

**A Hybrid Swarm Intelligence Convolution Neural Network for Object Detection  
and Tracking**

**Afiss Emiola KAREEM  
LCU/PG/001703**

**Being a PhD Thesis Submitted to the Department of Computer Science, Faculty  
of Natural & Applied Sciences, Lead City University, Ibadan, Oyo State, Nigeria**

**In Partial Fulfillment of the Requirements for the Award of Doctoral Degree  
(PhD) in Computer Science**

**2023**

### Certification

This is to certify that Afiss Emiola KAREEM with matriculation number LCU/PG/001703 carried out this research work titled “A Hybrid Swarm Intelligence Convolution Neural Network for Object Detection and Tracking” in the Department of Computer Science, Faculty of Natural and Applied Sciences, Lead City University, Ibadan, Oyo State, for the award of Doctoral Degree (PhD) in Computer Science and that this has not been previously submitted.

.....  
Dr. Wumi Ajayi  
(Supervisor)

.....  
Date

.....  
Dr. Wilson Sakpere  
(Head of Department)

.....  
Date

Do Not Copy Lead City University, Nigeria

## **Dedication**

This research work is dedicated to God Almighty.

*Do Not Copy, Lead City University, Nigeria*

## Acknowledgement

I want to thank this prestigious institution, Lead City University for the Opportunity given to me to learn and complete my PhD program. I equally thank the University Library for the provision of adequate materials needed.

I am grateful to the Department of Computer Science for giving me the privilege and opportunity to study in a conducive environment. I recall with appreciation the wholesome assistance of the Head of Department, Dr. Wilson Sakpere and efforts of my supervisor, Dr. Wumi Ajayi for his good leadership role that has brought out the beauty of this research work. Many thanks go to the Departmental PG Coordinator, Dr. A.A Waheed. God bless you immensely sir. I also acknowledge the invaluable contributions of my lecturers, Prof S. O Akinola, Dr Agbaje, and other departmental staff for their supports throughout this programme. God bless you sirs.

I greatly appreciate my friends, Messrs Awodele Simon Olukayode, Azeez Adeoye and Ezechi Chekwube. Lastly, my profound gratitude goes to my beloved wife and children for their immense support and understanding during the course of this programme, may Allah continues to ease our ways.

I thank you all.

“Even though the above-mentioned institutions and persons have assisted in the process of this research work, I alone stand responsible for the errors, if any, found in the work”.

## Abstract

In recent years, analysis and interpretation of video sequences to detect and track objects of interest had become an active research field in computer vision and image processing. Despite significant efforts in object detection and tracking, an efficient method which provides high computational efficiency has not been developed. Hence, in this work a Hybrid Particle Swarm Optimization Convolution Neural Network (CNN-HPSO) technique was developed to improve computational efficiency in object detection and tracking. The video datasets (MP4 and AVi video formats) used in this work were obtained from a conventional online database and on a real-time basis from YouTube. Multiple frames sampled from the video clips were pre-processed and then segmented. An Enhanced Particle Swarm Optimization (HPSO) was formulated from standard PSO and was applied to Convolution Neural Network (CNN). CNN-HPSO technique was used for edge detection and extraction of the boundary of the image and the object tracking was finally carried out. The work was implemented using MatLab R2016 software. The average results of CNN-HPSO, CNN-PSO and CNN on the videos with MP4 format yielded processing time, accuracy, precision, FPR, sensitivity and specificity of 165.89s, 97.08%, 98.41%, 7.75%, 97.82%, and 92.25%; 179.52s, 94.25%, 96.99%, 10.37%, 95.23% and 89.62%; and 189.19s, 89.95%, 93.64%, 15.56%, 91.95.33% and 84.44% respectively. For the videos in AVi format, CNN-HPSO, CNN-PSO and CNN produced similar average results with processing time, accuracy, precision, FPR, sensitivity, and specificity of 185.09s, 96.62%, 98.23%, 7.80%, 97.34% and 92.19%; 198.24s, 94.83%, 97.62%, 8.56%, 95.46% and 91.43%; and 216.59s, 91.30%, 93.98%, 15.09%, 93.67% and 84.91% respectively. In this research, a CNN-HPSO with associated high computational efficiency was developed. The developed technique can be used for solving other related optimization problems.

**Keywords:** Deep Learning Algorithms, Computer Vision, Moving Objects, Video Frame, Object Segmentation

**Word Count:** 293

## Table of Contents

|  |     |
|--|-----|
| Title Page   | i   |
| Certification  | ii  |
| Dedication   | iii |
| Acknowledgment   | iv  |
| Table of Contents                                      | v   |
| List of Figures  | ix  |
| Abstract   | x   |
| <b>Chapter One: Introduction</b>                       |     |
| 1.1 Background to the Study                            | 1   |
| 1.2 Statement of the Problem                           | 4   |
| 1.3 Aim and Objectives of the Study                    | 5   |
| 1.4 Significance of the Study                          | 6   |
| 1.5 Justification for the Study                        | 6   |
| 1.6 Scope of the Study                                 | 7   |
| 1.7 Limitations of the Study                           | 8   |
| 1.8 Operational Definition of Terms                    | 8   |
| Endnotes   | 10  |
| <b>Chapter Two: Literature Review</b>                  |     |
| 2.1 Introduction                                       | 13  |
| 2.2 Challenge in Detecting and Tracking Moving Objects | 14  |
| 2.3 Concept of Object Detection and Tracking           | 15  |
| 2.3.1 Environment Modeling                             | 17  |
| 2.3.2 Pre-processing                                   | 18  |

|         |  |    |
|---------|--|----|
| 2.3.3   | Moving Object Segmentation                 | 18 |
| 2.3.4   | Object Detection                           | 19 |
| 2.3.5   | Object Tracking                            | 20 |
| 2.3.5.1 | Region-Based Tracking                      | 21 |
| 2.3.5.2 | Contour-Based Tracking                     | 22 |
| 2.3.5.3 | Feature-Based Tracking                     | 23 |
| 2.3.5.4 | Model-Based Tracking                       | 23 |
| 2.4     | Tracking Process                           | 24 |
| 2.4.1   | Object Representation                      | 25 |
| 2.4.2   | Shape Representation                       | 26 |
| 2.4.2.1 | Points                                     | 26 |
| 2.4.2.2 | Geometric Shapes                           | 26 |
| 2.4.2.3 | Silhouette and Contour                     | 27 |
| 2.4.2.4 | Articulated Shape Models                   | 27 |
| 2.4.2.5 | Skeletal Model                             | 28 |
| 2.4.3   | Appearance Representation                  | 28 |
| 2.4.3.1 | Probability Densities of Object Appearance | 28 |
| 2.4.3.2 | Templates                                  | 28 |
| 2.4.3.3 | Active Appearance Models                   | 29 |
| 2.4.3.4 | Multiview Appearance Models                | 29 |
| 2.5     | Feature selection                          | 29 |
| 2.5.1   | Edges                                      | 30 |
| 2.5.2   | Optical Flows                              | 30 |
| 2.5.3   | Color                                      | 30 |

|         |   |    |
|---------|---|----|
| 2.5.4   | Texture   | 31 |
| 2.6     | Segmentation Techniques for Object Detection and Tracking | 31 |
| 2.6.1   | Background Subtraction Method                             | 32 |
| 2.6.1.1 | Gaussian Mixture Model                                    | 33 |
| 2.6.1.2 | Frame Difference  | 33 |
| 2.6.2   | Contour and Threshold Method                              | 34 |
| 2.6.2.1 | Active Contour Model                                      | 34 |
| 2.6.2.2 | Threshold Method  | 35 |
| 2.6.3   | Temporal and Spatial                                      | 35 |
| 2.6.4   | Edge Detection Algorithm                                  | 36 |
| 2.6.4.1 | Sobel Edge Detection                                      | 37 |
| 2.6.4.2 | Canny Edge Detection                                      | 37 |
| 2.6.5   | Optical Flow Segmentation Method                          | 39 |
| 2.6.5.1 | Optical Flow in Motion Analysis                           | 41 |
| 2.6.5.2 | Optical Flow Computation                                  | 42 |
| 2.7     | Optimization Techniques                                   | 45 |
| 2.7.1   | Firefly Algorithm   | 46 |
| 2.7.2   | Particle Swarm Optimization                               | 46 |
| 2.8     | Convolution Neural Networks (CNN)                         | 52 |
| 2.8.1   | CNN Architecture  | 54 |
| 2.8.2   | Performance Metrics                                       | 56 |
| 2.9     | Related Works   | 57 |
| 2.10    | Summary of Gaps in Literature                             | 88 |
|         | Endnotes  | 89 |

### **Chapter Three: Methodology**

|       |  |     |
|-------|--|-----|
| 3.1   | Research Approach  | 99  |
| 3.2   | Acquisition of Video Sequence File   | 100 |
| 3.2.1 | Frame Display  | 101 |
| 3.2.2 | Preprocessing  | 101 |
| 3.3   | Development of a Hybrid Particle Swarm Optimization Convolution<br>Neural Network Model for Object Detection and Tracking System | 103 |
| 3.3.1 | Development of an Enhanced Particle Swarm Optimization   | 103 |
| 3.3.2 | Formulation of Objective Function for Assignment of Optimal Weight   | 104 |
| 3.3.3 | Formulation of Acceleration Coefficient and Velocity Clamping<br>in Standard PSO   | 105 |
| 3.3.4 | The Procedural Steps in Achieving the Detection and Tracking<br>Process with CNN-HPSO  | 111 |
| 3.4   | Implementation of Object Detection and Tracking system   | 114 |
| 3.5   | Evaluation Measures  | 115 |
|       | Endnotes   | 117 |

### **Chapter Four: Results and Discussion**

|     |                     |     |
|-----|---------------------|-----|
| 4.1 | Results             | 118 |
| 4.2 | Results for Video 1 | 118 |
| 4.3 | Results for Video 2 | 128 |
| 4.4 | Results for Video 3 | 129 |
| 4.5 | Results for Video 4 | 131 |
| 4.6 | Results for Video 5 | 132 |
| 4.7 | Results for Video 6 | 133 |

|                                 |  |     |
|---------------------------------|--|-----|
| 4.8                             | Results for Video 7                            | 135 |
| 4.9                             | Results for Video 8                            | 136 |
| 4.10                            | Processing Time for the Techniques             | 137 |
| 4.11                            | Discussion of Results                          | 140 |
|                                 | Endnotes                                       | 150 |
| <b>Chapter Five: Conclusion</b> |  |     |
| 5.1                             | Summary of Results                             | 151 |
| 5.2                             | Recommendations                                | 151 |
| 5.3                             | Contribution to Knowledge                      | 152 |
| 5.4                             | Suggestion for Further Studies                 | 152 |
|                                 | <b>Bibliographies</b>                          | 153 |
|                                 | <b>Appendices</b>                              | 166 |
|                                 | <b>Biodata</b>                                 | 224 |
|                                 | <b>The University Compliance Certification</b> | 236 |

### List of Tables

| <b>Table</b> | <b>Title</b>                                       | <b>Page</b> |
|--------------|--|-------------|
| 4.1          | The Performance of CNN, CNN-PSO and CNN-HPSO       | 123         |
| 4.2          | Processing Time for the Techniques                 | 140         |
| 4.3          | Average of Various Techniques for the Eight Videos | 149         |

*Do Not Copy, Lead City University, Nigeria*

## List of Figures

| Figure | Title  | Page |
|--------|--|------|
| 2.1    | General Block Schematic of Video Surveillance System                                       | 16   |
| 2.2    | Different Approaches regarding Shape Representation  | 27   |
| 2.3    | Despite possible changes in pixel location, brightness in a particular area stays constant | 31   |
| 2.4    | Motion Form Recognition and Focus of expansion   | 43   |
| 2.5    | A simple CNN Architecture, comprise of five layers   | 55   |
| 3.1    | The Structure of the Object Detection System   | 100  |
| 3.2    | Architecture of the CNN-HPSO   | 114  |
| 3.3    | Flowchart of the CNN-HPSO  | 114  |
| 3.4    | Flowchart showing the Techniques for Detection and Tracking Object of Objects              | 116  |
| 4.1a   | Graph of Processing Time against each Technique with Mp4                                   | 126  |
| 4.1b   | Graph of Processing Time against each Technique with Avi                                   | 127  |
| 4.2a   | Graph showing the Accuracy of the Techniques with Mp4                                      | 128  |
| 4.2b   | Graph showing the Accuracy of the Techniques with Avi                                      | 129  |
| 4.3a   | Graph showing the Precision of the Techniques with Mp4                                     | 129  |
| 4.3b   | Graph showing the Precision of the Techniques with Avi                                     | 130  |
| 4.4a   | Graph showing the FPR of the Techniques with Avi   | 130  |
| 4.4b   | Graph showing the FPR of the Techniques with Mp4   | 131  |
| 4.5a   | Graph showing the Sensitivity of the Techniques with Mp4                                   | 131  |
| 4.5b   | Graph showing the Sensitivity of the Techniques with Avi                                   | 132  |
| 4.6a   | Graph showing the Specificity of the Techniques with Mp4                                   | 132  |
| 4.6b   | Graph showing the Specificity of the Techniques with Avi                                   | 133  |

### List of Plates

| <b>Plate</b> | <b>Title</b>   | <b>Page</b> |
|--------------|--|-------------|
| 4.1          | Graphical User Interface by CNN, CNN-PSO and CNN-HPSO Techniques during Training and Detection with online data                    | 115         |
| 4.2          | Graphical User Interface by CNN, CNN-PSO and CNN-HPSO Techniques during Training and Detection with local Acquired Data for Video2 | 116         |
| 4.3          | Graphical User Interface by CNN, CNN-PSO and CNN-HPSO Techniques during Training and Detection with local Acquired Data for Video3 | 116         |
| 4.4          | Graphical User Interface by CNN, CNN-PSO and CNN-HPSO Techniques during Training and Detection with local Acquired Data for Video4 | 116         |
| 4.5          | Graphical User Interface by CNN, CNN-PSO and CNN-HPSO Techniques during Training and Detection with local Acquired Data for Video5 | 116         |
| 4.6          | Graphical User Interface by CNN, CNN-PSO and CNN-HPSO Techniques during Training and Detection with local Acquired Data for Video6 | 116         |
| 4.7          | Graphical User Interface by CNN, CNN-PSO and CNN-HPSO Techniques during Training and Detection with local Acquired Data for Video7 | 116         |
| 4.8          | Graphical User Interface by CNN, CNN-PSO and CNN-HPSO Techniques during Training and Detection with local Acquired Data for Video8 | 116         |

## List of Algorithms

| <b>Algorithm</b> | <b>Title</b>                         | <b>Page</b> |
|------------------|--------------------------------------|-------------|
| 2.1              | Standard Particle Swarm Optimization | 50          |
| 2.2              | Modified Particle Swarm Optimization | 51          |
| 3.1              | Enhanced Particle Swarm Optimization | 109         |

*Do Not Copy, Lead City University, Nigeria*

## Appendices

| Appendix | Title       | Page |
|----------|-------------|------|
| i        | CNN         | 149  |
| ii       | CNN-PSO     | 156  |
| iii      | PSO-HPSO    | 158  |
| iv       | Source Code | 171  |

Do Not Copy, Lead City University, Nigeria