Computer Vision for Enhanced Gameplay: Image Recognition and Gesture Analysis

James Bianco
Institute of Information & Communication Technology
Malta College or Arts, Science & Technology
Corradino Hill
Paola PLA 9032
james.bianco.c10295@mcast.edu.mt

***Abstract*—Gaming has increased in popularity over the years and is constantly growing, with games being released on both consoles and computers. Controllers, keyboards, and mice are used to control the player's actions. However, the human body is never used as a controller by itself using a camera. This paper proposes an evaluation of the accuracy that using Computer Vision with Image Classification, a computer can read simple hand signs made by users, if trained with said hand sign, and can be used in later studies to make a hand controller using a camera. The aim of the research is to identify the weaknesses of Computer Vision and Image Classification, if the camera angle and lighting is changed after training with a clean and well-lit background. The evaluation metrics used are Clean Background, Bad Lighting, Bad Background, and Good Lighting. The results show that while the model was accurate when given good lighting and background, the other experiments had their own issues alongside a few unexpected issues showing up such as: low framerate and delay. The study provides a framework for evaluating and comparing different mixes of lighting and background and highlights the importance of a clean background and good lighting when working with Computer Vision and Image Classification.**

***Index Terms*—Computer Vision, Image Classification, Camera, Hand Signs**

I. Introduction

Gaming has grown as a hobby for many people around the world. It grew from computers to consoles with every generation making their own controllers and improvements on them. Computer Vision and Image Classification have also grown during the years with improvements being made and research being done.

The purpose of this study is to investigate the use of computer vision and image classification to make a new controller which uses hand signs being registered from a camera. The main dependent variable is the accuracy of the computer's confidence in the visible shape that the user will make and display in front of the camera's view, with the independent variables like the dataset and algorithm choice used to investigate the impact on the dependent variable.

Hence the hypothesis of this study is that using a camera and computer vision can be used to make a new controller that can be used for games to make certain actions in-game. But it would come with risks as changing the camera angle and/or the lighting will make a difference in the computer's performance to identify the shape made by the user. The research questions that this study attempts to answer are:

1) Is it possible to use computer vision and image classification with games as a new means of controlling certain actions in-game?
2) How much of a difference in confidence does different lighting and angle make on a computer trained with a clean dataset?

II. Literature Review

A. Computer Vision

Computer vision is a task that uses an algorithm and optical sensors to acquire, process, analyze and understand digital imagery and is used to extract high-dimensional data from the real world to be able to produce numerical or symbolic information. [1] Computer vision includes multiple sub-domains such as scene reconstruction, object detection, event detection, video tracking, object recognition, 3D pose estimation, learning, indexing, motion estimation, visual servoing, 3D scene modeling, and image restoration.

It is then combined with a lighting system to assist in image acquisition which is then continued with image analysis. In deeper detail, stages of image analysis are executed in this order:

1) image formation, where the image of the object is taken and stored in the computer;
2) image preprocessing, where the quality is improved to increase the detail of the image;
3) image segmentation, which is when the object image is identified and then separated from the background;
4) image measurement, where various important features are quantized;
5) image interpretation, where the images are then interpreted by the computer. [2]

B. Image Classification

Image classification is the ability to predict a semantic concept based on the visual content of an image. [3] It is a task that attempts to identify an entire image as a whole to a specific label. Typically, image classification is used to refer to an image where only one object is apparent and is analyzed, to then be set to a specific label. Pixels are separated into different sets based on the values of their data. [4] Traditionally, classification is done by using low-level or mid-level methods to represent an image. Low-level features are based on grayscale density, color, texture, shape and position information, which are defined by a human. Mid-level and learning-based features are then commonly distilled by algorithms, which have been used in image classification or retrieval framework in the past few years. [4] Once the features have been extracted, a classifier is then used to assign labels to different type of objects.

C. Use In Gaming

Computer vision and image classification can be used in collaboration with game development to make interactive games that use the player's camera as a means of controlling the game. Using simple shapes with the players hands, such as a rock, paper, scissors, etc... the player could be able to interact with games differently compared to the usual button pressing on a controller, keyboard, or mouse. But for this to function in a game, it must be quickly responsive and accurate. Aside from that, constant use of a webcam and computer vision will also the increase load on the central processor and video card. [5]

D. Related Work

Other studies used image classification and computer vision to identify pictures with car, aeroplane and bird using the technique named Convolutional Neural Network (CNN) [6] and to section websites based on the images gathered from them by using Optimally Pruned Extreme Learning Machine (OP-ELM) [7].

TABLE I
RESULTS OF OTHER STUDIES

Study	Usage	Methods used	Results
[6]	Images of planes, birds and cars	CNN	Validation Accuracy: 94.2%
[7]	Website Classification based on images	OP-ELM	Websites: 98% Images: 82%

III. RESEARCH METHODOLOGY

A. Problem and Hypothesis

The problem with existing research is that the datasets used for similar tests are either done by comparing with other static images or with cameras that never change angle and the same lighting. (reference 7) The hypothesis of this research is that every person that uses a computer and a camera uses different lighting and camera angles throughout the day, alongside possibly having other objects in the background. This will result in the computer having a harder time identifying what it is looking at, resulting in a lower confidence percentage and therefore taking a larger amount of time to identify objects in the camera view.

B. Aim and Objectives

This study intends to discover if computer vision and image classification can be used for computer games by testing if the computer can recognize three different hand shapes if the background and lighting of the room changes compared to what the images the computer has trained from as camera angles and lighting in a room can differ from time to time. The objectives of this research are:

1) To create a dataset of images consisting of rock, paper, scissors and nothing.
2) Review algorithms for object detection.
3) Develop an algorithm to be used on the dataset
4) Experiment with different lighting and camera angles
5) Evaluate the results of the different lighting and camera angles

C. Research Questions

The research questions that this study attempts to answer are:

1) Is it possible to use computer vision and image classification with games as a new means of controlling certain actions in-game?
2) How much of a difference in confidence does different lighting and angle make on a computer trained with a clean dataset?

D. Research Pipeline

The research pipeline for this study is shown in Figure 1.

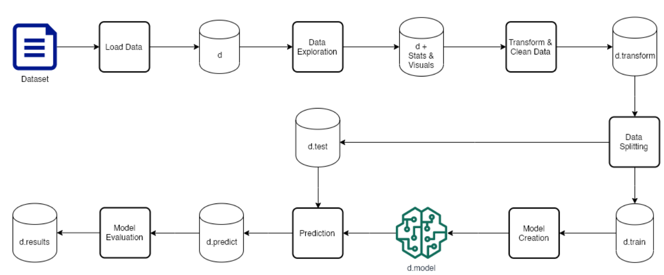

Fig. 1. Research Pipeline for evaluating the effectiveness of computer vision and image classification in the model

E. Technologies

The computer used for this research consisted of an Intel Core i5-8250U @ 1.6GHz (8 CPUs), 1.8GHz, 8GB of RAM, an Intel UHD Graphics 620, and a laptop camera. It is expected to have performance issues due to the low-end hardware of the computer used.

What is required to make a similar project is Python (Version 3.10.10 was used), Python Notebook, Tensorflow 2.2, OpenCV 4x, scikit-learn 0.23x, and NASNETMobile to train the model.

F. Research Method

During the development of the program, it will create its own graphs of the training regarding its accuracy and validation loss. Alongside that, while testing the shapes, the camera angle and lighting will be changed to compare the difference in confidence that the computer will have. The lowest and highest percentage confidence seen will be documented in a table for comparison.

IV. Findings & Discussion of Results

A. Unexpected Results

During the gathering of results, it was noted that distancing from the camera makes a difference in the computer's confidence and that the frame-rate of the camera and overall performance of the computer had a delay of around 0.5 seconds and had a frame-rate drop to around 15fps. This result mixed in with holding a game running in the background would leave the computer running at an even lower frame rate which would be detrimental to the player if they were attempting to play a game that required a camera and computer vision running at the same time. During the testing of the computer's confidence, it was noted that raising the arm up for the computer to recognize the shapes being made was physically straining and exhausting. This would result in much needed breaks if done for hours by users while playing a game which required such movements to commit certain actions.

B. Graph Discussions

Regarding the "Training Accuracy", the computer was able to spike up to 0.99 and have a few drops and jumps inside 0.95 and over, never dropping below 0.99 with its highest accuracy reaching 1.0. "Validation Accuracy" started off in 0.98 and swiftly raise up to 1.0 and then never dropping again for the rest of the accuracy training. See Figure 2 for more details.

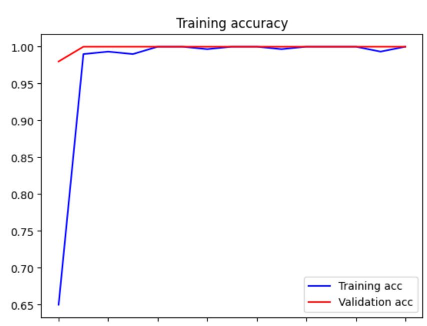

Fig. 2. Graph of Training Accuracy

Regarding "Training loss", it started off at 0.99 and rapidly decreased to 0.1 then lowering to 0.02 at the very end of the graph. "Validation Loss" started off at 0.4 and quickly decreased to 0.1, then to 0.05 and finally to 0.02 at the end of the graph. See Figure 3 for more details.

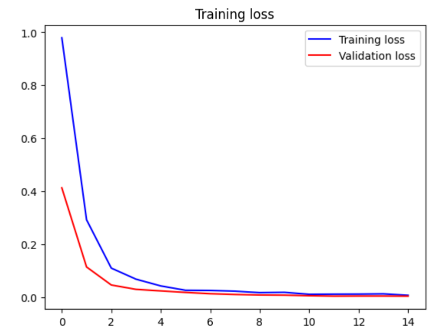

Fig. 3. Graph of Training Loss

C. Analysis of Tested Results

Four tables were created with the results gathered from the testing of the project. When given a good background like the ones used for the training of the project with good lighting, the lowest percentage was of the "Rock" at 85% confidence while the highest percentage of confidence was at 99% in "Nothing", "Paper" and "Scissors". Being given the same clean background but mixed with bad lighting results in lower percentages, with "Nothing" having an 83% confidence, but still having a confident 99% in both "Paper" and "Scissors". When the background has been changed and the lighting is returned to its better state, "Nothing" is read as "Scissors" with a low percentage of 58% and a high 79%. The rest of the shapes were read as their own selves and not recognized as any other shape, with the lowest percentage of confidence at 69% in "Paper" and the highest being 99% in "Scissors". A bad background and bad lighting compared together results in the lowest confidence with "Nothing" being read as "Scissors" at a low percentage of 47% and a high 63%, "Rock" was resulting in flickering between a low percentage of 56% as "Scissors" and a regular "Rock" at 74%. "Paper" suffered the least with the lowest percentage being at 92% and a high 99%. "Scissors" lowered the most at 78% and a high 94%. See Tables II, III, IV and V for further details.

TABLE II
Clean Background and Good Lighting Results

Shape	Lowest Percentage	Highest Percentage
Nothing	98%	99%
Rock	85%	96%
Paper	99%	99%
Scissors	98%	99%

TABLE III
Clean Background and Bad Lighting Results

Shape	Lowest Percentage	Highest Percentage
Nothing	83%	94%
Rock	94%	97%
Paper	99%	99%
Scissors	96%	99%

TABLE IV
BAD BACKGROUND AND GOOD LIGHTING RESULTS

Shape	Lowest Percentage	Highest Percentage
Nothing	Scissors at 58%	Scissors at 79%
Rock	83%	94%
Paper	69%	91%
Scissors	87%	99%

TABLE V
BAD BACKGROUND AND BAD LIGHTING RESULTS

Shape	Lowest Percentage	Highest Percentage
Nothing	Scissors at 47%	Scissors at 63%
Rock	Scissors at 56%	74%
Paper	92%	99%
Scissors	78%	94%

V. CONCLUSION

This study sought to evaluate if computer vision and image classification could be used as a new means of controlling certain actions in games by testing the confidence of the computer if given a clean and well-lit dataset. The analysis involved creating the dataset and training a model to be able to test whether different background and lighting will make a large difference in the computer's confidence.

The results showed that with a clean background and good lighting, the idea of using computer vision as a new means of a controller for games is possible. Unfortunately, when the angle of the camera and/or the lighting is bad, the computer's confidence lowers and, in some cases, identifies a different shape due to objects in the background interfering with the computer's vision. Due to users, specifically laptop users changing their camera background, it would be difficult to be able to develop a game that can register the correct shapes that the user makes. In addition, during the testing it was identified that the computer's performance reduced the camera's framerate to approximately 15fps alongside a 0.5 second delay. Running a game in the background would only make the performance of the camera worse as games use up a large amount of the computer's performance. Furthermore, straining of the arm was also noticed as the camera requires the user to use a specific corner of the display for the computer to identify which shape the user is displaying using their arm. Which results with the users choosing alternatives to using the computer vision if such an option was available to them. These results indicate that the hypothesis holds true and that research question one, which sought to investigate if computer vision and image classification can be used a new means of controlling certain actions in-game is answered to be no. And that research question two, which sought to identify how different the confidence the computer would have depending on the different angle and lighting given to the camera would make be, which the results show that it makes a large difference when the angle and/or lighting is changed. This study presented a framework that can be used for evaluation and comparison when other researchers conduct their own research.

REFERENCES

[1] R. Klette, "Concise computer vision," *Undergraduate Topics in Computer Science*, 2014.
[2] V. Wiley and T. Lucas, "Computer vision and image processing: A paper review," *International Journal of Artificial Intelligence Research*, 2018.
[3] T. Durand, N. Thome, M. Cord, and S. Avila, "Image classification using object detectors," in *2013 IEEE International Conference on Image Processing*, 2013, pp. 4340–4344.
[4] A. Mohamed, A. Hemeida, and M. Hassan, "Image classification based deep learning: A review," *Aswan University Journal of Sciences and Technology*, vol. 2, 06 2022.
[5] G. Chursin and M. Semenov, "Using computer vision in the gameplay of educational computer games," *Journal of Physics: Conference Series*, vol. 1989, 2021.
[6] A. Sharma and G. Phonsa, "Image classification using cnn," *SSRN Electronic Journal*, 2021.
[7] A. Akusok, Y. Miche, J. Karhunen, K.-M. Bjork, R. Nian, and A. Lendasse, "Arbitrary category classification of websites based on image content," *IEEE Computational Intelligence Magazine*, vol. 10, no. 2, pp. 30–41, 2015.