Contributions to multimodal interaction in 3D

Doctoral Thesis – Summary

for obtaining the scientific title of Doctor at the
Politehnica University Timisoara
in the field of doctoral studies Computer Science and Information Technology

author eng. Stelian-Nicolae NICOLA
Scientific Supervisor Prof. univ. dr. ing. Lăcrămioara STOICU-TIVADAR

July 2023

The doctoral thesis "Contributions to Multimodal Interaction in 3D" is structured into **7** chapters. It includes **12** tables, **58** figures, and **101** bibliographic references. The main objective of this thesis focuses on creating gestures and their utilization in various applications across different domains: educational, medical, and engineering. Another objective is to leverage new VR/AR/MR technologies for enhanced human-machine/computer interaction. All of these contribute significantly to the field of Computer Science and Information Technology.

Below is a brief summary of each chapter, highlighting the main contributions:

Capter 1 – Introducion

The doctoral thesis focuses on the development and interpretation of hand gestures to be detected by an input device called Leap Motion (LM). The gestures can be classified as static or dynamic through the application of specific techniques. To recognize these gestures, algorithms have been developed based on mathematical formulas for calculating distances, angles, directions, and surfaces in the 3D virtual space. An important aspect in creating gestures is the accuracy and quality of their recognition, which is why convolutional neural networks have been used for more precise classification. The accuracy of gesture detection and classification has been improved by addressing issues such as misidentifying the left or right hand.

These created gestures are applied in the educational field through 3D applications that have been developed to assist users. The use of these interactive applications, which involve gestures as a means of interaction, adds value compared to traditional methods such as a mouse, keyboard, or touch screen (for mobile devices). Various educational 3D applications have been developed in which users can learn about human skeletal bones, amino acids, and DNA nucleotides using gestures. By using gestures to manipulate these applications and integrating concepts of virtual and augmented reality, users feel more engaged, and their concentration increases due to the sense of physical presence within the application.

The main benefits brought to the field of Computer Science and Information Technology, as well as the educational and medical domains, are as follows:

- Development of 7 algorithms for dynamic gesture detection using the Leap Motion (LM) device.
- Comparison of various models and methods for gesture classification in the relevant literature.

- Identification of suitable technologies to enhance interactivity between humans and computers.
- Definition of parameters that describe hand gestures.
- Creation of a dataset for identifying 3 dynamic gestures.
- Classification of dynamic gestures using different models or classifiers based on neural networks.
- Improved quality of defined gestures for this purpose using classification algorithms.
- Enhanced data visualization through 3D representation and the use of virtual reality (VR).
- Increased interactivity between humans and computers.
- Dynamic measurement of interaction within a 3D application.

Chapter 2 - Current State of the Field

This chapter presents the concerns addressed in the specialized literature regarding the recognition/detection and processing of gestures, as well as the utility of virtual, augmented, and mixed reality in various domains.

A gesture is defined as a movement, usually of the body or limbs, that can express or emphasize an idea, emotion/state, action, or attitude.

Among the gestures made with the upper limbs (hands), they can be divided into two categories: static gestures and dynamic gestures. Dynamic gestures involve the movement of the hands or hand over a specified period of time and distance. Some of the well-known dynamic gestures are alphanumeric gestures (gestures that express numbers or letters). Static gestures convey a state or even a number and cannot be defined over a duration of time. Examples of static gestures include the gesture of approval or disapproval and the gesture of showing a number by raising a certain number of fingers.

The Leap Motion (LM) sensor is a device equipped with two monochrome cameras and three infrared LEDs, enabling 3D visualization. It can generate a maximum of 300 frames per second, depending on the performance of the connected computer. Typically, the initial setup of this device for an average computer allows capturing at 120 frames per second. The LM sensor has a USB output for connecting to a PC or laptop. The associated software of the Leap Motion device has the capability to recognize each finger and the joints between the fingers on the hands. Additionally, it comes with four predefined gestures: Screen Tap, Hand Swipe, Key Tap, and Circle. Furthermore, it is possible to define and recognize new gestures by the device. This allows for the creation of databases that provide a detailed description of a newly defined gesture. For example, these databases can store information such as positions, distances, or angles in the three-dimensional space.

On the topic of gesture recognition and classification, over 40 scientific papers have been identified, focusing on achieving more precise and higher-quality detection. The main gestures studied in the literature refer to alphanumeric static gestures described by American Sign Language (ASL).

Other types of gestures that have been identified include flexion and extension, finger opening and closing, finger touch, finger pressing, palm rotation, finger extension and flexion, pronation and supination, radial-ulnar movement, rotation, writing gestures, and gestures for the rock-paper-scissors game. The classification methods for these gestures have involved mathematical formulas for calculating angle and surface intervals, as well as neural network models.

The most commonly used models/classifiers include KNN, MLP, CNN, SVM, ANN, RF, and NB. The highest accuracy scores in gesture detection have been achieved by static gestures and the predefined gestures of the LM sensor, reaching over 99% accuracy in gesture detection.

The main problem identified after studying these articles was the low accuracy in recognizing dynamic gestures, especially when these gestures could lead to confusion regarding which hand, left or right, was being used. Hardware solutions identified in the literature include the use of multiple LM devices or combining the LM device with the Mayo Armband to identify muscle activity. Other solutions involve a rigorous description of the defined dynamic gestures based on features such as finger positions, palm positions, rotation angles, and distances.

In the second part of this chapter, the field of technology is explored, with a focus on virtual reality (VR). The main domains where VR is applied are presented, including the medical, educational, and industrial sectors. Definitions and explanations are also provided for other technologies that have emerged as a result of virtual reality development, such as augmented reality (AR), mixed reality (MR), and extended/experimental reality (XR). The identified goals of using these technologies are:

- Assisting medical personnel in performing interventions.
- Training employees in the manufacturing industry.
- Training medical students in performing surgical procedures.
- Upper limb mobility rehabilitation.
- Home-based recovery and providing a virtual coach to demonstrate exercises.
- Improving the preparation of medical students regarding the steps to be taken in the operating room.
- Training individuals who have suffered a stroke.

As seen above, VR is used in various fields with different purposes, highlighting its wide applicability. It has been observed that VR helps enhance the interactivity between humans and computers/phones, leading to improved outcomes in the aforementioned domains.

Chapter 3 - Definition and Classification of Gestures Recognized by Leap Motion

This chapter presents the key technical steps involved in creating a dynamic gesture for the Leap Motion device, as well as the theoretical basis for classification models for LM gestures. The chapter also introduces the theoretical basis for calculating metrics such as precision, accuracy, recall, and F1 score, along with how to construct a confusion matrix. All the classification models presented are accompanied by concrete examples applied to the scope of the doctoral thesis.
The steps to create a gesture for LM are as follows:

- Verify the connectivity of the Leap Motion device to the PC/laptop.
- Check for the presence of hands in the recorded frames.
- Initialize the position of the palm center for the detected hands.
- Check for the presence of fingers in the frames and initialize the position of the fingertips.

- Calculate distances, angles, and surfaces.
- Verify the range of values obtained from the calculations.
- Generate the gesture.

In this chapter, 12 classification models for gesture recognition were explained through examples. The following are the models used:

- Logistic Regression – LR
- Linear Discriminant Analysis - LDA
- Kneighbors Classifier – KNN
- Decision Tree Classifier - CART
- Gaussian Naive Bayes – NB
- SVC1 (Support Vector Classifier 1) - SVC
- SVC2 (Support Vector Classifier 2) - Liniar SVM
- SVM (Support Vector Machine) - RBF SVM
- Random Forest Classifier - Random Forest
- Ada Boost Classifier - AdaBoost
- MLPClassifier (Multi-layer Perceptron) – MLP
- DNNClassifier (Deep Neural Network)

In the continuation of this chapter, the main terms and formulas used in statistical analysis were presented. After defining these terms, they were accompanied by examples.

The contributions of this chapter to the field of Computer Science and Information Technology are reflected in the model for gesture construction, as well as the real-life examples provided for gesture classification using classification models. The theoretical part of this chapter serves as the foundation for the following chapters, which address the contributions I have made in the field of Computer Science and Information Technology.

Chapter 4 - Gesture Detection Algorithms

This chapter is part of the chapters where the author has made significant contributions in the field of Computer Science and Information Technology. Here, the algorithms created for the detection of defined dynamic gestures are presented. The created gestures include:

- Grasp Gesture 1, 2, and 3 (1 - involving the index finger and thumb, 2 - involving the index finger only, and 3 - involving the thumb and any other finger(s) of the hand)
- Flexion and Extension Gesture
- Hand Rotation Gesture - Pronation and Supination
- Hand Clenching and Opening Gesture
- Complex Gesture of Hand Proximity and Separation

The contributions made in the field of Computer Science and Information Technology presented in this chapter are represented by the created algorithms aimed at detecting the most commonly used gestures for controlling 3D interfaces and the most utilized gestures for hand motion and joint recovery.

The created algorithms are based on frame traversal from Leap Motion to detect the hand positions in the 3D virtual space. They involve mathematical formulas to calculate distances between two points, angles formed, and surfaces and/or semiperimeters formed by the detected points.

To assist users in controlling 3D interfaces, multiple approaches for detecting a gesture involving grasping a 3D object were addressed. Three algorithms for detecting the grasp gesture were described.

Additionally, three gestures with applications in recovery were described, forming the basis of exercises that users need to perform for hand motion recovery.
Since there are few approaches in the literature for creating gestures involving the use of both hands for the Leap Motion device, an algorithm was developed to detect a complex gesture. This gesture has applications in manipulating 3D objects to scale them up or down.

Chapter 5 - Classification Models Contributing to Gesture Precision Enhancement

In this chapter, the classification of three gestures - hand clenching and opening, palm rotation, and hand flexion and extension - is presented using a series of 12 classification models explained in Chapter 3 of this thesis. The idea of classifying the created gestures and achieving even better precision in their detection is based on the following hypothesis:

"The palm rotation gesture - the pronation and supination movement - is misdetected due to the orientation of the thumb, leading to confusion regarding the correct use of the virtual hand."

The wide variety of features describing the gestures in the dataset allows for the construction of new gestures that can be applied to hand mobility recovery. The 29 features describing the initial gestures have contributed to their improved classification. The classification models have shown higher performance after applying features such as the hand and thumb direction vector, hand rotation angle, and the semiperimeter formed by the geometric figure with points at the thumb tip, middle finger tip, and palm center. Lastly, the three features of rotation, tilt, and grace from the 14 features describing the flexion and extension gesture have shown high performance in its classification.

Since the three described gestures have applications in hand mobility recovery, all gestures have been divided into safety levels for monitoring the progress of recovery. The main benefits brought to the field of Computer Science and Information Technology in this chapter are as follows: diverse datasets that can be used in the future for creating new gestures and establishing standard value ranges for precise gesture level definition. Another benefit is the proper selection of the classification model for a gesture from a set of 12 classification models.

After a detailed analysis of the accuracy of the tested models, it can be concluded that the CART (Classification and Regression Tree) model can be used for gesture classification. This model ranks first in terms of classification accuracy, achieving an accuracy of 99%. Even though two different datasets are used for the first two gestures (9,345 data points) and the last gesture (over 2,500 data points), this model demonstrates the best performance. Furthermore, it was observed that there is no relationship between the number of tests and the classification accuracy of gesture levels for this model. Whether there are few or many tests, the classification precision remains the same.

The AdaBoost and DNN models had different results on the two datasets. The main reason why the AdaBoost model performed well on the second dataset is related to the weak classifiers it relies

on. As the AdaBoost model combines weak classifiers (gesture features) at the end of training to produce a strong classifier, the features defining Gesture 3 are optimal to be considered. Although there are more data and more features for Gestures 1 and 2, these cannot be effectively combined to produce a strong classifier.

The DNN model performed poorly on the first two gestures, similar to the AdaBoost model, because it is based on a neural network with a certain number of layers and interconnected neurons. The weak performance of this model on Gestures 1 and 2 is primarily due to the number of neurons and the diversity of the features in these gestures. Both the AdaBoost and DNN models had weak performances primarily due to the diversity of features in these gestures and the number of training data points.

The diversity of gesture features and the number of data points in the dataset are advantageous for the CART model. It achieved the best performance due to its construction. The CART model builds a decision tree based on features, with each node representing a feature. Thus, a dataset with numerous features leads to high performance in gesture classification.

Two models that exhibited different performances on the two gesture datasets are the DNN (Deep Neural Network Classifier) and AdaBoost models. They achieved accuracies of 47% and 60% on the first dataset, and 95% and 99% on the second dataset, respectively. It is evident that these models have varying accuracies in gesture classification.

The achieved accuracy performances on the first dataset ranged from 47% to 99% across the 12 classification models. On the second dataset, the performances ranged from 45% to 99%. The data from both datasets were divided into 75% training data and 25% test data.

To demonstrate and compare the results obtained from the 12 classification models in this chapter, nine tables were constructed. These tables present various comparisons on the test cases of the models, as well as their metrics. Additionally, for each model tested on the two datasets, confusion matrices were created to provide a clearer understanding of the correct and incorrect detections of gesture levels.

Chapter 6 - Gestures in 3D Applications based on Virtual Reality

In this chapter, the application of gestures in 3D applications based on virtual reality (VR) is presented, as well as how to interact with 3D objects through gestures. It demonstrates how hand gestures and head movements can be used in two different use cases.

The main benefits brought to the field of Computer Science and Information Technology are the utilization and processing of gestures for interaction in VR-based 3D applications, as well as the measurement of interactivity in 3D applications through Unity Analytics services.

The use of Leap Motion gestures and hand motion tracking in desktop 3D applications, with virtual hands, and the use of head movement gestures in smartphone-based 3D applications, with VR headsets, bring the user closer to the computer/smartphone. It can be said that by using these types of gestures in 3D spaces, the principle of virtual reality is respected, which states that the user should feel a physical presence within them. All these characteristics and principles are the foundation of mixed reality (MR), which combines the virtual world with the real world to bring

the user closer to the application. Additionally, MR utilizes input devices (controllers, sensors) to facilitate application control. Furthermore, in recent years, when referring to technologies such as virtual reality, augmented reality, or mixed reality, all of them can be encompassed by the concept of extended or experimental reality (XR).

Chapter 7 - Conclusions and Research Directions

The doctoral thesis entitled "Contributions to Multimodal Interaction in 3D" encompasses the research results conducted in the field of gesture-based interaction in 3D applications using AI algorithms that lead to precise solutions in interpreting hand movements.

The main aim of the research is to develop new algorithms that define gestures for the Leap Motion device. These gestures include three grasping gestures, wrist flexion and extension gesture, palm rotation gesture, hand squeezing and opening gesture, as well as the complex gesture of hand approach and separation. At the same time, the objective was to improve the recognition accuracy of these defined gestures. Machine learning techniques were employed to classify the gestures defined by the algorithms. The models used in gesture classification contribute to the improvement of classification accuracy. These defined gestures have applications in the educational and medical domains. In an educational context, gestures enhance the interaction between users and computers in virtual 3D environments. Additionally, the defined gestures can be used in medicine, particularly in hand mobility recovery, playing a role in the rehabilitation process.

Other original contributions in this work involve defining gestures that can be used for interaction in VR-based applications and defining metrics for measuring interactivity in 3D applications.

The main directions for further research are as follows:

- Building new LM gestures using existing datasets.
- Combining two different datasets and applying the initial 12 network models.
- Improving the quality of LM gestures by using end-to-end neural networks. In cases where the gestures cannot be recognized, the network takes over the responsibilities of the Leap Motion device and continues the gesture.
- Applying the constructed gestures to the new generation of the Leap Motion device, Leap Motion Controller 2.
- Utilizing new MR and XR technologies to create new experiences for users who use gestures in 3D applications.
- Applying the developed algorithms to the audio signal processing in virtual reality applications to screen individuals with speech impairments.

By summarizing the contributions made in the doctoral thesis "Contributions to Multimodal Interaction in 3D," a total of **19** research papers were produced during the years of study. Among these, **11** papers are indexed in ISI/WOS (Web of Science), and the remaining **8** papers are indexed in BDI (Bibliographic Databases), which are currently in the process of being indexed in ISI/WOS. Out of these **19** published scientific papers, **12** were published as the first author.

The published papers have received a total of **111** independent citations (excluding self-citations). The citations, grouped according to the international databases in which they are indexed, are as follows:

- **29** citations indexed in Clarivate Analytics Web of Science (ISI Web of Knowledge)
- **82** citations indexed in international databases

Below is the list of publications:

1. **Nicola, S**; Stoicu-Tivadar, L; Virag, I; Crisan-Vida, M, "Leap motion supporting medical education", in Proc. 12th IEEE International Symposium on Electronics and Telecommunications (ISETC 2016), pp. 153-156, Timișoara, România, WOS: 000390717800035 (ISI)
2. **Nicola, S**; Virag, I; Stoicu-Tivadar, L, "VR Medical Gamification for Training and Education", in Proc. 11th Annaul Conference on Health Informatics Meets eHealth (eHealth 2017), Vol. 236, pp. 97- 103, Schloss Schonbrunn, Austria, WOS: 000426828000013 (ISI)
3. **Nicola, S**; Handrea, FL; Crisan-Vida, M; Stoicu-Tivadar, L, "DNA Encoding Training Using 3D Gesture Interaction", in Proc. Special Topic Conference of the European-Fedeeration-for-Medical-Informatics (EFMI STC 2017), Vol. 244, pp. 63-67, Tel Aviv, Israel, WOS: 000450270500013 (ISI)
4. **Nicola, S**; Stoicu-Tivadar, L, "Hand Rehabilitation Using a 3D Environment and Leap Motion Device", Studies in health technology and informatics (ICIMTH 2018), Vol. 251, pp. 43-46, ISSN: 0926-9630
5. Mahmut, E-E; **Nicola, S**; Stoicu-Tivadar, V, "A Computer-Based Speech Sound Disorder Screening System Architecture", Studies in health technology and informatics (ICIMTH 2018), Vol. 251, pp. 39-42, ISSN:1879-8365
6. **Nicola, S**; Lupșe, OS, Stoicu-Tivadar, L, "Novel Gesture Interaction Using Leap Motion in 3D Applications", in Proc. 12th International Symposium on Applied Computational Intelligence and Informatics (SACI 2018), Timișoara, Romania, May 2018, pp. 113-118, WOS:000448144200020 (ISI)
7. **Nicola, S**; Stoicu-Tivadar, L, "Mixed Reality Supporting Modern Medical Education", in Proc. Special Topic Conference of the European-Fedeeration-for-Medical-Informatics (EFMI STC 2018), Vol. 255, pp. 242-246, Zagreb, Croatia, WOS: 000455957400047 (ISI)
8. **Nicola, S**; Stoicu-Tivadar, L; Patrascoiu, A, " VR for Education in Information and Tehnology: application for Bubble Sort", in Proc. 13th International Symposium on Electronics and Telecommunications (ISETC 2018), pp. 343-346, Timișoara, România, WOS: 000463031500076 (ISI)
9. **Nicola, S**; Chirila, OS.; Stoicu-Tivadar, L, "Enhancing Precision in Gesture Detection for Hand Recovery fter Injury Using Leap Motion and Machine Learning", in Proc. 18th International Conference on Informatics, Management and Technology in Healthcare (ICIMTH 2019), Vol. 262, pp. 320-323, Athens, Greece, WOS:000560388600073 (ISI)
10. Mahmut, E-E; **Nicola, S**; Stoicu-Tivadar, V, "CROSS-CORRELATION BASED AUTOMATIC SEGMENTATION OF MEDIAL PHONEMES", in Proc. 14th International Symposium on Electronics and Telecommunications (ISETC 2020), pp. 293-296, Timișoara, România, WOS: 000612681000070 (ISI)
11. Mahmut, E-E; **Nicola, S**; Stoicu-Tivadar, V, "Cross-Correlation Based Automated Segmentation of Audio Samples", in Proc. 18th International Conference on Informatics, Management and Technology in Healthcare (ICIMTH 2020), Vol. 272, pp. 241-244, Athens, Greece, WOS:000630065600062 (ISI)
12. **Nicola, S**; Stoicu-Tivadar, L "Evaluating Interactivity in VR Healthcare Applications Using Analytics", in Proc. 18th International Conference on Informatics, Management and Technology

in Healthcare (ICIMTH 2020), Vol. 272, pp. 225-228, Athens, Greece, WOS:000560388600073 (ISI)

13. **Nicola, S**; Chirila, OS; Stoicu-Tivadar, L, "Gesture Classification for a Hand Controller Device Using Neural Networks", 30th Medical Informatics Europe (MIE 2020) Conference, Vol. 270, pp. 756-760, APR, 2020, WOS:000630065600058 (ISI)

14. Mahmut, E-E; **Nicola, S**; Stoicu-Tivadar, V, "Word-Final Phoneme Segmentation Using Cross-Correlation", Studies in health technology and informatics (STC 2020), Vol. 275, pp. 132-136, Nov, 2020, ISSN:1879-8365

15. Varga, G; Stoicu-Tivadar, L; **Nicola, S**, "Serious Gaming and AI Supporting Treatment in Rheumatoid Arthritis", 31th Medical Informatics Europe (MIE 2021) Conference, Vol. 281, pp. 699-703, Mai, 2021.

16. **Nicola, S**; Stoicu-Tivadar, L, "E Sharing the IT Educational Experience of Developing 3D Applications for Medical Students Training", 19th International Conference on Informatics, Management and Technology in Healthcare (ICIMTH 2021), Vol. 289, pp. 204-207.

17. Mahmut, E-E; **Nicola, S**; Stoicu-Tivadar, V, "Support-Vector Machine-Based Classifier of Cross-Correlated Phoneme Segments for Speech Sound Disorder Screening", 32th Medical Informatics Europe (MIE 2022) Conference, Vol. 294, pp. 455-459, Mai, 2022.

18. Varga, G; Stoicu-Tivadar, L; **Nicola, S**, "Serious Gaming and Artificial Intelligence in Rehabilitation of Rheumatoid Arthritis", 20th International Conference on Informatics, Management and Technology in Healthcare (ICIMTH 2022), Vol. 295, pp. 562-565, Athens, Greece.

19. **Nicola, S**; Chirila, OS; Stoicu-Tivadar, L, "Comparison of Data Classification Results for Leap Motion Recovery Gestures", 20th International Conference on Informatics, Management and Technology in Healthcare (ICIMTH 2022), Vol. 295, pp. 189-192, Athens, Greece.