Learning-based Hand Gesture Sensing and Controlling Techniques for Automotive Electronics

Shang-Yi Cheng, Yu-An Wei, Min-Cheng Yen, Szu-Hao Huang*

Department of Industrial Engineering and Engineering Management, National Tsing Hua University,
No. 101, Section 2, Kuang-Fu Road, Hsinchu, Taiwan 30013, R.O.C.

*Corresponding Author: shhuang@ie.nthu.edu.tw

Abstract

The control interfaces of current automotive electronics are mostly designed in push-button or touch panel styles. In order to decrease the eyes-off-the-road time and increase the driving safety, a novel hand gesture sensing and controlling system is proposed based on various signal processing and statistical machine learning techniques in this paper. Replacing traditional high resolution camera or depth sensors, the proposed system uses a low-cost hardware design which is composed of proximity sensor chip, digital signal processor, microcontroller unit, and three programmable LED light sources. The learning-based method starts at the collection of large-scale gesture database with different labels under various environmental settings. Based on the dynamic time warping similarity measurements, the first gesture recognition method can achieve high accuracy with heavy computation. In addition, we also developed time-series feature extraction methods to transform the samples into low-dimensional space. Three different statistical learning approaches, including linear discriminant analysis, artificial neural networks, and support vector machines, can train the multi-class gesture classifiers in feature space. The demo system of automotive music player shows that the proposed system can efficiently and effectively recognize the pre-defined controlling gestures and improve driving safety dramatically.

Keywords: machine learning, gesture recognition, automotive electronics

1. Introduction

Human-computer interaction (HCI) is an interdisciplinary research field which focuses on the design

and implementation of the communication methods between human and machines. For current user interface of automotive electronics, the control panel of car audio, air conditioning, and navigation systems, may just fit the functional requirements. From the ergonomic perspective, a more intuitive control methods should be developed to decrease the operating time of automotive electronics and let the drivers focus on the road ahead. In this paper, a novel hand gesture controlling system is proposed based on various signal processing and statistical machine learning techniques. This new HCI design may decrease the eyes-off-the-road time and increase the driving safety dramatically.

Traditional researches of HCI focus on the design of user-friendly interface and comfortable user experience. With the rapid growth of novel information and communication technology, a variety of interdisciplinary applications and research issues had been developed for next generation HCI techniques. Along with the rapid development of audio signal processing and voice recognition techniques, a virtual assistant based on artificial intelligence technology is developed and integrated into the mobile devices, such as the famous Siri in iOS system. Kounoudes et al.(1) also proposed a user authentication procedure through analyzing users' specific voice.

HCI focuses on not only interaction between users and machines but human-centered design. Yang et al.(2) conducted face detection research to improve interaction between users and machines. Pantic et al.(3) surveyed more than 100 state-of-the-art facial expression analysis methods which are also important for further emotional design in human-centric HCI.

Currently, using touch panel is the most common method to control smartphones and tablet PCs. Several advanced operational functions, such as multi-touch and

customized gestures, are developed to improve the usability of touch panel devices. Moreover, users can control machines without touching but operating gestures in the air. Air-gesture sensing has been developed for a long time. Panwar et al.⁽⁴⁾ provided a real-time recognition system through analyzing basic gesture features including direction, finger patterns, and the position of each finger. Shimada et al.⁽⁵⁾ carried out a research about the relation among the rotational direction of hands. Panwar et al.⁽⁴⁾ used single camera to develop a hand gesture recognition algorithm with four major steps, which are named as segmentation, orientation detection, feature extraction and classification.

The RGB-D cameras and depth sensors are also widely applied in the gesture recognition problem. Ren et al.⁽⁶⁾ used Kinect sensor to capture the depth information of finger parts and developed a novel part-based hand gesture recognition system. They claimed that this system can work in cluttered background and lighting conditions. Ohn-Bar et al.⁽⁷⁾ developed a human-machine interface application of hand gestures in the car. The RGB-D data is collected for the pre-process in temporary segmentation module, and then be classified in gesture recognition module. Combining the researches of color images and depth information, Doliotis et al.⁽⁸⁾ applied dynamic time warping techniques to compare gesture recognition accuracy between these two different kinds input data. In addition, most early works had been surveyed in the research of Rautaray et al.⁽⁹⁾, which discussed about the techniques and applications of hand gestures in HCI.

In this paper, an air-gesture recognition system is proposed to intelligently analyze the time-series signals which are captured by multiple distance sensors. The design concept of air-gesture techniques is to create a more intuitive and convenient interactive methods for users to control the electrical devices, such as radio, mobile phones, table PCs, and other consumer electronics.

2. Proposed System Overview

This chapter describes an overview of the proposed gesture control system, including the sensing hardware and the decision system with dynamic time warping similarity measurements. The detailed definitions of eight control gestures are also recorded in the section 2.2. These gestures are used in the experiments and practical demo system.

2.1 Signal Acquisition Device

The proposed algorithm is developed based on

Fig. 1. Signal acquisition subsystem includes (a) Si1143 gesture sensor and (b) P8x32A QuickStart board

multiple time-series signals which are captured by the gesture sensing circuits. As shown in Fig. 1(b), the signal acquisition system is implemented in the P8X32A QuickStart board which is a simple and accessible development platform for the P8X32A Propeller microcontroller. In addition, Fig. 1(a) displays the Si1143 proximity sensor released by Parallax Propeller Inc. which is assembled onto the development board to capture the gesture signals.

Si1143 proximity sensor is composed of three LEDs light source, highly sensitive photodiodes to detect visible and infrared light, signal converter, and digital signal processor. The LEDs are soldered in three corners of sensor PCB and will be enabled sequentially during the signal acquisition process. As a low-power, reflectance-based, infrared proximity and ambient light sensor with I2C digital interface, Si1143 sensor can estimate the relative distance between the object and sensor chip.

P8x32A QuickStart board contains a multi-core parallel microprocessor, digital input/output pins, and temporary data storage space. Each program written in SPIN language can have 2KB instructions and data storage. In our signal acquisition subsystem, the distance measurements from the gesture sensor will be sequentially recorded in the temporary storage space in P8X32A QuickStart board. The computers will acquire these data through the serial port connection.

2.2 Gesture Design

After resampling the input data sequence into three independent signals by the light sources synchronization process, each training and testing gesture samples can be represented as three 1D time-series data. To define the target gestures is the first step of gesture recognition and control system design. These gestures should be designed in an ergonomic style which may let the user feel naturally and comfortably. In addition, the design has to avoid the ambiguous gesture definition which may lead to much lower classification accuracy. For example, it is very

difficult to correctly identify a complex gesture which is composed of another two simple target gestures. In this paper, eight different control gestures are defined as above, which include four fundamental sliding movements, three commonly-used symbol drawing, and a grabbing gesture in 3D space:

- (a) Upward Sliding (US): bottom-up palm movement.
- (b) Downward Sliding (DS): top-down palm movement.
- (c) Rightward Sliding (RS): sliding from left to right.
- (d) Leftward Sliding (LS): sliding from right to left.
- (e) Tick drawing (TD): draw a ticket in x-y plane.
- (f) Circle drawing (CiD): draw a circle in x-y plane.
- (g) Cross drawing (CoD): draw a cross in x-y plane.
- (h) 3D grabbing (3DG): Grabbing gesture from the sensor to higher z depth in 3D space.

Fig. 2 shows some example input signals captured by the proximity sensor after data synchronization and time-series signal pre-processes. In each sub figure, the x-axis is the time domain, y-axis represents the magnitude of signals, and the proximity signals captured with three different light sources are visualized by three lines in different colors.

Fig. 2. The example synchronized proximity signals of different gestures

2.3 Dynamic Time Warping

Dynamic Time Warping (DTW) is a temporal matching algorithm for measuring the similarity between two sequential data which may have speed variations. DTW algorithm had been widely applied in some audio and video processing systems. The gesture input are similar to multiple audio data, which are one-dimensional time-series signals. The similarity and distance between pairs of temporal signals cannot be calculated by simple Euclidean distance measurements because they may have different vector lengths.

The idea of DTW is to separate the temporal signals into multiple short units and to calculate the optimal unit matching method between two time-series sequences. One of the learning gesture samples in our database can be represented as a vector $\mathbf{t}(t_1^1, t_2^1, \dots, t_m^1, t_1^2, \dots, t_m^2)$, and the testing vector is $\mathbf{r}(r_1^1, \dots, r_n^1)$, where the m and n mean the numbers of signal units, respectively. The goal of DTW is to find a route $\{(p_1, q_1), (p_2, q_2), \dots, (p_k, q_k)\}$ with minimal distance measurement:

$$\text{route} = \arg \min_{(p_1, q_1), \dots, (p_k, q_k)} \sum_{i=1}^k \sum_{j=1}^3 |t_{p_i}^j - r_{q_i}^j|^2 \quad (1)$$

In addition, the estimated route should follow these constraints:

- $(p_1, q_1) = (1, 1)$, and $(p_k, q_k) = (m, n)$
- Assuming a point on the optimal path is represented by (i, j) , there are only three possible ways to reach this point from $(i-1, j)$, $(i, j-1)$, and $(i-1, j-1)$

$D(i, j)$ is defined as the DTW similarity measurement of the signals $\{t_1, t_2, \dots, t_i\}$ and $\{r_1, r_2, \dots, r_j\}$ with corresponding optimal route. To solve this problem, we can formulate the optimization equation as a recursive form to estimate the $D(m, n)$. The recursive relationship of $D(i, j)$ can be represented as the following form:

$$D(i, j) = \sum_{k=1}^3 |t_i^k - r_j^k|^2 + \min \begin{cases} D(i-1, j) \\ D(i-1, j-1) \\ D(i, j-1) \end{cases} \quad (2)$$

Another important issue is the time complexity, to calculate $D(m, n)$ needs $O(m \times n)$ computation. Fig. 3 uses the intensity to visualize unit-to-unit similarity, where the dark pixels mean the small between-signal difference. The DTW route calculated by recursive energy minimization process is displayed as the red line segments.

Fig. 3. The unit-to-unit similarity table and estimated DTW route

3. Learning-based Approach

In addition to the direct pattern matching method, the learning approaches sometimes can ignore the noises and achieve better performance. This first step of statistical learning system is to extract the features from raw signals. Then, three different kinds of learning algorithms are described to train the gesture recognition model.

3.1 Time-series Feature Extraction

In our research, two major feature representation methods are adopted. The first kind of features is to estimate the accumulative energy of each signal independently. The ratio in timeline will be recorded when each signal of three LEDs reached 25%、50% and 75% of total signal magnitude summation. From the fig. 2(c), we can observe that the red signal of RS motion will accumulate 25% earlier than the other two signals. It is because that the LED3 light source is located in the left side of sensor board and the left-to-right motion will enter the sensing area of LED3 first. Nine features can be extracted by this method.

The second feature extraction method is the binary comparison results among three LED signals at the specific time. There are also 9 features in this kind of feature.

The learning database includes 430 gesture learning samples which are captured from 20 subjects. The leave-one-out testing framework is adopted for accuracy experiments.

3.2 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a statistical analysis method that builds a better discriminant function with the least misclassifications. All the data will be classified into two categories in the feature space. The most popular methods are Fisher's linear discriminant, Bayes' s linear discriminant, and distance discriminant.

In our research, the training samples and the corresponding label vector are used as the input data for the LDA model construction. The distances between samples and population are estimated for the optimization process. Then, this model can predict the labels of the testing samples according to the rules learned from training dataset.

3.3 Artificial Neural Networks

Artificial neural networks (ANN) is a parallel computing system, consisting of individual artificial nodes, and imitates the neural network model of biologic systems. In artificial neural networks, simple artificial nodes, known as "neurons", are connected together to form a network which mimics a biological neural network. Each artificial neuron represents a kind of specific output function. All connections between two neurons are an adaptive weight activated during training and prediction. The output of an artificial neural network depends on the weights and output functions after convergence. Most of artificial neural networks are multilayer networks, including input layer, output layer, and hidden layers. The nodes of input layer take the responsibility to connect the time-series features of each learning sample. Output layer is the final decisions after transmitting, analyzing and weighing. Hidden layer consists of neurons and connections between input layer and output layer.

ANNs have been widely applied in various applications, such as pattern recognition, identification, classification, speech, computer vision, and control system. And in our application is gesture recognition. In this paper, the binary ANNs are extended to solve the multiple-class classification problem, such as the gesture recognition. Eight output nodes are adopted to connect eight binary decisions to determine whether this sample belongs to the specific targets or not.

The major process of artificial neural network learning is to adjust the weights of each layer by back propagation algorithm. The accuracy may be increased by repeatedly weight updating.

3.4 Support vector machines

In data classification and regression analysis applications, support vector machines (SVMs) is a supervised learning model with associated learning algorithms that analyze data and recognize patterns. Given a set of training data with specific labels, SVM algorithm builds a model that can assign new samples into the corresponding category.

Fig. 4. The hyper planes

Given a set of learning samples in high-dimensional space, SVMs algorithm constructs a hyper-plane which can separate data into two groups with the largest margin. Taking the examples in the fig. 4, the hyper plane with larger margin, as shown in fig. 4(b), may have more robust classification results than the other one.

Solid lines in fig. 4 are so-called support hyper-plane, which can be expressed as follows:

$$y_i(w^T x_i - b) - 1 \geq 0 \quad (3)$$

$$\text{margin} = 2d = 2/||w|| \quad (4)$$

w is the normal vector to the hyper-plane and b is a constant. To find optimal separating hyper-plane, we need to find two parallel support hyper-planes with largest margin first. d is the distance between separating hyper-plane and support hyper-plane, so the margin can be calculated as the equation (4). In addition, the largest margin can be estimated by the minimization system listed in the equation (5). In our research, we used C-support vector classification method and radial basis kernel function.

$$\begin{aligned} & \text{minimize } \frac{1}{2} ||w||^2 \\ & \text{subject to } y_i(w^T x_i - b) - 1 \geq 0 \end{aligned} \quad (5)$$

4. Experimental results

Various experimental results are listed in this section to prove the effectiveness of the proposed system. The most accurate but time-consuming DTW pattern matching approach is examined first. Then, the learning-based approaches with three different learning methods under different feature space settings are compared to find the optimal hand gesture recognition model. Finally, we found that we must adopt different strategy for eight different gestures.

4.1 Performance of DTW pattern matching

Table 1. The confusion matrix of DTW matching (%)

	US	DS	RS	LS	TD	CiD	CoD	3DG
US	99.3	0	2.7	1.6	0.3	0.3	0	0.4
DS	0	98.8	1.4	3	0	0	0	3.2
RS	0.7	0	94.7	0.3	0	0	0	1.2
LS	0	0	0	93.5	0.3	0	0	0.9
TD	0	0	0	0	97.9	1.2	4.2	0
CiD	0	0	0	0	0.3	97	0.3	0
CoD	0	0	0	0	0.5	0	94.2	0
3DG	0	0.4	0.9	0.9	0.4	0.6	0.9	94
Reject	0	0.8	0.3	0.7	0.3	0.9	0.4	1.2

We first collect a database with 3440 gesture signals as the matching pools. The leave-one-out testing framework is adopted to evaluate the system performance of DTW pattern matching method. In other words, each signal in the database will be predicted by the other 3439 signals with corresponding labels. Table 1 shows the confusion matrix of the prediction results by DTW pattern matching method. Each column in this matrix represents a group of testing gestures. Taking the example of US gestures, 99.3% are correctly recognized and only 0.7% are misclassified to RS gesture. In addition, reject decision means that the system judge the input signal does not belong to the eight target gestures.

All the accuracy of DTW algorithm is above 90%, and even half of them are higher than 95%. However, the relative high computation complexity may lead to longer response time. It is difficult to implement this kind of approach to the real-time applications. Then, the learning-based approaches are proposed to improve the system efficiency.

4.2 Performance of learning-based approaches

Three different learning-based approaches are examined in this section. Table 2 shows the system performance of these methods under different feature settings. From this table, we can find that the accuracy rate of dynamic time warping algorithm is almost higher than three learning-based methods in eight different gestures. In addition, we also examine the discrimination of different feature representation methods. In this table, F1 and F2 mean two different kinds of feature extraction methods which are described in the section 3.1.

Table 2. The system performance of three learning-based approaches under different feature settings

		US	DS	RS	LS	TD	CiD	CoD	3DG	Average
DTW		99.30%	98.84%	94.65%	93.49%	97.91%	96.98%	94.19%	93.95%	96.16%
LDA	F1	89.07%	95.12%	92.79%	91.86%	80.47%	84.88%	86.05%	84.88%	88.14%
	F2	76.28%	80.47%	94.19%	67.44%	77.21%	70.93%	86.98%	9.53%	70.38%
	F1+F2	85.12%	95.81%	93.72%	91.63%	79.77%	85.81%	88.60%	68.84%	86.16%
ANNs	F1	90.47%	96.98%	92.79%	96.28%	91.16%	94.42%	89.77%	90.93%	92.85%
	F2	88.84%	85.58%	90.23%	67.44%	75.81%	75.81%	90.00%	24.19%	74.74%
	F1+F2	94.88%	96.74%	92.09%	96.28%	94.88%	96.51%	93.26%	93.72%	94.80%
SVMs	F1	87.17%	94.57%	89.57%	87.61%	74.57%	63.70%	88.04%	85.84%	83.88%
	F2	86.30%	84.57%	90.87%	65.65%	75.43%	70.43%	87.39%	24.13%	73.09%
	F1+F2	85.65%	84.35%	92.17%	85.22%	76.96%	72.17%	87.39%	55.22%	79.89%

The experimental results of linear discriminant analysis (LDA) algorithm show that the accuracy rates of feature set F2 are lower than the accuracy of F1 in average. Even the richer feature representation method F1+F2 still cannot work better than F1. In these three kinds of learning-based approaches, ANNs with both kinds of features can achieve 94.80% recognition rate which is near the system performance of DTW matching.

To discuss about the feature representation methods, F2 cannot describe 3DG well because its accuracy with three learning method can only achieve 9.53% to 24.19%.

4.3 Hybrid Strategy

Table 2 also shows that we must adopt hybrid strategy to recognize different gestures. It is because that the feature representation methods and learning kernels may affect the recognition rate dramatically. From the table 2, the most accurate learning-based approach settings for different hand gesture recognition are chosen to form the final decisions. Table 3 shows the recognition rate with hybrid strategy.

Table 3. Gesture recognition rate with hybrid strategy

Gesture	
Upward Sliding	94.88 %
Downward Sliding	96.98 %
Rightward Sliding	94.19 %
Leftward Sliding	96.28 %
Tick drawing	94.88 %
Circle drawing	96.51 %
Cross drawing	93.26 %
3D grabbing	93.72 %
Average	95.09 %

5. Driving Safety Experiments

Finally, a demo system based on new HCI design of automotive electronics is developed to examine the improvement of driving safety.

5.1 Driving Experiments Overview

The demo system is designed to simulate the driving environments. There are 30 subjects with driver license join our driving safety experiments. In this experiment, the drivers are asked to control the music player by a sequence of events during driving. Three kinds of controlling system are examined, which includes the traditional push-button interface, 9.7 inches touch panel, and proposed air-gesture techniques.

In this experiment, the driving safety and the controlling fluency are compared to find the potential problems and to evaluate the improvement of proposed air-gesture techniques. The experimental environment is shown in fig. 5.

Fig. 5. The environment of driving safety experiments

Fig. 6. Experimental flow

5.2 Experimental flow

All the subjects will go through the same experiment flow in this experiment. They will be told about the object, flow of this experiment and events. They will also learn how to control the music player by three different interfaces before starting to drive. The experiments will be repeated three times for three devices. The experimental flow is shown in fig. 6.

Six different events will be announced during the subjects drive this car. These events include a music play requirement, volume increasing and decreasing, music selection under the demands of the passenger, system pause and resume with phone call event.

5.3 Numerical Evaluation of Driving Safety

Three indexes are measured in this human-computer interaction experiment. There are some indexes of the ergonomic experiment, including mission accomplishment time, eyes-off-the-road time, and average error operations. The mission accomplishment time recorded the time interval between the ends of vocal commands and event accomplishment. Eyes-off-the-road time is a widely-used index to measure the total time when subjects' eyes leave their frontal sight. Average error operations are the third index which counts the number of error trials.

Table 4. Mission accomplishment time

Task	Button	Touch Panel	Air-gesture
Play	3.708214	2.969643	3.616071
Volume up	4.439286	3.816071	4.581071
Music selection	7.973929	4.483214	11.215
Volume Down	3.973214	3.337143	4.776429

Phone call	15.61607	15.43286	16.27929
Resume	3.428929	3.334286	4.318571

Table 5. Eyes-off-the-road time

Task	Button	Touch Panel	Air-gesture
Play	1.793214	1.635714	0.715357
Volume up	2.161786	2.358214	0.7525
Music selection	2.653571	2.597857	1.4025
Volume Down	1.769643	2.133214	0.613571
Phone call	2.863929	2.8725	1.052857
Resume	1.927857	1.722143	0.761429

Table 6. Average error operations

Task	Button	Touch Panel	Air-gesture
Play	1.057143	1.228571	1.314286
Volume up	1.142857	1.085714	1.428571
Music selection	1.057143	1.2	2.042857
Volume Down	1.085714	1.2	1.371429
Phone call	1.028571	1.142857	1.885714
Resume	1	1.722143	0.761429

Table 4, 5, 6 shows the experimental results of mission accomplishment time, eyes-off-the-road time, and average number of error operations. The numbers of error operations are diverse according to the habits and dexterity of the users to control different devices. Observed from the table 4 and table 6, it is obvious that the air-gesture needs more trials to accomplish the missions. The main reason our research surmised is that the subjects may not get habit of waving the pre-defined gestures without touching any device. When the subjects want to finish the hand moving quickly and go back to the driving task as soon as possible, a short and dense gesture signals will be captured. It is very difficult to recognize this kind of gestures which is dissimilar to the learning dataset.

Table 5 shows that the proposed method can decrease the eyes-off-the-road time and increase the driving safety dramatically. Almost all these events can be accomplished with less than half eyes-off-the-road time.

6. Conclusions

In our research, Si1143 gesture sensor is used to capture three synchronized time-series signals. The pattern matching algorithm and machine learning techniques are adopted to develop an air-gesture controlling method for automotive electronics. The proposed method based on

dynamic time warping (DTW) can achieve most accurate recognition rate. But it also needs the longest response time because the high computational complexity. Then, two different kinds of feature extraction methods and three learning kernels are used to develop the learning-based approaches. These learning methods include Linear Discriminant Analysis (LDA), Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs). A hybrid strategy which uses different feature space and learning methods to recognize different gestures is used as the final strategy to recognize the hand gestures in an efficient and effective way. Furthermore, the gesture sensing HCI techniques are applied to the driving safety experiments. The proposed system can dramatically decrease the eyes-off-the-road time and increase the driving safety.

In the future, we believe this kind of air-gesture sensing technique can be extended to wearable device HCI design. It is because that the touch panels of some wearable device, such as the intelligent watches, are too small. Traditional user interfaces of smartphones and tablet PCs are not suitable for this kind of devices. In addition, it can also be used in some medical devices to avoid

Wearable devices combined with gesture sensor can simply operations and thus eliminating the inconvenience that screen is too small to control. In medical applications, this techniques can avoid the contagion.

Acknowledgment

The authors would like to thank the Ministry of Science and Technology of the Republic of China, Taiwan, for partially financially supporting this research under contract number MOST 103-2622-E-007-018-CC3 and MOST 103-2220-E-007 -020. This work was also supported by the Advanced Manufacturing and Service Management Research Center (AMSMRC), National Tsing Hua University.

References

- (1) Anastasis Kounoudes, Vassilis Kekatos, and Stephanos Mavromoustakos : "Voice Biometric Authentication for Enhancing Internet Service Security", *Information and Communication Technologies*, Vol. 1, pp. 1020-1025, 2006.
- (2) Ming-Hsuan Yang, David J. Kriegman, and Narendra Ahuja : "Detecting faces in images: A survey", *Pattern Analysis and Machine Intelligence*, IEEE Transactions on 24, No. 1, pp. 34-58, 2002
- (3) Maja Pantic, and Leon J.M. Rothkrantz : "Automatic analysis of facial expressions: The state of the art.", *Pattern Analysis and Machine Intelligence*, IEEE Transactions on 22, No. 12, pp. 1424-1445, 2000
- (4) Meenakshi Panwar, and Pawan Singh Mehra : "Hand Gesture Recognition for Human Computer Interaction", In *Image Information Processing (ICIIP)*, International Conference on IEEE, pp. 1-7, 2011
- (5) Atsushi Shimada, Takayoshi Yamashita and Rin-ichiro Taniguchi : "Hand gesture based TV control system—Towards both user-& machine-friendly gesture applications", In *Frontiers of Computer Vision,(FCV)*, 2013 19th Korea-Japan Joint Workshop on IEEE, pp. 121-126, 2013
- (6) Zhou Ren, Junsong Yuan, Jingjing Meng and Zhengyou Zhang : "Robust Part-Based Hand Gesture Recognition Using Kinect Sensor", *IEEE Transactions on multimedia*, Vol. 15, No. 5, pp.1110-1120, 2013
- (7) Eshed Ohn-Bar and Mohan Manubhai Trivedi : "Hand Gesture Recognition in Real Time for Automotive Interfaces: A Multimodal Vision-Based Approach and Evaluations", *IEEE Transactions on intelligent transportation systems*, Vol. 15, No. 6, pp.2368-2377, 2014
- (8) Paul Doliotis, Alexandra Stefan, Christopher McMurrough, David Eckhard and Vassilis Athitsos : "Comparing Gesture Recognition Accuracy Using Color and Depth Information", In *Proceedings of the 4th International Conference on Pervasive Technologies Related to Assistive Environments*, pp.20-27, 2011
- (9) Siddharth S. Rautaray and Anupam Agrawal : "Vision based hand gesture recognition for human computer interaction: a survey", *Artificial Intelligence Review*, pp.1-54, 2012