | Infrared temperature 34 degree  Trial duration 30s |
| 1.five hand gestures  2Optical flow is extracted for each two consecutive frames  3. An eight bin histograms based on the flow direction are created [16] | 60 videos. Out of 60 videos,  20 videos created by randomly,  3 videos from dataset | Average accuracy 80% |
| Extract HOG-LBP functionality, Use the function Radial Basis as kernel, Find the perfect parameter, Train the whole set using appropriate parameters received in the third step, Test the samples in the fourth step with the SVM model [39] | 2000 hand gesture color images | average accuracy  97.8% |
| Apply the threshold of depth,segment the region of interest [40] | Kinect input video feed. | 92.3% recognition rate |

iv. Dataset with accuracy

Table 6. Previous Research based on Datasets

|  |  |  |  |
| --- | --- | --- | --- |
| Dataset Name | Parameter Used | Numerical Estimation | |
| Unnamed Dataset [16] | Intuitive, easy to use, understanding ability and dynamic change and interaction. | | |
| Confusion matrix of the classification [17] | cut, shoot, pinch, scale, buckle, hold, grab, rotate, crawl, OK | | |
| Bioloid robot’s servo data using Dynamixel Library [18] | forward, right, left and stop | The precision rate between K-means and SVM is 4.45ms and SVM 10ms with an average recognition time using K-means. | |
| Hand gestures from the default database [19] | The model learned by the rules - based classifiers is represented by a set of if then rules. Exactness is then determined by (c / t) \* 100. | | |
| NUS Hand Posture Dataset-Ii [21] | The accuracy of hand gesture recognition is tested on subsets | The average accuracy of recognition for subsets A= 97.8% and B=95.07%. | |
| Hand Gesture Dataset LSP [22] | taken under different features such as scale, rotation, translation, illumination and noise. | The architecture of ADCNN achieved a precision of 99.73 percent, which was 4 percent higher than the performance of the basic CNN architecture | |
| Confusion matrix for NUS hand posture dataset. [23] | 25 samples in each of the class | 100% accuracy  Accuracy reduced in complex background | |
| Gesture Action Accuracies [24] | Gesture Sequence | On running the same algorithm, the beagle board is measured to be 3.2 times slower than a 2.24 GHz Intel dual core processor while the Beagle board XM is running on a 1 GHz OMAP processor. | |
| ChaLearn LAP IsoGD Dataset [25] | RGB-D depth sequences |  | |
| ChaLearn LAP IsoGD Dataset [27] | for feature extraction is C3D | The processing of feature extraction can reach 656 thanks to the speed of the implementation of C3D.fps. | |
| Mean values for the ISONORM usability questionnaire in 6 categories on a scale from-3 to+ 3. [33] | All ratings above 2,00 (highest ratings) most students believe in Construct3D | | |
| 1.MSR gesture 3D  2.SKIG  3.MSR action 3D  4.NTU [2] | 98.05%  97.31%  95.24%  68.66% | | |
| Nina pro database | CNN-65.4%, SVM-68.2% | | |
| Self-made dataset  (compared to GaS) [3] | Accuracy=4.7, Speed=4.3, Ease=3.9Perception=4.2 | | |
| Self-made video dataset [4] | 4 gesture | Good light=85%  Poor light=71.3% | |
| Self-made dataset [5] | Index finger, two fingers | Index=99.12%  Two=98.87% | |
| Self-made dataset [6] | 26 Alphabet  6 persons  1560 different gesture | 2 crosses=93.3%  10 crosses= 94.7% | |
| Self-made dataset [7] | 10 testers  330 cases  11 gesture | 95.1%  Processing time 55 ms | |
| Khusboo el data set [14] [8] | 10 subjects  600 examples  Stratified sampling | 98.88% | |
| Self-made dataset [9] | 5 subjects  15 hand gesture | 93.12% average accuracy | |
| CHEALEARN gesture dataset [10] | Harming distance  Euclidean distance  EJC  OCRM  MCM | From the data set every time it shows, from all the method cascading feature take least recognition time | |
| Self-made database [11] | 10 gesture | Right-left | 98% |
| Left-right | 95% |
| Bottom-top | 95% |
| Top L Bottom R | 90% |
| Bottom R Top L | 85% |
| Top R Bottom L | 85% |
| Bottom L Top R | 80% |
| Clockwise | 75% |
| Infinity | 75% |
| Anti-Clock wise | 80% |
| Self-made dataset [12] | 89333 of posture image  274068 fingertips image | Average accuracy 90% | |
| $1Unistroke Recognition Application [13] | 10 gesture | 91% accuracy | |
| Microsoft MSR dataset [14] | 10 gesture | 93% accuracy | |
| Self-made dataset [15] | 900 images | One finger 99.2%  Two fingers 97.5% | |
| Created 2 Datasets (Spotting framework is evaluated on two datasets collected from 11 users) [38] | Multiple user | accuracyDataset-I=71.5% and Dataset-II=69.2% | |
| NUS Hand Posture Dataset-II [39] | 4750 hand gesture images | Accuracy 97.8% (subset A) and 95.07% (subset B). | |
| 1500 × 5 vector dataset [40] | Accuracy rate obtained was  97.4%. | | |