Texas A&M University Qatar

ECEN 403

Benchmarking Report

Project Title:

Hand Gesture Controller

Instructor:

Dr. Khalid Qaraqe

Mentor:

Dr. Hussein Al-Nuweiri

Team 3:

Maha Al-Dehemi

Fatma Al-Mohannadi

Maryam Al-Safran

Lolwa Al-Kaabi

Due Date: 26/05/2024

Texas A&M University at Qatar, 2024

"An Aggie does not lie, cheat or steal or tolerate those who do."

Table of Contents

Introduction: 3

Relevant Standards and Constraints: 4

Existing Solutions: 5

- **1- Multivariate fuzzy decision tree (MFDT):** 5
- **2- Hidden Markov Model (HMM):** 5
- **3- 3D Skeleton-Based Hand Gesture (3D-BHG):** 6
- **4- Continual Gesture Activity (CGA):** 6

Benchmarking Criteria: 7

Benchmarking Table: 9

Benchmarking Study Analysis and Summary: 10

Conclusion: 11

References: 12

Introduction:

The challenges faced by individuals with conditions such as spinal cord injury (SCI) are indeed significant, impacting their independence and quality of life. It's heartening to see advancements in computational motion recognition offering potential solutions to assist these individuals in their daily lives. One of the key aspects of these advancements lies in the diversity of technologies being employed, from EOG sensors to camera-based eye tracking. Such technologies hold promise in enabling individuals with disabilities to interact with their environment more effectively, thereby enhancing their independence and reducing the need for constant external assistance. We focused on a group of target groups (MS, CP, elderly people, polio) However, it's crucial to acknowledge the limitations that these systems may currently face in catering to the diverse needs of individuals with physical disabilities. Tailoring these technologies to address specific challenges faced by individuals with conditions like SCI, including motor impairments and speech deficits, is essential for their widespread adoption and effectiveness.

Furthermore, integrating these computational motion recognition systems into the existing healthcare infrastructure and ensuring affordability and accessibility for patients and their families are vital considerations. Collaborations between researchers, healthcare professionals, and technology developers can help streamline the development and deployment of these solutions, ultimately improving the lives of those affected by SCI and similar conditions.

Relevant Standards and Constraints:

One of the major design constraints is the training time required to learn Python programming language to achieve the highest levels of optimal use of the programming language. It is difficult to predict the duration of the training period due to the complex interactive nature of the processes involved in designing such as SolidWork design programs to create the project illustration[1]. One of the reasons that prevents us from predicting the duration is the amount of information required for training. It is not possible to fully understand the quantity, type, or variety of information required [6]. Another reason is the lack of similar real-life applications of machine learning for programming and identifying the hand that appears in the camera and not dealing with any other inputs [6]. Since such a project does not exist in the market and if it does exist, it is very expensive to purchase, the criteria that our design and the program that we will write may not exist. However, we are unable to fit a specific framework of criteria at this early stage of design. The project we are working on could bring great benefits and progress in the field of medicine and physical therapy [2].

Existing Solutions:

The researchers in this paper defined four main categories of detection techniques that are used widely in Gesture Guide which are:

1- Multivariate fuzzy decision tree (MFDT):

Propose MFDT to learn and classify hand gestures based on decision tree learning method. In fuzzy decision tree method of cataloguing hand gestures, count of nodes is very large as the data is split using fuzzy membership; having large number of nodes causes lower performance rates. MFDT has smaller number of nodes as compared to fuzzy decision trees. In decision tree learning method, discrete valued target functions are approximated and learned functions are symbolized as a decision tree. ID3 and C4.5 are the well-known decision tree algorithms. As compared to fuzzy decision tree, in MFDT, decision tree is formed using multivariate concept, but fuzzy decision tree uses unilabiate split concept[2].

2- Hidden Markov Model (HMM):

HMM is one of the most widely accepted tool for time-series analysis as this method is capable of exhibiting temporal relationships among different models and samples in addition to segmentation and classification. This capability is extensively used for training and interference by a different variation of HMM. Also it is considered as one of the most popular approaches for dynamic hand gesture recognition. In HMM, hidden factors ensure different states and variants of these state transitions are visible [4]. The output of HMM also consists of hidden information of state sequence, and though each state is not noticeable, potential output has possibility distribution [4].

3- 3D Skeleton-Based Hand Gesture (3D-BHG):

3D skeleton-sequences are being increasingly used as inputs for action and gesture recognition tasks due to their robustness to background interference, illumination and viewpoint changes as well as reduced training complexity as compared to RGB-D inputs. Several deep learning based methods have been used to model skeleton-based hand gesture sequences. Authors in propose a two-stage CNN and LSTM framework to learn spatial and temporal joint features respectively [7]. More recent approaches for action recognition employ Graph Convolutional Networks to model a spatio-temporal graph of the skeleton sequence [8]. STST and DSTA-Net design specialized transformer blocks to learn spatial and temporal features in a decoupled man ner. DG-STA is a fully connected graph transformer. It applies multi-head spatial attention over the spatial skeletal graph, followed by multi-head temporal attention on the graph's temporal edges [7]. We use DG-STA as our architectural backbone due to its simplicity of construction, feasibility of model inversion and code availability.

4- Continual Gesture Activity (CGA):

Recently, there has been an active interest in making real-world gesture and activity-based human-robot interaction systems continually learn new user classes. However, most existing works focus on sensory data from accelerometers, ambient sensors, or surface electromyographic (sEMG) signals. Authors in propose a lifelong adaptive learning framework that processes motion sensor-based HAR datasets in a task-free continual fashion using experience replay and continual prototype adaptation. More recently, proposes an exemplar memory enhancement strategy for class-incremental learning (CIL) of static, single-image gestures such as in NUS II and Sign

Language MNIST [3]. CIL has also been explored for action recognition in videos. Both these works address video continual learning using regularization and episodic memory replay based methods. Authors in CatNet are the first to attempt class-incremental hand gesture recognition. They use the EgoGesture dataset and propose a two-stream RGB and depth framework, which replays previous class exemplars based on the iCARL algorithm [4]. Our work differs from these works in two key aspects. Firstly, incremental learning has not yet been explored for skeleton-based dynamic hand gesture recognition. Secondly, unlike these methods which majorly rely on replay of stored exemplars from previous classes to mitigate forgetting, we circumvent user privacy, data security, and scalability concerns by proposing a novel data-free class-incremental framework.

Benchmarking Criteria:

Our design is expected to meet specific needs with specific criteria in mind, including public health, public safety, welfare, global, cultural, and social, and more. We believe that some of these criteria do not apply exactly to our design, such as medical criteria. Although not obvious, most of the criteria mentioned above may be indirectly relevant to our design. The purpose of our project is to help people with mobility disabilities make their daily lives significantly easier. By using scientific and medical networks to try to optimize our project, we anticipate a faster detection rate at the end of the future, and thus a more stable, reliable, and cost-effective communication system. Thanks to the project having a stable and reliable communication system, the person with mobility disabilities will be able to complete all their basic tasks smoothly, etc., faster in an emergency. In a way, we take into account public health, safety, welfare, and global needs, but indirectly.

Among the specific needs criteria, only environmental and economic needs apply to the proposed design. Therefore, we added two more criteria, performance and feasibility, which we felt were relevant to our project design and existing solutions.

For environmental needs, the use of the project will be a breakthrough in the future, instead of the existing very expensive solutions, meaning that the system will facilitate the lives of people with disabilities and make their lives less complicated, and it is worth noting that the solution we propose is based on software and some simple devices. There will be no need to implement and manufacture new expensive and complex equipment as the device can be handled simply and efficiently, since we do not have to manufacture new equipment, we save energy.

For economic needs, as we mentioned earlier, our proposed design will not require the manufacture of new equipment, and it can be implemented with the least capabilities available in the local market such as shopping sites such as Amazon and AliExpress, and therefore we assume that the implementation should be at a relatively low cost. In fact, this proposed solution can be considered economically beneficial from the manufacturer's side because it requires a low production cost, but it is profitable due to the expected performance. Also, very low hardware maintenance will be required. Once the project is made and tested through supervised learning, it can work without supervision and operate on its own. There will not be much software maintenance required either. The only economic downside is that the device would be purchased by the consumer, as it would be more expensive due to the higher performance expected from it.

In terms of performance, we expect it to provide better performance in terms of camera quality, capacity, throughput, hand-to-noise ratio, and hand recognition error rate from the camera. We aspire to make a more reliable system with higher capacity adjustments at the transmitting and receiving ends of the handicapped or user, all while maintaining a relatively low error rate or noise to signal-to-receive ratio. Obviously, current projects are already available in the market to realize the possibility of achieving this. However, for projects that work with the help of artificial intelligence, it is still in research as training and testing the neural network can take a long time.

Benchmarking Table:

		Criteria			
		Environmental	**Economical**	**Performance**	**Attainability**
Existing Solutions	**Our Project**	Fewer procedural operations after training and testing, Lower power consumption and excellent user assistant.	Easy to use, technically inexpensive and content based project.	High Performance more than MFDT and HMM	High
	MFDT	Low impact	Moderate Considerations	Moderate performance	High
	HMM	Fewer uses than MFDT, lower power.	It needs training from the user first and then dealing with it	Similar performance of MFDT	Moderate
	3D-BHG	Low impact	High Considerations	High performance	Moderate
	CGA	Low impact	High Considerations	High performance	Moderate

Table 1 Benchmarking criteria comparison

Benchmarking Study Analysis and Summary:

After considering environmental, economic, performance and Attainability factors, we compared our solution with existing market solutions. We assessed the environmental impacts based on computational complexity, reasoning that higher energy consumption and increased environmental impact result from more project processes.

- Environmental Impact: All techniques have a relatively low environmental impact since they primarily involve computational processes rather than physical resources. However, the choice of hardware used for implementation can influence energy consumption.
- Economic Considerations: HMM, 3D-BHG, and CGA tend to have higher economic considerations due to the need for sophisticated hardware and potentially expensive training processes. MFDT may be more economical due to its simpler algorithmic approach.
- Performance: HMM, 3D-BHG, and CGA generally offer high performance in gesture recognition tasks due to their ability to capture temporal dynamics and spatial features effectively. MFDT may have moderate performance depending on the complexity of the decision tree and feature representation.
- Attainability: MFDT stands out for its high attainability since decision tree algorithms are well-established and easily implementable. HMM, 3D-BHG, and CGA may require more specialized knowledge and resources for implementation, making them moderately attainable.

Conclusion:

Upon thorough examination, we closely analyzed the design and constraints of our project in relation to other solutions and products available on the market. We meticulously sifted through a wide array of criteria to determine their applicability to our project. Additionally, we delved into researching criteria that were more pertinent to our solution. We then identified various existing solutions, ultimately selecting four for comparison across different criteria to gauge the standing of our solution in the landscape of existing alternatives. From the insights gleaned in this report, it became evident that our focus should be on enhancing the feasibility and performance of our design.

References:

1- Salonee Powar, Shweta Kadam, Sonali Malage, Priyanka Shingane "Automated Digital Presentation Control using Hand Gesture Technique", ITM Web Conf. 44 03031 (2022)DOI: 10.1051/itmconf/20224403031.

2- Damdoo, Rina & Kalyani, Kanak & Sanghavi, Jignyasa. (2020). "Adaptive Hand Gesture Recognition System Using Machine Learning Approach". Bioscience Biotechnology Research Communications. 13. 106-110

3- Shangchen Han, Beibei Liu, Randi Cabezas, Christopher D Twigg, Peizhao Zhang, Jeff Petkau, Tsz-Ho Yu, ChunJung Tai, Muzaffer Akbay, Zheng Wang, et al. Megatrack: monochrome egocentric articulated hand-tracking for virtual reality. ACM ToG, 2020.

4- Shangchen Han, Po-chen Wu, Yubo Zhang, Beibei Liu, Linguang Zhang, Zheng Wang, Weiguang Si, Peizhao Zhang, Yujun Cai, Tomas Hodan, et al. Umetrack: Unified multiview end-to-end hand tracking for vr. In SIGGRAPH Asia, 2022

5- Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D Bagdanov, and Joost van de Weijer. Classincremental learning: survey and performance evaluation on image classification. IEEE TPAMI, 2022.

6- Zheda Mai, Ruiwen Li, Jihwan Jeong, David Quispe, Hyunwoo Kim, and Scott Sanner. Online continual learning in image classification: An empirical survey. Neurocomputing, 2022.

7- Yuhan Zhang, Bo Wu, Wen Li, Lixin Duan, and Chuang Gan. Stst: Spatial-temporal specialized transformer for skeletonbased action recognition. In ACM Multimedia, 2021

8- Zhengwei Wang, Qi She, Tejo Chalasani, and Aljosa Smolic. Catnet: Class incremental 3d convnets for lifelong egocentric gesture recognition. In CVPRW, 2020.