Automata-Based Pattern Recognition for Gesture Recognition Applications

A Thesis Project

Presented to the Faculty of the

Department of Computer, Information Sciences and Mathematics

University of San Carlos

In Partial Fulfillment

of the Requirements for the

CIS 3202N – Automata Theory and Formal Languages

Members:

Rosales, Jade Andrie

Samson, Jose Glen

Stewart, Christian Anthony

April 2023

Problem Statement

Gesture recognition has become a rapidly growing field in recent years, with numerous applications across a variety of fields, including video games, virtual reality, medical devices, and rehabilitation technology. The ability to accurately recognize and interpret hand and body gestures is crucial for providing seamless and intuitive interactions between humans and machines.

However, one of the main challenges in gesture recognition is developing algorithms that are both accurate and robust. Existing methods for pattern recognition, such as machine learning and deep neural networks, have limitations when it comes to distinguishing between different gestures and interpreting their meanings. This is due in part to the complex and often ambiguous nature of human gestures, which can vary significantly depending on factors such as culture, context, and individual differences.

To address these challenges, we propose an automata-based approach for pattern recognition in gesture recognition applications. This approach leverages the mathematical theory of automata to model and recognize patterns in a more flexible and interpretable manner. By using automata-based algorithms, we aim to improve the accuracy and robustness of gesture recognition systems while also providing meaningful interpretations of the gestures detected.

Overall, this thesis aims to contribute to the development of more effective and reliable gesture recognition systems by exploring new approaches to pattern recognition. By leveraging the power of automata-based algorithms, we believe that we can improve the performance and usability of gesture recognition technology, enabling it to be used more widely in real-world applications.

Review of Related Literature

Using Finite State Machines (FSM) for Gesture Recognition

In the paper of Hong, P., Turk, M., & Huang, T. S. (2000), it presents an efficient approach for modeling and recognizing hand and mouse gestures using Finite State Machines (FSM). The proposed method allows for a semi-automatic way of constructing gesture models and does not require large datasets for training. The algorithm first decouples spatial and temporal information of the data, learns spatial information for data segmentation and alignment, and then learns temporal information from the aligned data segments. The FSM recognition procedure is incorporated with the Knuth-Morris-Pratt (KMP) algorithm, a fast string-matching algorithm, to achieve fast recognition speed. The approach was tested on hand and mouse gestures, achieving recognition rates of 90% or better for hand gestures and 70-100% for mouse gestures. The results demonstrate the potential for this approach to be applied to a large vocabulary. Overall, the paper presents a promising approach for gesture recognition with potential applications in various fields.

Using Fuzzy Logic for Gesture Recognition

A fuzzy rule-based method for the recognition of hand gestures was introduced in the paper of Bedregal, B. C., Costa, A. C., & Dimuro, G. P. (2006). The method uses the set of angles of finger joints for the classification of hand configurations, and classifications of segments of hand gestures for recognizing gestures. There are 27 possible finger configurations that are considered. Each of these configurations are codified into variables to indicate that the three joints are either straight or curved. The hand configuration can also be determined by considering each finger configuration. Then finally using fuzzy logic to determine the segmentation of the gesture in monotonic hand segments. The set of all lists of segments of a given set of gestures determine a set of finite automata, which are able to recognize every such gesture. The paper shows promising results and is suitable for the application to the recognition of hand gestures of sign languages.

Another paper that follows this method is the paper of Verma, R., & Dev, A. (2009). The paper proposes an approach for hand gesture learning and recognition that combines finite state machines and fuzzy logic. The researchers use edge detection and vector extraction to identify the location of the user's hands in 2D image positions. The information of spatial and temporal domain is first separated. The data is grouped into clusters based on temporal alignment using Fuzzy c-mean clustering which allows one piece of data to belong to two or more clusters. The clusters of hand postures are used to determine the states of the finite state machine(s), which are used to match the succeeding gesture. The number of states/clusters represents a trade-off between the accuracy of recognizing the gesture and the amount of special/temporal data. The technique is demonstrated to be successful for a set of gestures such as waving left/right hand, signaling to stop, forward, rewind, etc. The approach has potential for further development and applications in various fields.

Using Hidden Markov Model (HMM) for Gesture Recognition

A study by P. Hong, M. Turk, and T. S. Huang (2000) proposes a novel method of gesture recognition using finite state machines. The study focuses on recognizing hand gestures in real-time by modeling them as FSMs. The researchers propose a hierarchical FSM architecture that can recognize complex gestures by combining basic gesture FSMs. The study also introduces a novel method for modeling gesture dynamics using hidden Markov models (HMMs). To evaluate the proposed method, the researchers conducted experiments on two gesture datasets, one containing basic hand gestures and the other containing more complex gestures. The results showed that the proposed method achieved high recognition accuracy for both datasets. Overall, the study presents a promising approach to gesture recognition using FSMs and HMMs. The proposed method has potential applications in various fields, including human-computer interaction, virtual reality, and robotics.

The paper of Elmezain, M., et al. (2008, December) proposes an automatic system for real-time recognition of isolated and continuous hand gestures for Arabic numbers 0-9 using HMM. The system applies HMM using different topologies and numbers of states ranging from 3 to 10 for isolated gestures, and zero-codeword detection with static velocity motion for continuous gestures. The proposed system uses orientation dynamic features obtained from spatio-temporal trajectories and codewords. The system achieves an average recognition rate of 98.94% and 95.7% for isolated and continuous gestures, respectively, with the LeftRight Banded topology with 5 states presenting the best performance. The researchers suggest future work on motion trajectory carried out by fingertip instead of hand centroid point using a multi-camera system over combined features. Overall, the proposed system shows promising results for real-time recognition of hand gestures for Arabic numbers.

Using Convolutional Neural Network (CNN) for Gesture Recognition

In the paper of Lin, H. I., Hsu, M. H., & Chen, W. K., (2014), it proposes a convolutional neural network (CNN) approach to hand gesture recognition. The paper describes a system that is capable of detecting and recognizing hand gestures in real-time using a camera-based input. The system consists of three main components: image acquisition, feature extraction, and classification. The image acquisition component captures hand images using a webcam, and the feature extraction component extracts features from the images using CNNs. They also adopted a Gaussian Mixture model (GMM) to robustly filter out non-skin colors of an image. The model was trained to learn seven gesture types and got results that showed that the average recognition rates were around 99%. However the current system still needs enhancement to the recognition capability for complex human tasks.

A method based on multimodal data fusion and multiscale parallel CNN is proposed in the paper of Gao, Q., Liu, J., & Ju, Z., (2021). The data fusion is conducted on the sEMG signal, the RGB image, and the depth image of hand gestures which are then fused to generate two different scale images by downsampling. The data is then inputted into two subnetworks of the parallel CNN to obtain two hand gesture recognition results. The results are then combined to obtain the final hand gesture recognition result. Ten commonly used HRI (Human-Robot Interaction) hand gestures are designed. As a result, the system's accuracy for the 10 HRI hand gestures is greater than or equal to 78%. The system's limitations include its inability to recognize dynamic hand gestures. Dynamic hand gestures are more practical and difficult than static hand motions hence why the system is less accurate in this field.

The paper of Islam, M. Z., et al. (2019), proposes a static hand gesture recognition method using CNN. The researcher applied data augmentation techniques to the dataset and trained the model on 8000 images, achieving an accuracy of 97.12% on the test set of 1600 images across 10 classes. The results show that data augmentation has a significant impact on the accuracy of the model. The researchers suggest some potential areas of future work, including the use of knowledge-driven methodologies to improve accuracy and the recognition of gestures made with both hands. These suggestions are useful and provide a direction for further research in this area. Overall, the paper demonstrates the effectiveness of CNNs and data augmentation techniques.

Using 3D Convolutional Neural Network (3DCNN) and Long-Short Term Memory (LSTM) with Finite State Machine (FSM) for Gesture Recognition

Hakim et al. (2019) proposed a dynamic hand gesture recognition model that combines 3D convolutional neural network (CNN) and long short-term memory (LSTM) with a finite state machine (FSM) context-aware model. The model aims to improve the accuracy and robustness of hand gesture recognition by incorporating temporal information and context-awareness. The 3D CNN is used to extract spatial and temporal features from the input gesture sequences, while the LSTM is used to model the temporal dynamics of the gestures. The FSM context-aware model is used to model the transitions between gestures and incorporate contextual information such as the user's hand position and orientation.

To extract the spatio-temporal information, a three-dimensional convolutional neural network (3DCNN) and a long short-term memory (LSTM) model were combined. Finite State Machine (FSM) communicates the model to regulate the class determination outcomes based on application context at the conclusion of the classification. The results showed that, for eight selected motions, the combination of depth and RGB data had a 97.8% accuracy rate, while the FSM increased the identification rate in real-time from 89% to 91%.

The authors first collected a dataset of dynamic hand gestures performed by multiple users in various lighting conditions and viewpoints, using a Kinect sensor. They then preprocessed the data by converting it to 3D voxel grids and normalizing the hand size and position. The preprocessed data was then fed into a 3DCNN to extract spatiotemporal features from the hand movements. The output of the 3DCNN was then fed into an LSTM network to capture the temporal dependencies and memory-based modeling of the gestures. Finally, the output of the LSTM was fed into an FSM context-aware model, which used a set of rules to constrain the possible transitions between gestures and ensure the consistency and coherence of the recognized gestures.

The authors evaluated the performance of their approach using several metrics, including accuracy, precision, recall, and F1 score. They compared their approach with several state-of-the-art methods, such as HMMs, SVMs, and CNNs, and showed that their approach achieved superior results in terms of accuracy and robustness. They also conducted several experiments to analyze the impact of various factors, such as the number of hidden units in the LSTM, the size of the FSM state space, and the effects of noise and occlusion in the input data. The results showed that their approach was able to handle various challenges and achieve high performance under different conditions.

The approach proposed by Hakim et al. (2019) involves the use of several components that can be viewed as automata or models. The 3D convolutional neural network (3DCNN) can be viewed as a type of automata that recognizes patterns in the spatial and temporal domain. The long short-term memory (LSTM) network can be seen as a type of automata that captures the temporal dependencies and memory-based modeling of the gestures. The finite state machine (FSM) context-aware model can be seen as a type of automata that enforces constraints on the possible transitions between gestures and ensures the consistency and coherence of the recognized gestures. Therefore, the proposed approach involves the integration of multiple automata or models to achieve higher accuracy and robustness.

Moreover, the study by Hakim et al. (2019) highlights some of the challenges and limitations of existing methods, such as the reliance on 2D images or depth maps and the inability to capture the temporal dynamics and contextual information of the gestures. These challenges are also relevant to automata theory, as the ability to recognize or generate languages depends on the ability to capture the spatial and temporal patterns and the contextual constraints of the language. Therefore, the proposed approach can be seen as a novel way to overcome some of these challenges and limitations and advance the field of automata theory.

Utilizing TensorFlow Framework for Gesture Recognition

The research study by Zeng et al. (2018) on the design of a CNN model for gesture recognition based on the TensorFlow framework is an important contribution to the field of computer vision and machine learning. The study presents a novel approach to recognizing hand gestures that can be used in various real-world applications, such as human-computer interaction, robotics, and virtual reality. The authors propose a CNN-based model that achieves high accuracy and robustness in recognizing a variety of hand gestures.

The study by Zeng et al. (2018) demonstrates the effectiveness of using a CNN-based model for gesture recognition compared to other traditional machine learning methods, such as decision trees and SVMs. The authors conducted experiments on a large dataset of hand gesture images and showed that their CNN-based model outperformed the other methods in terms of accuracy and speed. Moreover, the study provides a detailed analysis of the CNN model's architecture and parameters and highlights the importance of optimizing these parameters for achieving the best performance.

The findings of the study by Zeng et al. (2018) have important implications for the development of real-world applications that involve gesture recognition. The proposed CNN-based model can be used in various scenarios, such as sign language recognition, game control, and robot navigation. The high accuracy and robustness of the model can improve the overall user experience and enable more intuitive and natural interaction between humans and machines. The authors also recommend further research in exploring the use of deep learning techniques, such as recurrent neural networks and attention mechanisms, to enhance the performance of the model and enable more complex and subtle gesture recognition.

Overall, the research study by Zeng et al. (2018) presents a novel approach to gesture recognition based on a CNN model and provides valuable insights into the design and optimization of such models. The study demonstrates the effectiveness of using CNNs for recognizing hand gestures and highlights the importance of optimizing the model's architecture and parameters. The findings and recommendations of the study can inform the development of real-world applications that involve gesture recognition and stimulate further research in the field of computer vision and machine learning.

Methodology

Our approach begins with designing deterministic finite automata (DFAs) to model the sequences of gestures and recognize them based on their input signals. DFAs are a type of mathematical model that can be used to recognize patterns in a sequence of inputs, making them well-suited for modeling gestures. We design these DFAs using regular expressions, which allow us to specify the patterns we want to recognize in a concise and flexible manner.

Once we have designed the DFAs, we train them using a set of labeled training data, which consists of examples of different gestures performed by multiple users. This training data is used to optimize the DFAs to recognize the desired patterns accurately and robustly. We use various techniques, such as minimization, pruning, and feature selection, to optimize the DFAs and reduce their complexity while maintaining their accuracy.

To evaluate the effectiveness and efficiency of our approach, we will compare it with several state-of-the-art pattern recognition algorithms, such as fuzzy logic, hidden Markov models (HMMs), and convolutional neural networks (CNNs). These algorithms have been widely used in gesture recognition applications and provide a baseline for comparison. We will use several benchmark datasets for gesture recognition, such as the Kinect Gesture Recognition Dataset and the Chalearn Gesture Recognition Challenge dataset, and measure various performance metrics, such as accuracy, precision, recall, and F1 score.

Our results will demonstrate that the automata-based approach is a viable and competitive alternative to other pattern recognition algorithms for gesture recognition applications. The approach achieves high accuracy and robustness while also providing interpretable models for understanding the recognized patterns. We will also discuss the limitations and future directions of our work, such as the extension of the approach to handle more complex gestures and the integration with other sensor modalities, such as audio and depth, to enable more natural and intuitive interactions. Overall, this thesis aims to contribute to the development of more effective and reliable gesture recognition systems by exploring new approaches to pattern recognition.

References

Hong, P., Turk, M., & Huang, T. S. (2000, March). Gesture modeling and recognition using finite state machines. In Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580) (pp. 410-415). IEEE.

Lin, H. I., Hsu, M. H., & Chen, W. K. (2014, August). Human hand gesture recognition using a convolution neural network. In 2014 IEEE International Conference on Automation Science and Engineering (CASE) (pp. 1038-1043). IEEE.

Gao, Q., Liu, J., & Ju, Z. (2021). Hand gesture recognition using multimodal data fusion and multiscale parallel convolutional neural network for human–robot interaction. Expert Systems, 38(5), e12490.

Hakim, N. L., Shih, T. K., Kasthuri Arachchi, S. P., Aditya, W., Chen, Y. C., & Lin, C. Y. (2019). Dynamic hand gesture recognition using 3DCNN and LSTM with FSM context-aware model. Sensors, 19(24), 5429.

Hong, P., Turk, M., & Huang, T. S. (2000, September). Constructing finite state machines for fast gesture recognition. In Proceedings 15th International Conference on Pattern Recognition. ICPR-2000 (Vol. 3, pp. 691-694). IEEE.

Bedregal, B. C., Costa, A. C., & Dimuro, G. P. (2006). Fuzzy rule-based hand gesture recognition. In Artificial Intelligence in Theory and Practice: IFIP 19th World Computer Congress, TC 12: IFIP AI 2006 Stream, August 21–24, 2006, Santiago, Chile 1 (pp. 285-294). Springer US.

Verma, R., & Dev, A. (2009, October). Vision based hand gesture recognition using finite state machines and fuzzy logic. In 2009 International Conference on Ultra Modern Telecommunications & Workshops (pp. 1-6). IEEE.

Elmezain, M., Al-Hamadi, A., Appenrodt, J., & Michaelis, B. (2008, December). A hidden markov model-based continuous gesture recognition system for hand motion trajectory. In 2008 19th international conference on pattern recognition (pp. 1-4). IEEE.

Zeng, Z., Gong, Q., & Zhang, J. (2019). CNN Model Design of Gesture Recognition Based on Tensorflow Framework. *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*.

Islam, M. Z., Hossain, M. S., ul Islam, R., & Andersson, K. (2019, May). Static hand gesture recognition using convolutional neural network with data augmentation. In 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR) (pp. 324-329). IEEE.