

Hand Gesture Recognition

Jay Prakash, Uma Kant Gautam

Abstract: As our society and technology is growing day by day, the use of gadgets like mobile phone, computers are also increasing. In our daily life we use numbers of way of communication which include speaking, writing, by some body movement but while in case machine we still stuck with typing or speaking so we need some advancement so that we can communicate with machine in some body movement also. This way of communication in which any type of body movement is involved called Gestures. In other words, Gesture is non-vocal way of communication which uses hand motion, different postures of body, face expressions. So to make machine smart we are enabling our machine to take command by recognizing different hand gestures. Hand gestures is used as a input in our system.

Index Terms: Computer Vision Based Approach, Hand Gesture Recognition, Human Computer Interface (HCI), Instrumented Glove, Non-Verbal language.

I. INTRODUCTION

In this whole world, there is a vast development of computing techniques and due to ubiquitous methods of computing, current user interaction with the pointing and positioning devices such as mouse, keyboard and pen are not that sufficient. These devices are only limited so the commands set is also limited. Making use of body parts for interaction such as use of hands, is a better option. Hands can be used as an input device for providing natural interaction (1)

To detect hand gesture recognition (2) there are basically two ways. First one is using instrumental and other one is vision-based recognition. In instrumental glove, there is a glove (3) in which some sensors are fixed on the glove and to take input from that glove, some person must wear it and then according to hand movement or finger movement the input for hand gesture are produced. This instrumented glove work on photocell. As the light hits on photocell cause of movement of hands and fingers and by this input also vary. Instrumented glove is less used as its cost is high and while taking inputs in this technique more chances of error as to hold hand in same for long time is difficult. To overcome this approach vision based approach came into existence.

In vision based (4), inputs are taken from a webcam and but this method is more complex than glove method because in this number of camera and their position all matters. In this approach, visibility is main component. Capturing image from stream then detecting skin and differentiate it from background as background could also have same color as skin so all these things make this vision based complex. But it is still quite popular because chances of error less and efficiency is high. After taking input from either

by these two we save that inputs to databases by assigning some task. Now when we give some commands to our system by hand gestures then first machine capture our command as a image then compare this with database and if any image found in database then task assigned to that will be perform. This idea of hand gesture recognition is very useful for normal user but it is very difficult to implement by programmers but still many programmers working on it and we can see many gadgets are in market which work on hand gesture. Basically the idea is to make a machine intelligent and develop a routine that the machine understand the human language efficiently. Gestures could be say easy way of HUMAN COMPUTER INTERFACES (HCI), so that the computer can understand the speech, facial expression and human gestures. Gestures are non-vocal way of communication like posing a victory sign from fingers in front of smart phone camera for clicking photos.

II. RELATED WORK

Till now there is many hand gesture recognition technologies have evolved. Each technology has its own advantages and disadvantages. First technique uses wires in which for taking input users tied up with wires to connect with computer system. Disadvantage of this technology is user cannot move freely and it was difficult to work with wires without getting any harm. Instrumented gloves are used in this technology for hand gestures. In instrumented glove some sensors are used which gave reading to the system when the hand or fingers do any motion. Data glove optimal results but they are very expensive that's why normal person cannot own it and also it can easily disturbed so managing it also difficult. As wire are very difficult to handle so these wires are replaced with optical markers. In this also user wear glove but working in glove is now different now an infra-red light source placed at one end and a screen at other end. when hand or finger moved then intensity and direction of light changes and that's how for different gestures different light intensity and according to that gestures are recognized. This optical way is also good but it's configuration is very complex. After this Image based hand gesture recognition technique introduced in which image is processed like image shape, size, color etc. these features are used to recognize the gestures. But working with is bit difficult as compare to previously proposed one because in this to detect color, shape, size is bit difficult as background and skin color could be same and different person could have different size of hand. Illumination also have important role in image based because of illumination some noise could be removed by image while may be on less light the noise could be added to image which change the result.

Revised Manuscript Received on December 22, 2018.

Jay Prakash, Department of Computer Science and Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur.

Uma Kant Gautam, Department of Computer Science and Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur



Hand Gesture Recognition

After all these techniques the Vision Based approach introduced which worked mainly on shape. As every person have one thumb and four fingers in normal condition so this technique is more successful for hand gesture recognition than other techniques. This method is very successful also described in this paper Mehra et. al. (5) for hand gesture recognition.

According to Anant Atray (6) in his paper "*Automatic Hand Gesture Recognition*" he is using best algorithm for hand gesture recognition which give optimal result and complexity is very low but there is one limitation in his project that if the skin color is same as background or environment where hand present then his system is not able to distinguish between skin color and background which is major issue cause it can change our entire result and can lead to false output. So in my project we are using YUV color space and camshift algorithms which help to differentiate between skin color and background.

According to T. Freeman et. al. (6) they are using the orientation histogram to classify different gestures and interpolation. Histogram is very simple and powerful as even in different lights and colors the gesture recognition will be always correct. The main advantage of histogram is in dynamic gestures because in dynamic gesture the movement of hand is changed over time period so according to this paper histogram helps to find dynamic gestures very easily. But only limitation in using histograms that similar gestures can have different types of histogram orientations and similar histogram orientation can represent completely different gestures so this is bit confusing in gestures using orientation histogram (6).

As gesture recognition is very useful for nowadays so some researchers also forming Gesture recognition using recurrent neural networks. According to Murakami et.al. (7) recurrent neural network also has its own advantage in gesture recognition as in sign language to differentiate between words like trained-untrained, mom-dad, present-absent are very hard. So in this neural network help a lot cause using this our system can capture past history to identify the word are used in pairs that are confusing. But recurrent neural network in gesture recognition has its own disadvantages it is very slow means even to learn only 10 to 20 words while training the system, the system can take 4 to 5 days which is large amount of time that's why in my system we are not using neural network for gesture recognition (7).

Some researchers used algorithms based on Hidden Markov Model and some use the algorithms based on Genetic Algorithms. According to X. H. Wang et.al.(9) Genetic Algorithm(GA) convert capture input as a discrete points and after that convert the problem of recognition gesture image into problem of combinatorial optimization that include discrete points and then gesture recognition algorithm applied to detect gestures. But This genetic algorithm have disadvantage that it require high amount of training data and long period of time. So, we are using Hidden Markov Model(HMM) is our work because learning algorithms used for HMM is very efficient. Using HMM learning can take place directly from sequence of raw data. Our work using HMM which also advantage over genetic algorithms that it uses less number of training data as compare to GA and also training time is less in HMM.

A. Research Status

Many researchers and students work on Hand Gesture Recognition so they use different approaches to make their system efficient but in each work there is always some limitation which cannot be nullify so to make our system efficient we are taking help of other research papers and analyze where the author made errors. Like some researchers used algorithm based on Genetic Algorithms (8) and also algorithms based on neural network (12) these two methods are good in gesture observation purpose but inefficient as compare to Hidden Markov Model. Both of take huge amount of time to train the system and also large amount of training data that's why we are using algorithm based on Hidden Markov Model as it comparatively faster than both GA and Recurrent Neural Network.

There is also some problem with many system which are using histograms and also skin detection problem (13). Skin problem issue is major because in this noise increases drastically if skin color and background color is similar that's why we are using YUV color space and Camshift algorithm which to distinguish between background and skin color.

III. HAND GESTURE RECOGNITION BASICS

A. Human Computer Interface(HCI)

Developing some technology that let human to interact or communicate with machine is Human Computer Interface. To interact with computer human can use many things like hand gestures, eye recognition or some other device like instrumented glove etc.

B. Gestures

Gesture (14) is also a way of communication like we communicate in daily life. But in daily life, we used verbal and vocal communication while gesture is non verbal way of communication or we can say it is a silent communication.

Gestures constitute every type of communication except verbal and written. The other way of communication apart from that is making hand movement, giving facial expressions, giving different type of body postures these all are gestures.

We can say gesture is the body language that means to communicate with others by just making the movement and motion of body parts. Gesture is very quick communication if both parties understands the gesture meaning same. By this we can also show our emotions by just some movements.

According to movement gestures are of two types :

a. Dynamic Gesture - Change over period of time_(16).

Examples of dynamic gesture :

Waving of hand means "goodbye".



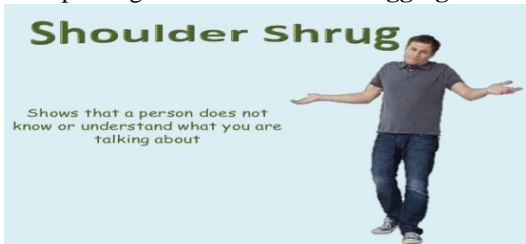
b. Static Gesture -- observed over a spurt of time_(17).
It is type of static gestures_(12) which means "stop".



C. Gesture Recognition

Understanding a full message, interpretation of all static and dynamic gestures is necessary for over a period of time. This complex process is **GESTURE RECOGNITION**.

example of gesture: **Shoulder Shrugging**



IV. PROPOSED WORK

There are many use of Hand Gesture Recognition and that's the reason significant number of Gesture Based application evolved, Some of them are listed below :

A. 3 Dimensional Geometry Design

Auto CAD (computer aided design) is an Human Computer Interface which is used for designing and drafting of 2 dimensional and 3 dimensional images. By using mouse and keyboard it is difficult for an programmer or user to design 3D design because making a 3D images involves all 6 Degree of Freedom(DOF) and allocating points in space using mouse is very hectic and complex. Now CAD provides facility to translate points or rotate points of image in any direction. Using this we can also see image from each and every direction according to our requirement to analyze it (9). *Massachusetts institute of technology* (10) has come up with the 3D RAW technology that uses a pen embedded in polhemus device to track the pen position and orientation in 3D.

B. Telepresence

Tele Presence (2) is way of communicating to person which are not in same room by using internet services like digital video or 3D visualizer. Tele Presence advantage is that any user can remotely present on location where he is permit to connect itself. For instances, in many companies, if employees of company in different countries or in different places then meeting still can be done by using video camera, microphone and video screens. The Tele Presence include collaborations especially instructions often, depends on the physical act of one person showing another person how to do something and even if your Tele presence robot has an arm of two it may not be at all intuitive for a remote user have effective direct interactions.

C. Virtual Reality

Virtual reality (11) is computer generated 3 dimensional environment with the help of software for users so that user can run its program, test its system by assuming it be a real

environment. The virtual environment that are using in present days can be displayed on screens and allow user to implement all of system applications through it. We can divide Virtual reality into:

- a. Forming a real environment copy or simulation for testing and training of software , project or systems. We can also use this for education purpose.
- b. We can develop a environment which is like real life places but actually doesn't exist . For example in Games like Pubg and GTA Vice city we see many maps or places which is same as real life places but in actual don't exist.

D. Sign Language

This language is different from our daily language we use verbally or which we speak. This sign language (2) includes all the non- verbal way of communication (5) likewise the movement of body on any statement made, gesture made by hands to illustrate things to some person or making expressions by face, All these things are included in sign language. This language is frequently used by deaf and mute persons.

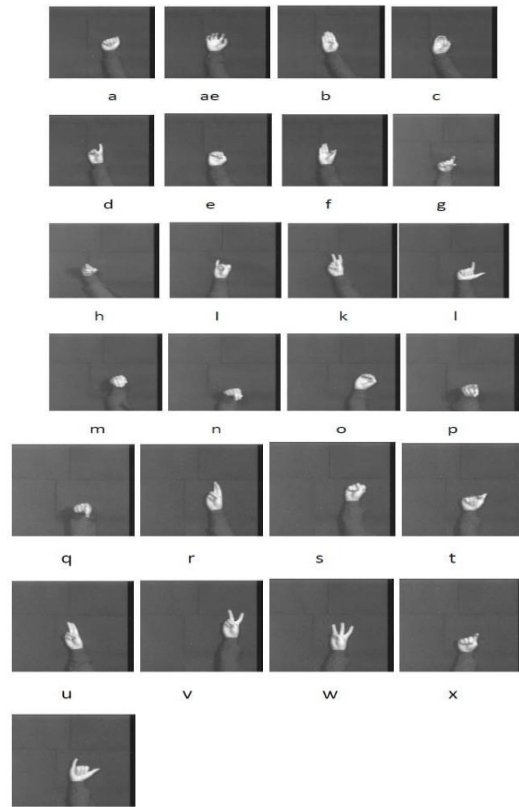


Fig 1. Gestures used for English Alphabet

Before this language it is very difficult to communicate with some physically challenged persons. Some time they understand you but cannot express themselves and other time they can express themselves but cannot understand you.

After this language, by watching, a deaf person can easily understand what you want to say and a mute person can even express themselves.



Hand Gesture Recognition

Even in some news channels we can see one side reporter speak and according to that reporter report a person besides it use sign language to sparse that report in deaf person too and also by that a normal person can learn how to correctly use that sign language.

V. ALGORITHMIC TECHNIQUES USED FOR RECOGNITION OF HAND GESTURES

To collect raw data we use either vision or glove based data collection system and various algorithms (14) used in order to collect the raw data smoothly and correctly.

Various algorithm used are :

A. Template Matching

The template matching method for hand recognition postures and gestures recognition used experimental method to know the required number of template of a certain gesture to be taken that should be saved on the database for the matching process of the algorithm. If the system will not be able to detect and recognizes the gesture given with the templates an additional templates must be trained and stored in the database until the system accurately recognize the gestures. The proponents will sum up all the time in second under a certain number the same number of template gesture.

B. Analyzing Extracted Features

When we recognizing some image or some pattern then first we gather information on low level or we can say we identify features of raw data and according to features we categorize the inputs is called Feature Extraction.

A tough feature will be invariant, meaning that if the image is rotates shrunk or enlarged, or translated, the value for the recognize hand gestures and postures.

C. Active Shapes Model

The Active Shape Model is training is done by manually drawing 3 dimensional surfaces images. The Active shape Model (9) find the difference in the data by which it is training using Principal Component Analysis .

Active shape model find best match for object or input outline according to data stored in database during training process. It iteratively checks every image and frame then try to find the best image which have similar control points as training data image.

D. Analyzing Principal Component

Analyzing principal component is a important technique to understand in the field of statistics and data science interrelated variables is called Principal Components Analysis. When putting the students online to technical didn't fully address our needs and provided conflicting reducing the dimensions of the features spaces is called dimensionality reduction information. PCA as accessible as possible the algorithm well cover is pretty technical. Some all of the following will make this article and PCA as a method easier to understand: matrix operation/ linear algebra (matrix multiplication, matrix transposition, matrix inverse, matrix inverses, matrix decomposition, eigenvectors / eigen values) .

When the dealing image the image have position sensitivity , the joint angle between hand and fingers their orientation all these come in Principal component analysis.

E. Linear Fingertip Model

The linear finger model are finger movement are assume the linear rotational movement. Finger tissue modeling requires linear deformation models The Fingertip model posses capability of linear movement of finger pad, suitable fingertip model which can respond for every type of hand and fingers because person to person size of hand, size of fingers , their joint angle and their movement could vary. Linear Fingertip model enforced to various type of deformation effects which are taken through the cameras during computer vision based tracking, like the video stream is captured with less control by which boundaries condition are hard to extract.

Gesture Recognition Steps

Step 1: First image is captured taken from stream using camera which is working by OpenCV.

Step 2: Now the image is processed that what data it is containing.

Step 3: On data algorithms are applied to find optimal gesture.

Step 4: Obtained gesture is now compared with gesture dataset.

Step 5: When if any dataset gesture is matched with input gesture then the result will displayed on screen or gesture is recognized.

VI. HAND DETECTION AND RECOGNITION MODELS

A. Hidden Markov Model

In Hidden Markov Model(HMM) (14), gestures are captures from every picture that makes a video and in this skin color blobs are tracked corresponding to hand-face likely on the face of user. Hidden Markov Model is a Markov Model which is made by using unobserved states. These unobserved states is also called Hidden states. In hidden Markov model ,the output is visible which dependent on the states while state is not visible directly.

B. YUV color space and camshift algorithm

How hand gestures are going to be recognized is dealt with by this algorithms. Following steps are included in order to recognition of hand gesture :

- 1) System Camera or digital camera first takes input of frames which are making video stream of hand motion gestures.
- 2) The existing fames into video stream that are going to be our input are grasped and then process of segmentation is performed and this is based on YUV color space.
- 3) The YUV color space system is used for distinguish intensity and chrominance of frames that are captured from video frames. In word YUV, Y specifies the intensity in video frame and UV indicates chrominance components in video frames.
- 4) After all this CAMSHIFT algorithm is used for hand to bifurcate it from body as hand and other body parts colors are same.



To segment the hand from other body parts we use logic that is the hand is largest connected region.

- 5) Hand position is calculated in each frame of video stream and to calculate its centroid of hand is calculated. To calculate centroid, from initial to last position is calculated.
- 6) To find the path of our input i.e. hand movement we have to join the all centroid points which form a trajectory and by using all these procedures we can track the hand movement.

C. Naive Bayes Classifier

Naive Bayes Classifier basically works on Bayes theorem which is collection of different classification algorithms. This is collection of different algorithms which is based on same logic that is every feature should be classified and these features should not be depend on each other.

The Bayes Theorem (18) used to find the probability of an event if any other given event already has been occurred.

$$P(E|F) = P(F|E)P(E) / P(F)$$

This method is used for recognition of static hand gestures. It is very efficient and quick method for recognition. As Naive Bayes collect different input according to their features so here different gestures are classified according to their different geometry and orientation. Every frame in the video sequence is used to extract gestures in which background is static.

VII. COLLECTING DATA FOR HAND GESTURES

We can collect input for this by mainly Three type.

First Way is that user can wear the device which take input. This device detects every angle making by hand and all the degree of freedom. This input device is type of glove which detects the position of hand in space and according to that raw data produced.

The Computer Vision Based technology is another way of collecting data in this camera takes images of hand gesture which are given by during input. These stream of images is sent to routine which is for image processing by which what input given by user by their gesture could be identify.

The hybrid way is last approach is used to take input as in this Vision and Glove Based both are used simultaneously which make raw data more accurate. Main reason using this hybrid way is when there is some error by one method the other will compensate by its correct data.

A. Instrumented Gloves

While designing these glove (3) there are some circuits and some sensors are fixed on glove which detects every single motion made by hand or by even fingers.

These instrumented gloves design and sensors used in it can be differ according to use or by companies made it. That's why it is not necessary every glove work same as other.

In this glove there are some sensor whose working is Light based. In light based a tube is used which is flexible so that it can cooperate with hand movement in placed at one side and on the other side photocell is used. As we move or make movement by finger the light that strike on photocell changed, by this finger flexion could be measured. Meta carpophalangeal of finger which is basically joints of all our

finger in one hand and our hand's thumb which have joint with our index finger is also detects by this sensor based glove



Fig. 2 Instrumented Glove

B. Computer Vision Based Approach

While using Instrumented gloved there are many difficulties faced by user like user have to always wear that glove while collecting the hand gestures. It is very difficult to give same posture same again and again because angle of hands and finger could be vary each time which make input gestures two different gestures.

On time of collecting the data the user is not allowed to make movement for some amount of time until input is not recorded correctly. So, Computer Vision Based (19) came into existence. To collect data by using Vision Based Approach for Hand gesture and movement there are four constituent which make system as whole:

Number of camera used and where to place them is first challenge in this approach. As we have to place cameras such that each and every movement of hand could be register. Main problem that encounter in this type of approach that hand movement is not visible by cameras and they also fail to record hand movement so this is primary concern in this method.

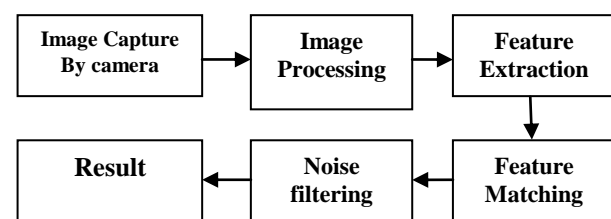


Fig. 3 Flow Diagram of Working of Computer Based Approach

The next (1)thing that constitute in this technology is extracting the simplest hand gesture data that have accurate visibility in camera.

Hand Gesture Recognition

Next constituent in this approach to captures every image from stream of images which are running. The last constituent to work this device efficiently some efficient algorithm should be applied in it which can take input, compare it with entries in database and then give result in minimum time. Here we are showing how a Vision Based Approach work(20).

C. Comparison Between both Models(2)

Methods	Glove Based Model	Vision-Based Model
Price	High	Less
Alleviation for User	Less	High
Hand Shape	To be same	Can differ
Calculation	Complex	Less complex than GB
Efficient	Lesser	High

VIII. CHALLENGES IN HAND GESTURE RECOGNITION

Hand gesture recognition confronts many challenges (20), these challenges are:

- A. **Changed illumination:** Change in light effects can affects our gesture input as it can change extracted skin region.
- B. **Rotation Problem:** Problem of Degree of freedom. if degree of freedom is changes then our gesture input may differ and by this output could vary
- C. **Distinguishing Problem:** If with hand there is other things which have shape and color like our skin but that is not our gesture input then it can create problem for system to distinguish between our hand gesture input and background.
- D. **Size Problem:** This As per human being , we have different shapes and sizes of hand like small child have small hands and adult have big hands so it can create problem for system.
- E. **Position Problem:** While giving input if hands position differs like hand placed in corner of screen or all dots which detecting hand position doesn't lies on hand then it can create problem to capture input from user.

IX. CONCLUSION

Hand Gesture Recognition is an important Human Computer interface for interaction between living thing like human and machine system. Hand Gesture Recognition System works like this : first user give input to the system by making hand gestures, then system scanned the gestures by using cam or sensor and deducts it into signal and passes the program, now its program responsibility to first accept the signal then examine what is the input given using gestures, then check if there is any corresponding data is saved into dataset then we will get our result.

System performance enhance when we train our system with maximum number of datasets. As we want to be our system more and more reliable for that we should train it with maximum number of datasets.

REFERENCES

1. Matthias Rehm, Nikolaus Bee, Elisabeth André. Wave Like an Egyptian – Accelerometer Based Gesture Recognition for Culture Specific Interactions. s.l. : British Computer Society, 2007. <https://ewic.bcs.org/content/ConWebDoc/21329>
2. Sanjay Meena. A Study on Hand Gesture Recognition Technique(Thesis Submitted in partial fulfillment of the requirements for the degree of Master of Technology in Telematics and Signal Processing)
3. Thomas G. Zimmerman, Jaron Lanier, Chuck Blanchard, Steve Bryson and Young Harvill. A HAND GESTURE INTERFACE DEVICE. (published in CHI '87 Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface Pages 189-19 <https://dl.acm.org/citation.cfm?id=275628>)
4. Toward a vision-based hand gesture interface. Quek, F. Singapore : s.n., August 23–26, 1994. Proceedings of the Virtual Reality System Technology Conference. pp. 17-29.
5. Hand Gesture Recognition for Human Computer Interaction. Mehra, Meenakshi Panwar and Pawan Singh. Wagnaghat, India : s.n., November 2011. IEEE International Conference on Image Information Processing (ICIIP 2011).
6. AnantAtray. Automated Hand Gesture recognition.
7. William T. Freeman, Michal Roth., Orientation Histograms for Hand Gesture Recognition. Cambridge : s.n.
8. Kouichi Murakami, Hitomi Taguchi. Gesture Recognition using Recurrent Neural Networks. Japan : s.n.
9. X. H. Wang, H. J. Bao., Recognition Based on Adaptive Genetic Algorithm. Journal of Computer-aided Design & Computer Graphics. 2007.
10. A framework for hand gesture recognition and spotting using sub-gesture modeling. M. R. Malgireddy, J. J. Corso, S. Setlur, V. Govindaraju, and D. Mandalapu. 20th International Conference on Pattern Recognition (ICPR '10), p. 3780.
11. R. Pradipa, Ms S. Kavitha. Hand Gesture Recognition- Analysis of Various Techniques and their algorithms. 2006.
12. Dynamic hand pose recognition using depth data. P. Suryanarayan, A. Subramanian, and D. Mandalapu. August 2010. 20th International Conference on Pattern Recognition (ICPR '10). pp. 3105–3108.
13. Real-time hand status recognition from RGB-D imagery. A. D. Bagdanov, A. Del Bimbo, L. Seidenari, and L. Usai. November 2012. 21st International Conference on Pattern Recognition (ICPR '12). pp. 2456–2459.
14. Vladimir I. Pavlovic, Rajeev Sharma, Thomas S. Huang. Visual Interpretation of Hand Gestures for Human-Computer Interaction;. s.l. : IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997.<https://pdfs.semanticscholar.org/8dd3/d5ef2e1c40433417cd90e21ce9a0468a7626.pdf>
15. Simon, Haykin. Neural Networks: A Comprehensive Foundation. s.l. : 2nd edition Prentice Hall PTR.
16. Virtual Reality. agentic.ca. [Online] <http://www.agentic.ca/virtual-reality>.
17. Stuart, A. and Ord, K. Kendall's Advanced Theory of Statistics: Volume I—Distribution Theory. s.l. : Edward Arnold, 1994.
18. Xu, Yanan and Dai, Yunhai. Review of Hand Gesture Recognition Study and Application. Contemporary Engineering Sciences. 2017, Vol. 10. Available: <https://doi.org/10.12988/ces.2017.7326>
19. Model-based hand tracking using a hierarchical Bayesian filter. Stenger B, Thayananthan A, Torr PH, Cipolla R. Sep 2006. IEEE Transactions on Pattern Analysis and Machine Intelligence. pp. 1372-84.
20. Shigeo, Abe. Support Vector Machines for Pattern Classification. Kobe University.