2016 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)
19-21 December 2016, AISSMS, Pune, India

A Survey of 2D and 3D Imaging Used in Hand Gesture Recognition for Human-Computer Interaction (HCI)

Rajeshri R Itkarkar
E & C Department
BVB, Hubli, Karnataka, India
itkarkarrajashri@yahoo.com

Anilkumar V. Nandi
E & C Department
BVB, Hubli, Karnataka, India
anilnandy@bvb.edu

***Abstract*— Interaction of human body movements through any machines or computing device is a man-computer interaction (HCI). Human hand movements or gestures are the best ways of interacting with any machine, which allows the user to communicate and control any remotely placed device. This includes control of any home appliance, to control a robot, track a player or its body position, control movement and orientation of games and also control any man-machine interaction. With the rapid development in a 3D application and virtual environment, there is a need of devices which interact between the human and the computer system. This paper presents the study of the techniques applied for recognizing gesture in real time application. It also highlights on various techniques and the conditions under which the methodologies are employed. Further, the methodologies in terms of rate of recognition and limitations are explored. Finally we highlight the challenges in real-time gesture recognition.**

***Index Terms*— Hand gestures, Human-computer interaction, gesture recognition**

I. INTRODUCTION

Gesture recognition is the simplest method of interacting with machines through hand postures. As a result these gestures are used to improve the man-machine interaction. The movement and orientation of the hands help the consumer to easily interact with the machines, at different lighting conditions and background. Hand gestures are scientifically classified as static and dynamic [1] [3]. The static gesture is orientation and position of the hand in space during which no movement of hands is observed. The dynamic gesture is the orientation and position of the hand in space during movement of hand over time is observed. The main aim is to develop a system that identifies human postures and utilize them to interact or control any device. Two approaches are commonly used to identify the hand gestures, use of data gloves and vision based. The data glove method makes use of sensors for hand orientation and finger motion. However the device is expensive and restricts the movement of the hands with the use of gloves. This kind of method can be applied for 2D as well as 3D images. Though the vision based method requires only a camera, making the interaction between the human and machines adaptable without any use of the extra device. The problem faced with the recognition system is the background needs to be invariant, light problems, the distance of the users from the camera and the orientation of the camera [2] [3].

With the development of consumer products and mobile device, the interaction of hand gestures with mobiles, tablets, PCs has been increased. This trend is seen not only in 2D screens but also in the 3D world [14]. However, for a real-time implementation, hand tracking from a sequence of videos is important where movement is considered. The movement of the hand is generally observed in 3D space. Thus the limitation of 2D space is hand movement in the Z-axis. The recent technology in gesture recognition uses three dimensional cameras to capture the information in the form of RGB and depth data.

This article presents the theoretical background of hand gesture recognition in Section 2 survey in section 3 followed by the proposed methodology based on survey work.

II. BACKGROUND

Hand gesture recognition can be achieved for both glove based and vision based methods. As glove based has drawbacks with sensors and hand motion restriction, most research is based on vision based method. The vision-based method is further categorized as model-based and appearance based [2] [3]. The 3D model based has three different techniques, such as 3D textured volumetric, 3D geometric model and 3D skeleton model. One or more camera is needed to focus the target and the spatial parameters are calculated to recognize the gesture [3] [4]. This is the exact strategy as the model parameters are redesigned while identifying the matches of move in the temporal model [3]. The usually used method in three dimensional model based is textured kinematic/volumetric model as it contains high details of human body skeleton and skin information [3]. While the three dimensional geometric models are less precise as they contain high skeleton information and less skin information [3]. Appearance-based hand gesture representation strategies are classified as a 2D static model based and motion-based method. 2D static based method includes the use of markers to track motion, silhouette geometry based in which geometric parameters such perimeter, convexity, surface, rectangularity, Centroid and so on are utilized to identify the gesture [3,5]. Motion based models are used for identifying of an object or the motion of an object in an image succession.

2.1 Vision-Based Hand Gesture Recognition.

Vision-based hand gesture identification framework needs to meet the prerequisites which include real-time performance, accuracy, and robustness. The steps involved in vision-based two-dimensional recognition system are segmentation, feature

extraction and recognition. In the process of segmentation of the image, hand gesture recognition system uses skin color detection, shape detection, pixel detection and motion detection.. With accurate segmentation, the recognition system becomes more powerful and successful. The segmentation in gesture recognition system can be detected by *Skin-color based segmentation* [6] [7]. The RGB color image is transformed into YCbCr, YUV, HSV etc. form. RGB is default color space available. Image formation from RGB to any color space is acquired using linear or non-linear transformation. The RGB color space is changed over into another color space which is not sensitive to light variation. Variations can be introduced due to changing illumination conditions and/or camera attributes. Therefore, color-based strategies for hand detection need to employ some means for making up for this variability [3].

The shape of the hand can be used to identify the gesture termed as shape identification. The contour of the hand identifies the shape. Extracting the contour of hand gives more information in shape detection. Contour can be evaluated by edge detection, but with the hand edges there are background object edges appearing in the image. Pre-processing or post processing can be used to avoid the irrelevant background edges. Pixel detection is another method of finding out hands in gray level images which is based on their appearance and texture [3]. Another method used for detection is 3D Model based detection. The dimensions of the hand are obtained from the degree of freedom of the hand. Point and line features are used in kinematic hand models to recover angles formed at the joints of the hand. Hand postures estimated provides image features corresponding to the 3D model features [3].

Feature extraction is the frame-to-frame correspondence of the segmented hand regions or elements towards understanding the observed hand motion [3]. *Correlation-based feature tracking is* correlation-based template matching is used to track hand features across the frames. *Contour based tracking is used* to track hand regions in successive image frames. Harr-like Feature is one of the simple algorithms for feature extraction compared to the use of intensity values of pixels where a change in contrast value is considered between the adjacent rectangular groups of pixels. The ratio between the dark and bright areas within a kernel is evaluated. Four different features such as line features, edge feature, center surround feature and diagonal features are considered [21] [29]. Thus the Harr-like features extracted from the gray image depends on the-the value of the difference of the sums computed over the rectangular region. The merit of Harr-like feature extraction is that it is faster than the pixel based system and also robust to noise indifferent light condition. HOG (histogram of gradient) is the feature extraction method in which the optical changes, illumination and the rotation in the image are considered. In this method the mage is divided into a number of cells, with a cell consisting of a number of pixels/ cell (eg. 8*8 pixels in a cell), then the gradient of the histogram is evaluated called as HOG descriptor [30] . The numbers of cells form a block and number blocks form a window. Thus the HOG gradient vector is calculated and further given to the classifier [29]. The binary silhouette is a feature highlights method in which binary silhouettes are generated from a sequence of video frames using background subtraction and binarization. The statistical method of feature highlights for high dimensional data is Principle component analysis (PCA). It is the method in which the statistical features such as standard deviation, covariance, eigenvectors and eigenvalues are evaluated. PCA reduces the original high dimensional variable into lower dimension orthogonal components.

Feature extraction is followed by gesture recognition. To detect static gestures a general classifier or a template-matching algorithm can be used. However, dynamic hand gestures have a temporal aspect and require techniques that have to handle higher dimensions like Hidden Markov Models (HMM), Support Vector machine(SVM), K-Near neighbor, Dynamic time wrapping method are employed. *HMM-Hidden Markova Model (*HMM) is defined as a set of states of which one state is the initial state, a set of output symbols, and a set of state transitions. Each state transition is represented by the state from which the transition starts, the state to which transition moves, the output symbol generated, and the probability that the transition is taken. In the context of hand gesture recognition, each state could represent a set of possible hand positions [22]. The state transitions represent the probability that a certain hand position transitions into another; the corresponding output symbol represents a specific posture and a sequence of output symbols represents the hand gesture. Firstly the multi-dimensional vector sequences are converted to one-dimensional symbol sequences. The multi-dimensional feature vector sequence is quantized into a finite symbol sequence [22]. A left-right banded structure, in which a state can only go to the next state or go to itself, is adopted for the HMMs. The observation probability and the state transition probabilities are described with $P(Oi \mid Si)$ and $P(Si \mid Si-1)$ respectively, where Si and $Si-1$ represent the current state and the previous state respectively and Oi represents the observation symbol at state Si [22] [23]. For the hand gesture recognition, 'M' Hidden Markov Models corresponding to 'M' hand gestures are trained to obtain a vector called as HMM vector. Next a Baum-Welch algorithm is used to estimate the parameters of the HMMs in the training hand gesture trajectories. During the recognition phase, an incoming hand gesture trajectory is converted to an observation symbol set O, which is the input to the HMM vector H, then the hand gesture is matched with the vector O [22] [23]. The rate of recognition using HMM [22] is 85%. The limitation in recognition rate in [22] was that they did not have enough training data to make a good estimate of the HMM model parameters. The recognition accuracy by HMM in [23] is of 90.6 % and 91.6 % under the dark and strong illumination, respectively. This is because the illumination condition directly influences the accuracy of gesture extraction, which leads to incorrect gesture classification results. Overall, the system yields an average accuracy of 92 % under different illumination conditions and complicated background and in [23] it is 92.1%.

SVM (Support Vector Machine) is another supervised classifier which works on high dimensional data. It selects the

features called as support features from each class and plots an n-dimensional vector space with the value of each feature being the value of a particular coordinate. The SVM classifier is a binary classifier which searches for a hyper plane as a decision function for optimal result. Many algorithms perform linear classification, but SVM can also perform non-linear classification where kernel method is used. In the non-linear classification method, the data points are represented on a 3-dimensional plane and a hyper-plane is used to classify it. In linear classification with two classes, the feature vectors in training samples are compared with the feature vectors in the classification phase. While in non-linear classification, as previously said a kernel is used. Mathematically, any function K(x, y) is a kernel and can be written as $K(x, y) = \Theta(x) \cdot \Theta(y)$, where Θ is the function that maps an instant into the higher-dimensional feature space. The kernel function represents a dot product in the feature space created by Θ [24]. With non-linear SVM classifier one-against-all classification is also performed where ***c*** classifiers have to be designed for each one of them used to separate one class from the rest [24]. The recognition rate in [24] is 99.4% with limitations of the camera used for the creation of the database, while in [25] the accuracy achieved is 98.5% but for complex or two hand gesture the accuracy dropped to 75.17 %.

k-Nearest Neighbor Classification Algorithm(kNN Algorithm) is one of the classifier based on distance metrics. The distance of training feature vectors when compared with feature vectors of testing images the minimum distance, called as the k-minor distance is evaluated. The method can be implemented with different data structures and distance metrics, to achieve good evaluation time [26]. Stefanet al. [27] used the nearest neighbor technique to classify dynamic gestures based on their feature vectors, with data acquired from video streams. A recognition rate of 96.3% was achieved.

The detection of hand gesture can be also achieved by an efficient technique i.e. neural network. The neural network is a function made up of simple elements inspired from the biological nervous system operating in parallel. It is a network determined by connections of values of the elements which are called as weights. The neural network is trained by varying the values of the weights between the elements to perform a particular function. The networks are adjusted, so that a particular input leads to a specific target output. The neural networks are developed through learning rather than programming and are flexible in a changing environment. Neural networks can handle very complex interactions and can easily model the data. The limitation of neural networks is their inability to explain the model built and takes the time to train a model. Tin Hninn Maung [27] implemented the hand gesture using the learning rule in which the rule provided with a set of the training set of proper network behaviour which is applied as an input to the network, and the corresponding target output is compared and observed. The learning rule is then used to vary the weights of the network in order to move the network outputs closer to the targets.

Dynamic Time Wrapping: DTW is an algorithm which evaluates the similarities between two sequences which differ in time. With hand gestures the two sequences X and Y at two different time instants are observed and the Euclidean distance is evaluated to recognize the gesture.

Template matching is the simplest algorithm used as a classifier in which the auto-correlation, cross-correlation and cross-covariance of the given image and the template image are matched. The matching evaluates the minimum difference called as mean square error (MSE). The minimum MSE represents the recognized gesture.

III. RELATED WORK

A lot of research since 1997 is carried out on 2D hand gesture recognition, which is summarized as shown in table 1.

3.1 Survey of 3D Imaging Gesture Recognition System

The limitations of existing 2D hand tracking system are majority related to camera position, its orientation, light conditions and background aspects. Now a day's gesture recognition is enabling humans and machines to interact and control TV, lights, devices etc. There is lot of research carried with 2D technology, but with introduction of 3D technique lot of applications can be controlled such as in gaming, consumer and mobile products, in industry to determine manufacturing flaws, remote and virtual care in medical, vehicles can be manufactured with 3D vision of preventing accidents, both day and night., 3D sensors enable high-quality video processing, facial recognition, 3D imaging, noise cancellation etc. Special 3D cameras with principle such as time of flight (ToF), structured light, stereo vision, Fresnel Hologram, Optical interferometry technique, Shape from Shading, 3D integral imaging, shape from focus/defocus and shape from texture are used in which ToF and structured light based are only two of the several existing 3D imaging techniques [32]. The Kinect sensor developed by Microsoft in 2010 which works on the principle of structured light plays an important role in the applications described. The ToF cameras are costly than the structured light cameras. Kinect consists of RGB camera and a depth camera, where the depth camera allows to build a 3D view of the environment. Frol et.al highlights on different existing 3D approaches based computational cost and reliability. It also presents the challenges and opportunities in HCI [32]. Kam et. al used a Kinect camera for real-time gesture recognition using skeleton based feature extraction and nearest-neighbour as a classifier with a close accuracy of 100%. The limitation is with Kinect camera which gives poor performance in outdoor applications [28]. Lubo et. al in [33] proposed spherical coordinate Feature with location Feature and recognized the gesture using SVM and extreme learning machine(ELM) neural network method for Chinese language recognition with the rate of 61.46% and 69.32 %. Due to high frame rate the joints produced redundant points and the invalid frames reduced the accuracy. Xi Chen et. al proposed a gesture recognition system using HOG, LBP and Gabor filter based extraction. The ELM classifier was implemented with a combined accuracy of 85.5% accuracy. The system is applicable for sign language recognition which 3D appearance based [34]. Harrison Cook et.al [35] implemented gesture recognition by using Kinect v2 sensor. First valid matrix by using the closest pixel and the furthest pixel is calculated from

the depth data. Then contour is evaluated by Moore-neighbour algorithm and fingertip detection is done by using the k-curvature algorithm. The system is implemented for a game called as fruit slicing game with the help of gestures. Limitations of the system are a lot, precision is needed in terms of angle and position of the hand and also it does not work with different hands. E. Escobedo-Cardenas et. al [36] proposed a new approach for dynamic hand gesture recognition using intensity, depth and skeleton joint data captured through Kinect sensor. Local and global features are extracted using SIFT descriptor from the 3D trajectories and the depth data from the Kinect sensor. The classification is done by using SVM with the rate of 88.3 %. Time invariability was the limitation due to the speed of the gestures. Yanmei Chen et.al in [37] used vector quantizer to convert the hand orientation to a code word to form a discrete vector. The vector further was classified using SVM and HMM, with a recognition rate of 95.42 % and 82.86%, but for some gestures such O, T and 2 orientation feature did not work as accuracy was less for these gestures. Javier Molina et. al [19] used ToF camera in which features such as geodesic center, length and orientation and minimum depth point of the hand silhouette were extracted. First static hand posture was detected from which the dynamic hand gestures were detected using SVM and FSM with an accuracy of 93.9%.

The study reveals that the feature extraction and recognition methods have some limitations, like HOG extraction method, same gesture may have a different histogram and same gestures may have different histogram but still it is computationally faster. Whereas extraction is done by using PCA has higher computational cost. Colour image extraction is less reliable if the background contains skin color, while the nearest neighbor is sensitive to noise. Thus good features can be extracted from the hand such contour, fingertip, finger direction etc. which are not reliable due to self-occlusion and lighting. Hence we have to rely on features such as color, texture etc. The classification methods commonly used are a neural network, K-NN, SVM, HMM and template matching. Neural network though efficient method takes the time to train the model. SVM and HMM take less time train the models and also give high accuracy. Thus ANN is more prone to over fitting than SVM and HMM.

3.2 Proposed Methodology

So far the research work is carried out in two phase, in phase-1, recognition for 2D image in static and dynamic is implemented. 3D work is in progress. With static gestures, the PCA (principal component analysis) algorithm is used for feature extraction and Euclidian distance evaluation for recognition. Accuracy obtaine is 97.22%. Figure 1 shows the results for number 10 in sign language recognition. Futher the gesture recognition is implemented for dynamic gestures. In dynamic gesture recognition the system is implemented in four modules, segmentation, contour tracking and estimating the convex hull, feature extraction and classification. Features such as centroid, minor axis length, major axis length and area are calculated and classifier used is Harris correlate algorithm. The the overall accuracy obtained for different signs is 81.66% in real time under dark illumination. Figure 2 represents the result for dynamic gesture for alphabet “C”.

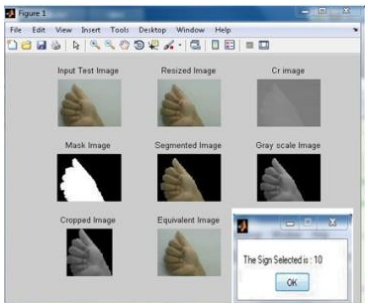

Fig 1. Result for number 10 in sign language recognition.

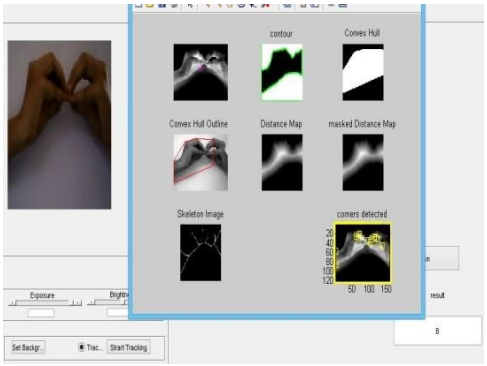

Fig 2. Result for Alphabet ‘C’ for sign language recognition

Phase-2 of the research work is in progress which is being implemented for 3D images. The methodology proposed for 3D is contour extraction with K-NN and SVM classifier to increase the accuracy.

3.3 Challenges in Gesture Recognition

One of challenge in vision-based gesture recognition is the background. The system should be robust to different backgrounds. Next is the complex gestures i.e multiple gestures. Another challenge is due to multiple cameras, which creates occlusion problems. With 3D imaging the resolution of the camera, the cost of the camera and the space occupied by the setup are the major challenges

IV CONCLUSION

In summary, the vision-based gesture recognition is more comfortable than the glove based, as the sensors attached to the gloves restrict the gestures. In vision based the 2D research has reached a better level in terms of accuracy and algorithm complexity. For 3D imaging the challenges such as the environment condition, the hardware setup compactness, the resolution of the camera matters and the work has to be concentrated on. The methodologies employed in recognition have proved that there is no specific method which is better than the other. Hence the performance is not methodology dependent. With the application of hand gesture recognition in the consumer and the advanced devices, the accuracy and the performance of the system may vary and the research may progress accordingly.

Acknowledgement: I sincerely thanks management of JSPM’s Rajarshi shahu college of Engineering, Pune, where I am presently working as Assistant Professor in E & TC department for supporting and allowing me for the research work at BVB Hubli, Karnataka . I sincerely thank my guide Dr. Anilkumar Nandi for his continuous and motivational support.

REFERENCES

1] Henrik Birk , homas Baltzer Moeslund, Recognizing Gestures From the Hand Alphabet Using Principal Component Analysis, Master's Thesis, Laboratory of Image Analysis, Aalborg University, Denmark, 1996.

[2] G.R.S Murthy et.al A review of vision based hand gestures recognition, International Journal of Information Technology and Knowledge Management July - December 2009, Volume 2, No. 2, pp. 405-410.

[3] Siddharth S. Rautaray et.al Vision-based hand gesture recognition for human-computer interaction: a survey, Artif Intell RevDOI 10.1007/s10462-012-9356-9

[4] McNeill D (1992) Hand and mind what gestures reveal about thought. University of Chicago Press. ISBN:9780226561325

[5] Bretzner L, Laptev I, Lindeberg T (2002) Hand gesture recognition using multi-scale color features, hierarchical models and particle filtering. In: Fifth IEEE international conference on automatic face and gesture recognition, pp 405–410. DOI:10.1109/AFGR.2002.1004190

[6] Choi J, Park H, Park, J-I (2011) Hand Shape Recognition Using Distance Transform and Shape Decomposition. 18th IEEE International Conference on Image Processing (ICIP)

[7] Qing Chen, et.al , Vision-based Hand Gesture Recognition Using Haar-like Features, Instrumentation and Measurement Technology Conference – IMTC 2007Warsaw, Poland, May 1-3, 2007

[8] Zhu X, Wong KK (2012) Single-Frame Hand Gesture Recognition Using Color and Depth Kernel Descriptors. 21st International Conference on Pattern Recognition (ICPR)

[9] Asadollah Shahbahrami et. al., Human Computer Interaction Using Vision-Based Hand Gesture Recognition, Journal of Advances in Computer Research 2 (2010) 21-30.

[10] Wong, F., Gaus, Y.F.A.; Hidden Markov Model-Based Gesture Recognition with Overlapping Hand-Head/Hand-Hand Estimated Using Kalman Filter, Third International Conference on ,Intelligent Systems, Modelling and Simulation (ISMS), vol., no., pp.262,267, 8-10 Feb. 2012.

[11] Hong-Min Zhu, Chi-Man Pun, Real-time Hand Gesture Recognition from Depth Image Sequences, 978-0-7695-4778-7/12 $26.00 © 2012 IEEE DOI 10.1109/CGIV.2012.13

[12] Dung-Hua Liou, Chen-Chiung Hsieh , Novel Haar features for real-time hand gesture recognition using SVM, J Real-Time Image Proc (2015) 10:357–370 DOI 10.1007/s11554-012-0295-0

[13] Emil M. Petriu, Nasser H. Dardas, Hand Gesture Detection and Recognition Using Principal Component Analysis, 978-1-61284-925-6/11©2011 IEEE.

[14] Yuan Yao, Yun Fu, Contour Model-Based Hand-Gesture Recognition using the Kinect Sensor, IEEE transactions on circuits and systems for video technology, vol. 24, NO. 11, Nov 2014.

[15]Sameem Abdul-Kareem, Haitham Hasan Human–computer interaction using vision-based hand gesture recognition systems: a survey, Neural Comput & Applic (2014) 25:251–261 DOI 10.1007/s00521-013-1481-0

[16] WeidongWang, YuehaiWang, Yafei Zhao, A Real-time HandGesture Recognition Method, 978-1-4577-0321-8/11 ©2011 IEEE

[17] Miskovic, N.; Vukic, Z., Menix, M.; Interpretation of divers symbolic language by using hidden Markov models, 37th International Convention on , Information and Communication Technology, Electronics and Microelectronics (MIPRO), vol., no., pp.976,981, 26-30 May 2014

[18] Chih-Hung Wu &Wei-Lun Chen & Chang Hong Lin, Depth-based hand gesture recognition, Multimedia Tools Appl DOI 10.1007/s11042-015-2632-3

[19] Javier Molina, Marcos Escudero-Viñolo, Alessandro Signoriello, Montse Pardàs, Christian Ferrán, Jesús Bescós, Ferran Marqués, José and Springer 2012.

[20] Paulo Trigueiros · Fernando Ribeiro ·Luis Paulo Reis, Generic System for Human-Computer Gesture Interaction Applications on Sign Language Recognition and Robotic, Soccer Refereeing, J Intell Robot Syst DOI 10.1007/s10846-015-0192-4 Machine Vision and Applications (2013) 24:187–204 DOI 10.1007/s00138-011-0364-6 from time-of-flight camera video using static and dynamic models, M. Martínez, Real-time user independent hand gesture recognition

[21] Alexandre Savaris, Aldo vonWangenheim, Comparative evaluation of static gesture recognition techniques based on nearest neighbor, neural networks and support vector machines, J Braz Comput Soc (2010) 16: 147–162 DOI 10.1007/s13173-010-0009-z

[22] Feng-Sheng Chen, Chih-Ming Fu, Chung-Lin Huang, Hand gesture recognition using a real-time tracking method and hidden Markov models, Image and Vision Computing 21 (2003) 745–758, 2003 Elsevier Science B.V

[23] Dan Xu · XinyuWu, Yen-Lun Chen, Yangsheng Xu, Online Dynamic Gesture Recognition for Human-Robot Interaction, Springer Science+Business Media Dordrecht 2014, J Intell Robot Syst (2015) 77:583–596 DOI 10.1007/s10846-014-0039-4

[24] Paulo Trigueiros, Fernando Ribeiro, Luis Paulo Reis , Generic System for Human-Computer Gesture Interaction: Applications on Sign Language Recognition and Robotic Soccer Refereeing, Springer Science Business Media Dordrecht 2015

[25]E.M.P.S.Edirisinghe,P.W.G.D.Shaminda,I.D.T.Prabash,N.S.Hettarachchige , L.Seneviratne and U.A.A.Niroshika, Enhanced Feature Extraction Method for Hand Gesture Recognition using Support Vector Machine, 2013 IEEE 8th International Conference on Industrial and Information Systems, ICIIS 2013, Aug. 18-20, 2013, Sri Lanka

[26] Alexandre Savaris · Aldo vonWangenheim, Comparative evaluation of static gesture recognition techniques based on nearest neighbor, neural networks and support vector machines, The Brazilian Computer Society 2010, J Braz Comput Soc (2010) 16: 147–162

[27] Athitsos v, Stefan A, Alon J, (2008), Translation and scale invariant gesture recognition in complex scenes. In: Proc 1st conf. pervasive technol relat assit env, pp 1-8 doi: 10.1145/1389586, 1389595

[28] Kam Lai, Janusz Konrad, and Prakash Ishwar, A gesture-driven computer interface using Kinect, 978-1-4673-1830-3/12/$31.00 ©2012 IEEE

[29] Haitham Badi, Sameem Abdul Kareem and Sabah Husien, Feature Extraction Technique for Static Hand Gesture Recognition, Science Gate Publishing, DO 10.15579/gcsr.vol3.ch2, GCSR Vol. 3, pp. 19-41, 2015

[30] Kai-ping Feng, Fang Yuan, Static Hand Gesture Recognition Based on HOG Characters and Support Vector Machines, 978-1-4799-2716-6/13/$31.IEEE

[31] Marco Maisto, Massimo Panella, *Member, IEEE*, Luca Liparulo, and Andrea Proietti, An Accurate Algorithm for the Identification of Fingertips Using an RGB-D Camera, 2156-3357/$31.00 © 2013 IEEE

[32] Frol Periverzov • Horea T. Ilieş, 3D Imaging for Hand Gesture Recognition: Exploring the Software-Hardware Interaction of Current Technologies, 3D Research Center, Kwangwoon University

[33] Lubo Geng, Xin Ma, Bingxia Xue, Hanbo Wu, Jason Gu, Yibin Li, Combining Features for Chinese Sign Language Recognition with Kinect, 978-1-4799-2837-8/14/$31.00 ©2014 IEEE

[34] Xi Chen and Markus Koskela, Using Appearance-Based Hand Features For Dynamic RGB-D Gesture Recognition, 1051-4651/14 $31.00 © 2014 IEEE DOI 10.1109/ICPR.2014.79

[35] Harrison Cook, Quang Vinh Nguyen, Simeon Simoff, Tomas Trescak, Dean Preston, A Close-Range Gesture Interaction with Kinect, 978-1-4673-7343-2/15/$31.00 ©2015 IEEE

[36] E. Escobedo-Cardenas, G. Camara-Chavez, A Robust Gesture Recognition Using Hand Local Data And Skeleton Trajectory, 978-1-4799-8339-1/15/$31.00 ©2015 IEEE

[37] Yanmei Chen, Bing Luo, Yen-Lun Chen, Guoyuan Liang, Xinyu Wu, A Real-time Dynamic Hand Gesture Recognition System Using Kinect Sensor, 978-1-4673-9675-2/15/$31.00 © 2015 IEEE

TABLE 1. Summary of Literature Survey

Author	Static/Dynamic Gesture	Method	Accuracy	Limitations
Hong-Min Zhu et.al.[11]	Dynamic	Hidden Markova model	92%	Accuracy can be improved for complex hand structures.
Yuan Yao et.al.[14]	Dynamic	3D model based and pixel based	74.65% 52.31%	Due to camera setup, feature extraction and recognition is limited for long arm and contour matching algorithm limits the accuracy.
Qing Chen et.al.[7]	Dynamic	Haar features and Adaptive Boosting	90%	For context-free grammar gesture a new methodology has to be developed to improve the recognition rate.
Chen-Chiung Hsieh et.al.[12]	Dynamic	Haar feature and Support vector machine.	95.37%	The algorithm can be further used integrate all the captured data from multi-cameras to recover more motion information accurately.
Nasser H. Dardas et.al.[13]	Static	Principal component analysis and projection method	80% (PCA)	-
Yafei Zhao et.al.[16]	Dynamic	Histogram of oriented gradients(HOG) and Nearest neighbor	91%	Accuracy affected for some of the gestures due to bad localization caused by shadows in the hand.
Menix, M et.al.[17]	Static and dynamic	HMM based on mixture of Gaussian	72%	Less accuracy as the system is implemented for drivers under sea water.
Chih-Hung et.al [18]	Static and dynamic	Uses depth information with SVM classifier	87.6%	The accuracy is affected due to the distance from the camera and the depth of the hand.
Gaus, Y.F et.al [10]	Static	Kalman filter and HMM	83%	The limitation with real-time processing.
Paulo Trigueiros et.al.[20]	Static and dynamic	SVM and HMM	99.4 % (SVM) 93.7% (HMM)	Limitations were due to camera
Javier Molina et.al.[19]	Static and dynamic	Low-level extraction with time- of- flight camera	90%	Improvement in the algorithm to make the system more robust to noise. The system must be used indoor, since the selected camera does not work well under sunlight conditions.
Alexandre Savaris et.al.[21]	Static	k-NN, NN& SVM	90% (SVM)	