MAT JOURNALS

Journal of Data Mining and Management
e-ISSN: 2456-9437
Volume 03, Issue 02
4th National Conference on Advancements in Information Technology

Innovation of Touchless Touchscreen Technology in Automotive User Interface

Gagana H, Mrs. Neha Singhal, Kavyashree S
Department of Information Science and Engineering
Rajarajeswari College of Engineering
Bengaluru, India
Email: gaganah1996@gmail.com, neha.banasthali@yahoo.co.in, kavyashree306@gmail.com

Abstract

Inside a car environment music system plays a major role for entertaining people. In this paper, a music system with touchless and vision based GUI is granted. This menu-driven UI is offered by controlling actions of the fist. Both these algorithms can be brought by the deep learning technique. Convolutional Neural Network (CNN) is used in the hand posture recognition technique to make the user interface interactive to perform the actions initiated by the user. Long-Term Recurrent Convolutional Neural Network (LRCNN) algorithm is used to ignite the touchless interface by the gestures. When a fist movement is carried out, a sequence is captured in the form of multiple image frames. So this can be accomplished using the deep learning technique. Sampled images are taken from the video sequence that is captured during the gesture recognition. Key frame extraction technique is adopted to obtain finer images from the video sequence using sparse learning. Sparse dictionary learning is used as it is individually optimized for the video sequence but, is expensive computationally.

Keywords: *Feature extraction, Gesture recognition, Image sequences, Machine learning, Training, User interface*

INTRODUCTION

A few decades before, the mobile phones had buttons on it to exercise any of the operations. Later, the era of touchscreen phones emerged with the virtual keypad present on the screen and this eradicated the phones with manually pressing buttons. Touchscreen is a user interface which navigates the user to the different areas on the screen with the sensation of touch. This technology is used in every technological component. For example: Mobile phones, laptop, in car entertainment system. The society got adapted to the touchscreen technology sooner and its pleasure thrilled them. Later it was adopted in the mobiles phones and other graphical user interfaces. When people started using touchscreen phones their status in the society was upgraded. Such phones were also operated using the stylus stick. But, years after the invention people found many of the flaws in touchscreen technology like; there is a chance of different key being pressed on the virtual keyboard in the touchscreen interface, for example, music system of a car. The screen of touchscreen mobile phones might break when it is trashed as in case of mobile phones. In the presence of sunlight the screen cannot be seen, also as the brightness of the screen is raised the battery of the phones and other electronic gadgets gets deduced. Also sometimes touch surface may lose its sensation. Due to many such flaws in this technology a better invention came to light which is the touchless touchscreen technology.

Touchless Touchscreen technology is also known as Touch-me-not technology. Such system's structure can notify the movement of hand in front it and diagnose the kind of action performed. This technology can be controlled with the fist postures and gestures and uniquely, with the absence of touch. It also allows the user to turn any suitable surface into multi-touch interface. Touchless interface detects the hand postures and gestures presented in front of it and certainly perform the actions on the device. So, this helps the user in handling the interface easily.

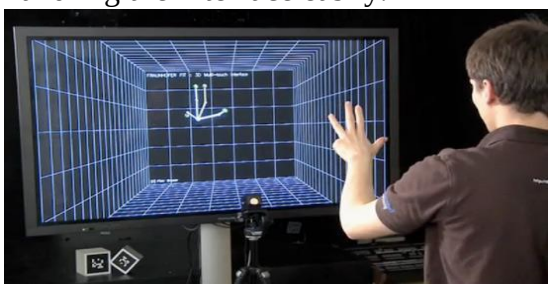

Fig.1: *Touchless control system that uses gestures to control user interface*

The touchless touchscreen is similar to the Nintendo Wii screen. Here, it does not contain the Wii controller. Certainly, this screen is assisted with the help of user's hand. This system is skilled of recognizing the movement of hands in 3-dimensions and especially, does not require placing of any fingers on such displays. Touchless interface doesn't need any peculiar sensors, so that it can be worn onto the fingers or on the hands. User has to make aim towards the UI (nearly from 5 feet away), so this can modify the objects in 3-D.

Optical pattern is recognized with a lens and the optical solid state matrix sensor to identify the motion of hands. Digital image processor is attached to the sensor to illuminate the motion patterns. Output results, as the signals to control component, appliances, machinery, or any other devices which are manageable through electrical signals are sent.

In automotive industry just like the car comprises music system whichever was prior designed to the touchscreen technology is now trending with the touchless touchscreen technology. As shown in figure 2, only through hand postures and gestures the music system can be operated. This may well be a propitious technique as this would not clutch the eye of the driver off the road.

Fig 2: *Images of touchless technology in automotive user interface (car)*

LITERATURE REVIEW

D. Wu proposes a semi-supervise ranked progressive framework. According to the Hidden Markov Model (HMM) that's recommended for a mutual gesture segregation and identification where skeletal club report, along with deep and RGB images, are the multimodal testimony observations. This method learns unusual spatiotemporal representations which use Neural Networks fitted to the input data technique: a Gaussian-Bernoulli Deep Belief Network (DBN) to perform and pay attention to the bony gesture, along with 3-D Convolutional Neural Network (3DCNN) to take care of and mix many such deep and RGB figures. This is achieved in the course of the modeling and information of the radioactivity chances of the Hidden Markov Model needed to ascertain the gesture array. Deep Neural Networks (DNNs) happen to be determined to outwit the cutting-edge in more than one area, similar to visual object recognition, and speech recognition [1]. Hand indication based gesture identification techniques work in the car user interface to amplify the user's delight by externally compromising their security. In a conventional LRCN based proceeding analyzer, more than one image is sampled from the film captured, secure as data to the structure, to carry out distribution. V. John proposed to cope with the flaws by obtaining lesser descriptive frames from the film, and inserting the system to the LRCN chain. Novel tiled figure patterns incorporate more than one distinct block to describe the broadcast arrangement of gestures in a one slated image. Image patterns suggest evidence to de-convolution structure along with arise of the broadcast array. Dual novel tiled patterns further consist of more than one non-overlapping block. These dual patterns show the results of one's de-convolution structure. The paired training patterns are derived from the priming filmed progressions together with the dictionary learning and sparse modeling groundwork. This approach to interaction again makes a bigger delight to the owner, without compromising the car owner's security [2]. An approach for action identification and localization in accordance with multi-ratio and multi-modal deep learning is utilized by N. Neverova. Each optical modality captures spatial instruction at this spatial scale, and the entire arrangement operates at two physical scales. This approach is usually a priming approach that exploits i) igniting of individual methods; and ii) progressive blending of procedures deriving out of most powerful to weakest cross-modality formation. In hand motion identification, a range of prominent image characteristics is well-known in way for segregating mannerism. Identification of gestures, starting with presence, spatiotemporal filming characteristics is used for segregating expression. These characteristics after which pre-owned inside a number of allotment frameworks for recognition of gesture labels. The Markov designs are primed upon skin colour primarily to analyze the gestures [3]. P. Molchanov employs Convolutional Deep Neural Networks to fuse data originating at more than one sensor and also to segregate signaling. Here, this algorithm distinguishes the images efficiently which are captured indoors and outdoors of the car during daylight and the darkness. It consumes lesser power than any other techniques. Hand positioning primarily used in graphical interfaces like in cars can lessen foresight and intellectual interruption, may also get better assurance and luxury [4]. J. Donahue provides a part of designs which has compliance impending interest a kind of vision strains. It provides the outline of image, activity and video obtained in vision task. He forms novel Recurrent Convolutional

composition advisable for large-extent optic learning that is end-to-end skilled, and testify to the cost of one's varieties on standard video recognition tests, image characterization and retrieval problems. In contradict to current models whatever affects the established spatio-temporal interested retrieve recurrent convolutional modes that are doubly deep so they might be compositional in contiguous [5].

EXISTING SYSTEM

Current interfaces don't get in contact with the touchless graphical user interface. So the car proprietor has to sensibly touch the buttons to administer or perhaps come in contact with the touchscreen arrangement within the car. This ends up in interference of one's car owner and feature limited concentration on driving. Further this can result in casualties.

Organizing multiple sensors for accurate and strength-coherent progressive car owner, fist movement identification having substitute tracking system, colour 35mm, and also a deep camera, that in combination perform the process physically powerful opposed to fluctuating light setting.

Fig 3: *Current interfaces in cars distract drivers from the concentration on road*

There is an operation to unitedly evaluate the tracking system and sensors. Researchers hire CDNN for integrating the inputs starting with more than one sensor and to sequence the gestures. It consumes moderately fewer management than essential vision-based arrangements. Also RGB-D image is just a sequence of a RGB image with depth images. So, depth image is definitely a figure channel in where every single pixel pertains to a size enclosed by the image plane and the comparable object inside the RGB drawing. A Kinect can be used to pick up such RGB-D drawings. If using Kinect-like metalware isn't handy under the user's charge, then it needs to predict the depth originating figures of the equivalent culture occupied starting with a couple of video pictures, coming right down to a mainframe vision complication. Another result is to bring together training picture consist of RGB-D figures and to use neural network techniques.

WORKING

The existing organization creates an effortlessly organized situation, where enterprise is continually monitored and usually distributed, definitely to the topics which are realized. The arrangement uses quadruplet calibrated cameras equipped within the suite which is being monitored and also a body-mounted mobile accelerometer on each individual, exploiting the characteristics of different

sensors to widen recognition certainty, get better scalability and reliability. The algorithm on which the system relies, in addition its network, are aimed towards analyzing and classifying convoluted movements (prefer walking, sitting, jumping, running, collapsing, etc.) of likely a couple of other people simultaneously. Here, we characterize a preliminary appeal, wherein claimed coordination is usually aimed toward locating slumps. Several instances of a combination classifier in response to Support Vector Machines and Hierarchical Temporal Memories, a latest bio-inspired computational model, are used to discover possibly fatal activities of every individual within the situation. If similar work is located and if the individual "at risk" is suited up with the accelerometer, the system localizes as well as activates it to gain information after which performs a further reliable ease discovery having in particular prepared classifier.

When the car owner makes hand gestures in front of the multi-sensor system, it is captured in different sensors like optical camera, time-of-flight depth, radar. This also includes the kinect. Then using Deep Neural Network the images get classified and the output is obtained as gesture classification. Finally, images are classified as behavior understanding and interaction images.

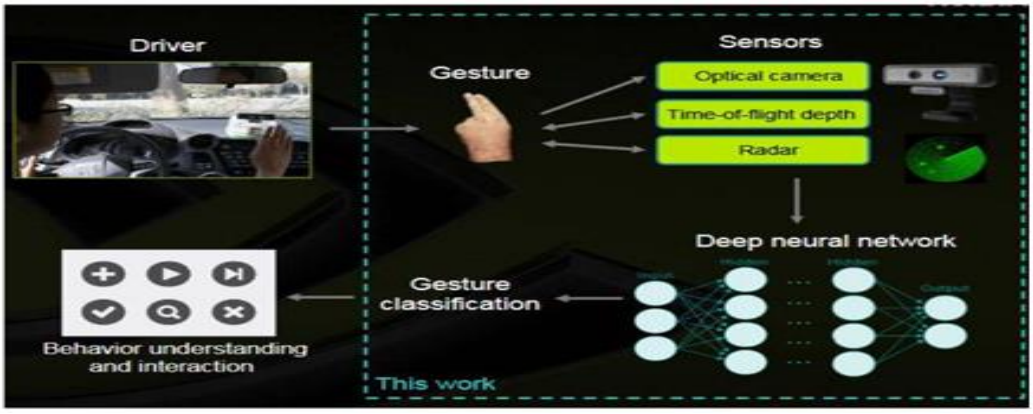

Fig 4: *Workflow of multi-sensor system*

Some of the advantages of the existing system include; can capture images during anytime of the day and also using Deep Neural Network images can be classified by recognizing the gestures. The flaws in the system led to flourishing of the proposed system; requires multiple sensors and the sensors may capture blur images when there is fast hand gestures made.

PROPOSED SYSTEM

Touchless touchscreen technology works as seen in the figure 5. This process senses the gestures made in front of the user interface as in our example, music system interface in car.

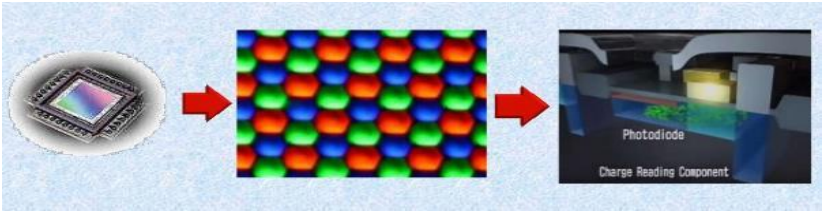

Fig 5: *Optical matrix sensor, matrix of pixels, photodiode*

Fist motions made facing the sensors are recorded by the cameras hence, these images are recognized. Then, light enters the optical matrix detector and hits its matrix. Signals are processed to provide output to the devices. Sensor generates the electric signals. Light sensitive diodes placed inside the sensor transmute light hitting the sensor into charge. Electrical waves are transformed to generate the desired results on a particular touchless screen. The optical matrix sensor is used where it senses the 3-dimensional gestures. Pixels matrix is present in each one of the sensors used. For engulfing charge storage regions light sensitive diodes are paired with the detectors. This is also called the charge reading component.

User makes gestures and hand postures which is captured in the form of video sequence by the sensor. A single recognizable image can be seen as shown in the figure 6. Single gesture or posture creates 16 samples in the database. Each image in the sample is different from each other. The sequence of arrangement of the images from the beginning till the end makes the sensor discover the kind of action made. Then, accordingly the action is labeled. As the type of gesture or posture is found further the action is implemented on the system. These sequences are seen in figure 7.For example: In the automotive system like car, phone calls can be attended through the music system interface. While driving if the driver gets a call the person should either stop the car and answer the phone call or ignore it, as it is risky to attend the call while driving. To overcome one such case, gesture recognition and hand posture recognition techniques can be implemented.

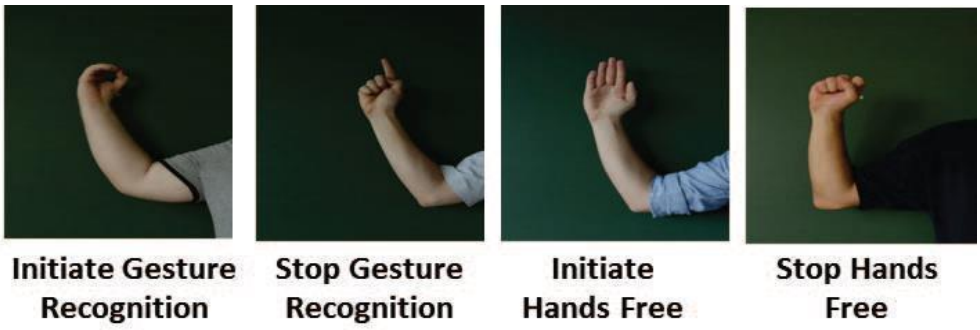

Fig 6: *Hand poses along with labels*

The static images that are captured using a camera are been sequenced. Each sequenced image is explained according to the gestures made. This initiates the touchless interactive interface. So, accordingly the interactive interface responds to the user. Analysis of an aggressive palm nod calls for processing a spatio-transitory impression arrangement. The exact range of your arrangement varies upon every single instantiation of one's body language. The solution for solving the issue is to convert the bounty of the individual gestural conversation law to a mechanical device to get a better man– mechanical device communication. The approach of recommending electricity is a recognizable assess that is prevalent proposition starting with a figure perpetuity. To, in attaining the specified ensue, we plan the notion of modeling the aggressive part expression with a limited speak mechanical device. The earthly identification is in consequence analyzed all fixed voice mechanical device to portray systematically the performed action.

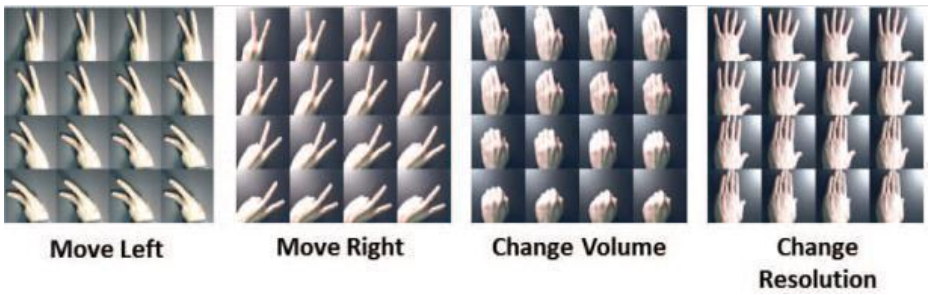

Fig 7: *Dynamic poses of the image along with labels*

The actions characterized are approximated by making use of the features from beginning of the frames in recurrent method. CNN also absorbs characteristics starting from the first till the last frame, and the Long Short-term Memory recognizes the hand motion. Here, recurrent is strengthened by the use of fewer input frames. Especially, the evaluated gesture tag by the use of three pointed out essential frames. Regression designs are indicated using these essential frames with the help of CNN. The semantic network is formulated in order to be told the dictionary learning. Given the extracted key frames, the CNN extracts the analogous theory features. The memory network due to this fact predicts the gesture tag.

Training Phase

The training phase is as shown in figure 8 includes the following four phases,
Fine-Tuning CNN Dataset Generation
Sparse Modelling approximation Fine-Tuning LRCN

Hand poses and pose label are stored in the pre-trained CNN. This takes place in the Fine-tuning CNN phase. The original pose sequence is sent for uniform sampling and the output of this stage is the sampled sequence. This sampled sequence images are further sent to the video feature extraction process. The null images are deleted from the video sequence. Only the required images are kept in the video feature stage is later sent to the sparse modeling representative frames. Key frames are obtained from the sparse coefficient. So, this creates a dataset needed for next phase. Video feature and sparse coefficient are integrated to form the neural network based SMRF (Sparse Modeling for finding Representative Feature) approximation. In the fine-tuning phase final key frames and the gesture labels are merged to form LRCN (Long-term Recurrent Convolutional Network).

Testing Phase

The testing phase is as shown in figure 9 includes the following three phases,
Hand Pose Estimation Key Frame Extraction
Hand Gesture Recognition

Sample image of the hand posture is finely tuned in the fine-tuned CNN so this process determines the hand posture estimation. In the key frame extraction process, original gesture sequence is uniformly sampled to obtain the sampled sequence. Sequence of images is obtained from the video feature extraction. A particular video feature is diverted to neural network based SMRF approximation to obtain the sparse coefficient factor. This results in the key frames. Hand gesture recognition process consists of the key frames that are sent to the LRCN so, from this process the gesture labels are obtained. A reverse process of training phase leads to the testing phase.

RESULTS & DISCUSSIONS

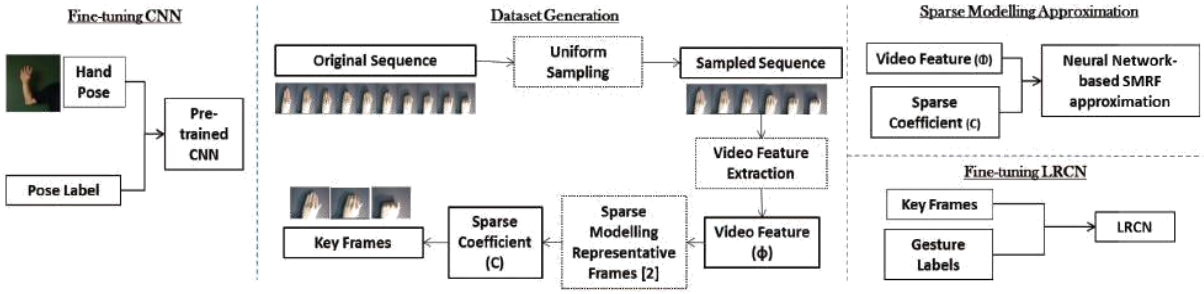

Fig 8: *The flow diagram of priming phase*

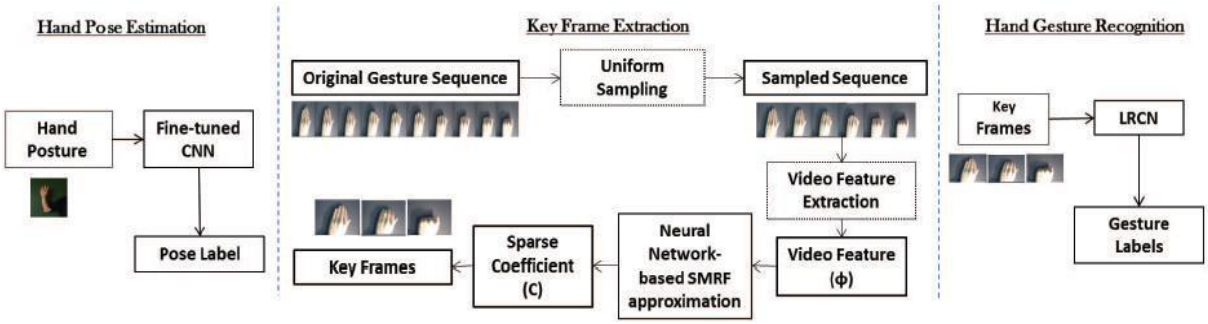

Fig 9: *The flow diagram of testing phase*

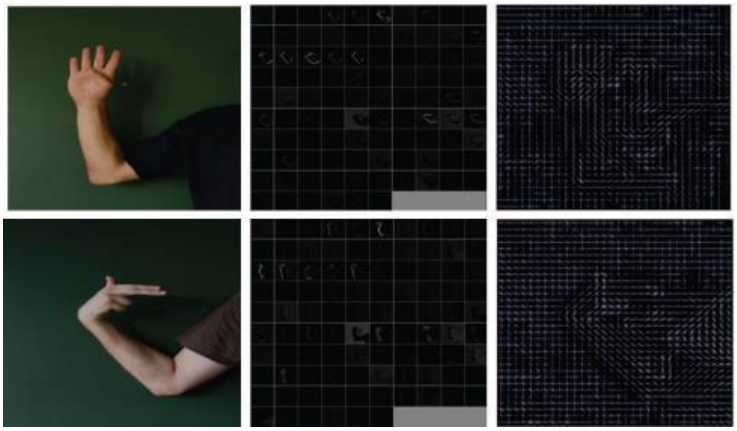

Fig 10: *Hand gesture in the form of Histogram of Oriented Gradient feature*

The image shown in the figure 10 indicates the stages of image segregation to identify the hand gesture. Captured video contains sequence of images of hand pose; the best image is taken from the sequence of image. To select the best image the hand posture and hand gesture algorithms are used. Then it is carried on to the HGR dataset where different postures of the images selected are put in the database. Finally, clear image from this dataset is obtained. The Histogram of Oriented Gradient feature is formed from the image selected in the HGR dataset. This gesture formation is found and it is labeled accordingly. Hence, the user interface responds back to the user by the gesture he/she made.

CONCLUSION

Hands mannerism and gesticulate distribution algorithm for touchless automotive graphical user interface is initiated in this paper. The signaling made is pointed out with the machine learning implementation. The data which means the

images captured by the sensor is indicated with the recurrent serious culture frame implementation. The input data to the recurrent framework is lesser making this algorithm an efficient one. Such frames are identified by the CNN and Sparse framework.

Sparse culture is computationally overpriced, that's addressed separately by approximation of neural network implementation and curvilinear implementation. We endorse the presented structure on communal datasets and we inform sharp regulation efficiency including problem solving and time computational expertise.

REFERENCES

1. D. Wu, L. Pigou, P. J. Kindermans, N. D. H. Le, L. Shao, J. Dambre, and J. M. Odobez, "*Deep dynamic neural networks for multimodal gesture segmentation and recognition*," IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 8, pp. 1583–1597, 2016.
2. V. John, A. Boyali, S. Mita, M. Imanishi, and N. Sanma, "*Deep-Learning based fast hand gesture recognition using representative frames*," in DICTA, 2016.
3. N. Neverova, C. Wolf, G. W. Taylor, and F. Nebout, "*Multi-scale deep learning for gesture detection and localization*," in ECCV 2014 Workshops, 2015.
4. P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "*Multi-sensor system for driver's hand-gesture recognition*," in FGR, 2015.
5. J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S.Venugopalan, K. Saenko, and T. Darrell, "*Long-term recurrent convolutional networks for visual recognition and description*," in CVPR, 2015.