The Building, Use, and Importance of Intelligent Cameras in the 21st Century

ENDRIT NGJELINA[1*]

[1]*Innovation Lab, Saint Petersburg College, Seminole, FL 33772, USA*

*endritngjelina2000@gmail.com

Abstract: The field of computer vision has experienced drastic technological breakthroughs since the end of the 20th century. Technology is becoming more prominent in people's lives each year; therefore, intelligent cameras are attracting their deserved attention. Their power and ability are unseen and fit for plenty of demands. The AIY Vision Kit is an intelligent camera that has been experimented with and researched carefully. This research illustrates the fundamentals of intelligent cameras, how they are used, and their importance in the future. Information has been detailed with its respective explanation, picture, and additional knowledge to aid in a deeper sense of understanding of the topic. This paradigm is conducted and concluded through extensive research, examples, and references.

Keywords: artificial intelligence, camera, Python, image recognition

Introduction

In the 21st century, the world is surrounded by astonishing technological advancements, machines, and gadgets that were only thought of as science fiction a century prior. This technical sophistication has prompted the public with innovative devices alongside great features that accompany them. An intelligent camera, just like the name implies, is a camera with versatile functions. Behind the lens that all are familiar with hides complex software that is trained to equip the camera with an array of features. They are customizable, but all fall under the umbrella of computer vision and artificial intelligence. The universal advancement towards automated machines strongly indicates the importance of intelligent cameras in the future.

In [1] the authors note that "The camera is a key sensor to achieve a reliable environment perception," referring to its role in the car's navigation system. In addition, they make work more efficient, precise and they are extremely reliable. Considering regular cameras' impact on all human landscapes, one can expect the same result with intelligent cameras. They are like standard cameras but with tweaks that make them multifunctional and more desirable. This is a great feature because not only does it save storage space and resources, but it also saves time for the appropriate entities. In [2] a surveillance system from intelligent cameras is described as follows, "This system is reliable and meets the aim of a modern intelligent surveillance system by combining multiple approaches to detect intrusions and to inform users effectively ." The camera system would implement face recognition software that would regularly update and have its database for employees. An intruder's face would not be recognized; hence it would be captured and reported to the appropriate authorities. Such technological improvement is escalated by artificial intelligence and its ability to learn, progress, and keep going. While intelligent cameras are still relatively young and have not been fully explored, their influence and ability portray their importance in the future.

To set the stage for anyone interested in computer vision, one must first be familiar with the fundamental concept. Computer vision works by training machines on large amounts of visual data and creating an algorithm through which the computer classifies the object. This concept is based on machine learning,

which is the potential or capability of machines to replicate intelligent behavior through extensive training, exploring, and studying.

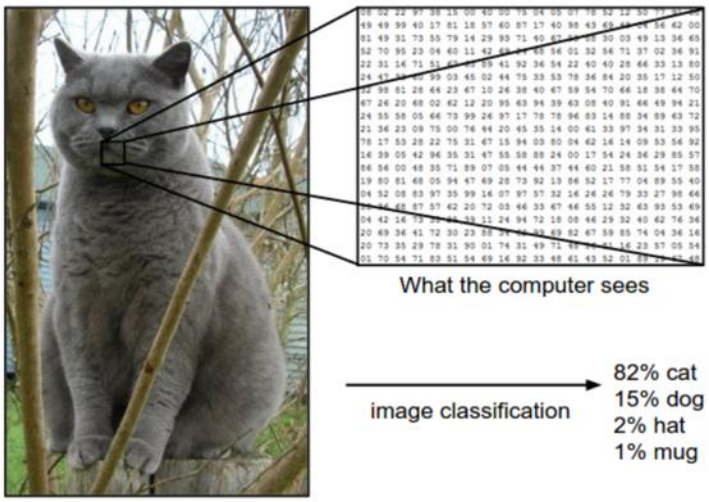

Fig.1. The photo is analyzed, and the computer classifies the image. [3]

The picture above sets an example of how computer vision works. Before classifying the photo, the processing unit examines the pixels, their placement, and other factors. The computer calculates the probability through trial and error and then classifies the object. In [3] the authors note that" Thanks to the use of deep learning in image recognition and classification, computers can automatically generate and learn features – distinctive characteristics and properties. And based on several features, machines predict what is on the image and show the level of probability".

Methods

The AIY Vision Kit is an intelligent camera created by Google LLC. It is crafted in 17 parts that, through careful handling, come together to form an intelligent camera. The main parts of the intelligent camera are the Vision Bonnet, the Raspberry Pi Zero WH, and the Raspberry Pi Camera v2. The Vision Bonnet is a minicomputer specializing in computer vision and artificial intelligence. The Raspberry Pi Zero WH is a tiny board that enables intelligent cameras with wireless and Bluetooth connectivity. The Raspberry Pi Camera v2 is a custom-designed 8-megapixel camera with an image sensor and a fixed focus lens. When powered up, this device can capture and recognize random objects in sight. It distinguishes their color, form, and size. It is embedded with facial recognition software, recognizing when someone is happy, angry, or shows other emotions. The instructions for building this intelligent camera can be found on their dedicated page [4]. Any enthusiastic builder can build and perform the demos available on the web page. They include the Joy Detector, the Image Classification Camera, the Face Detection, and more. Aside from the intelligent camera, one needs a monitor, a smartphone, chargers, and a display port cable to recreate this research. One can use other devices to enhance the research experience, but the devices mentioned above are the minimum requirements.

Results and Discussion

Building an intelligent camera can be challenging. Below are some problems one can face while putting everything together.

Fig. 2. Cable is not properly secured to the latch.

The Raspberry Pi Camera v2, the Vision Bonnet, and the Raspberry Pi Zero WH are connected through a long flex cable and a short flex cable, being tightly secured in their respective latches. If this cable is not secured, the Vision Kit will not work. All parts must be connected to communicate with the camera.

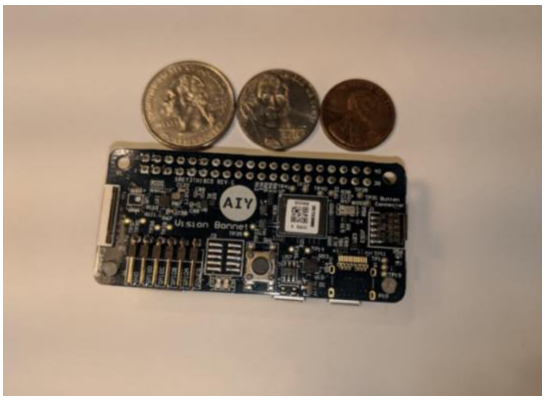

Fig. 3. Parts are tiny and hard to deal with. [5]

When dealing with the boards or other electronic parts of the Vision Kit, it is suggested to be careful and aware of how fragile they are. Above is a depiction of the Vision Bonnet, the main board which powers up the device, and it is as big as three coins.

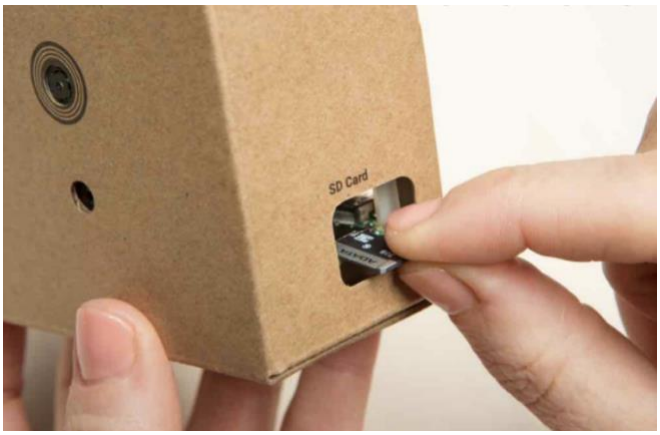

Fig. 4. The SD card is not working. [4]

Another technical issue that can be expected is a corrupt SD card. The files within it may not run properly; therefore, the Vision Kit does not work. Thankfully, there is an easy solution to this, which is reformatting the SD Card.

Fig. 5. Cable flex is turned around.

If one thoroughly follows step three from the dedicated web page, one soon discovers that the kit will not run. To fix this, the copper strip should be turned the other way. This step is incorrectly displayed on the web page, and it should be fixed by the team.

Fig. 6. The power source is not enough.

Fig. 7. The power source is enough.

If the power output is not big enough, the kit will not start. The power plug on the left has an output of 350mA, which was insufficient to start the kit. The kit booted up when switched to the one on the right with an output of 3A.

The first demo that is put to the test is the Joy Detector. As mentioned, the vision kit can detect someone's face and show their feelings. It does so by lighting the main button in different colors based on the person's facial expressions. This effect works on people, and surprisingly it works on photos or videos of people. One must point the camera at anyone and watch the color change according to their joy. Below is an example of this demo in action.

Fig. 8. Button turns bright in color.

Fig 9. Joyful face is detected.

In the picture on the right, someone is depicted through an intelligent camera. On the top left, the program displays the number of faces in the frame and their joy. The joy index is 82% (or 0.82), and when the camera

detects a joyful face, the main button changes color to bright pink or yellow, as depicted in the picture to the left.

Fig. 10. Button turns blue color.

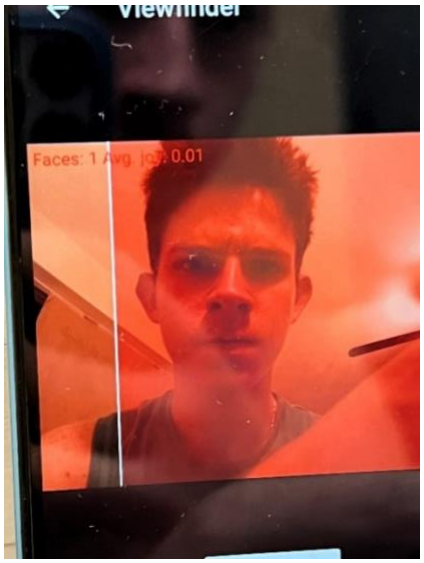

Fig. 11. Sad face detected.

The joy index has dropped to 1% (or 0.01) in this example. The intelligent camera quickly reads off the facial expression; when the face is not joyful, the color in the main button shifts to blue or dark blue.

Fig. 12. Button turns bright in color.

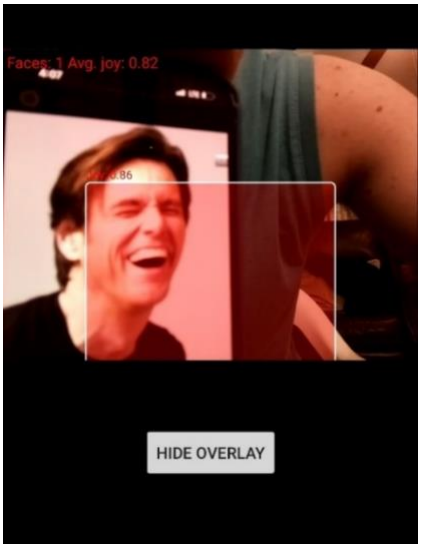

Fig. 13. Joyful face detected.

Here, another face is detected but from a photo on the internet. As usual, it detects the joy index, which is 86% or 0.86, and the light correctly turns pink or yellow, to indicate joy.

Another mind-blowing feature of this intelligent camera is a demo called "Image Classification Camera," it identifies, classifies, and displays data about any object just by pointing the camera toward it. This is done through many files in the source code that apply their data set to the input from the camera, and whichever matches is returned as output in the terminal. In other words, there is already a database for common objects in the source folder, and if you display it in this demo through its shape, size, and color, the database will

return those statements that match your object. Below is the output from the terminal when a keyboard was placed in sight.

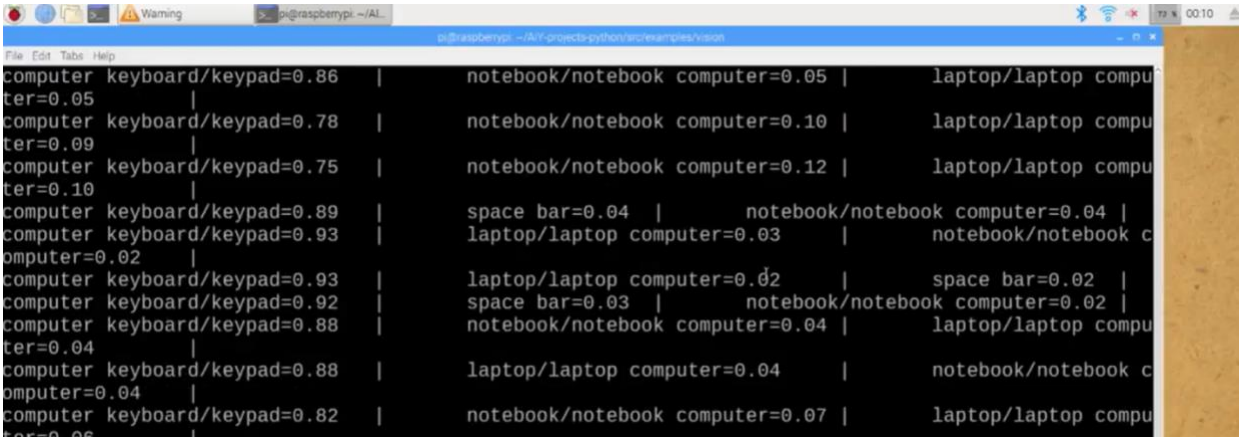

Fig. 14. Terminal outputs correct assumption for object.

On the left side of the screenshot, the type of object is displayed and the certainty of the program. It classified the object as a keyboard with a certainty of 75% (or 0.75) and as high as 93% (or 0.93). On the middle and right side are other options, but they are less likely to be correct since the certainty of the program is extremely low, ranging from 12% down to 1%. This certainty percentage is displayed because the camera does not directly examine the object but compares it to the data of common objects already stored in its source files and then makes an assumption. With some thinking, one might derive that it is possible to display an object not in the database, which is the case for the following example in the terminal.

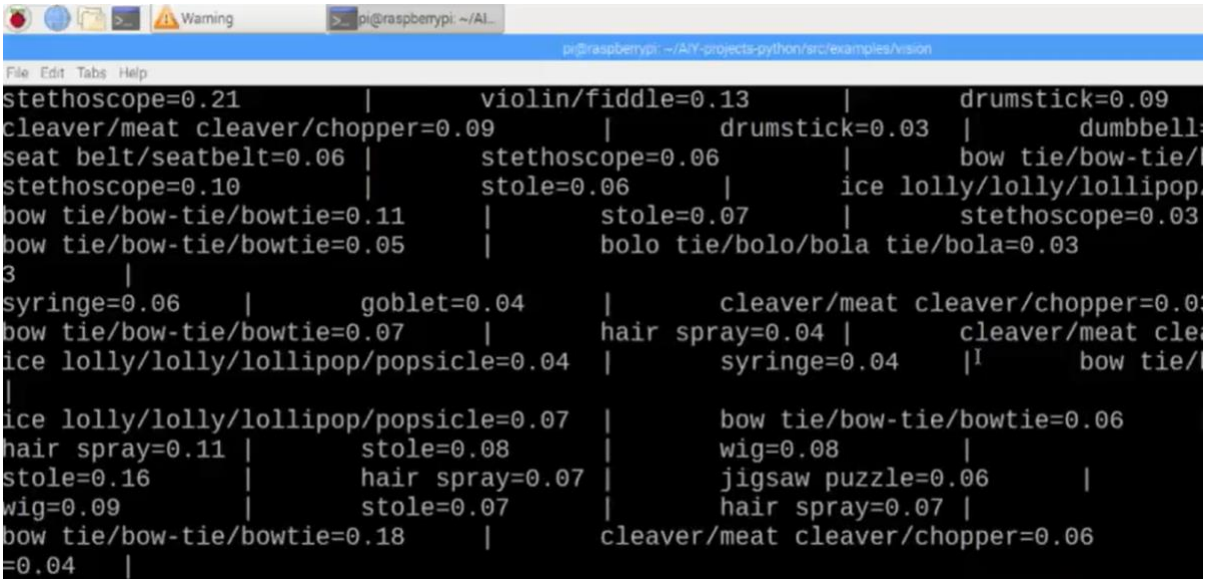

Fig. 15. Terminal outputs incorrect assumption for object.

The object in the instance above is a pair of pliers, but the program does not recognize them. This is most likely since an object named ‘pliers’ is not in the kit’s source code. Since there is not one, the program tries to find similar matches to this object. Pliers, lollipops, bow ties, and seat belts all share a long – vertical

shape. Another detail to determine that the program is not sure about the object in question is the certainty that it displays. The most significant value regarding certainty is 21%, assigned to a stethoscope. Compare these answers to those shown in the previous example with the keyboard; drastic differences can be noted.

The last demo tried is the face detection feature of the kit. It may not be as exciting as the one prior, but it plays a significant role in the overall functioning of the camera. Alongside the face detector code, a face tracker feature is programmed in the source files. This enables the vision kit to keep track of the people in the shot and simultaneously record data about their face's geometry and expression. Below, the output of this demo is displayed.

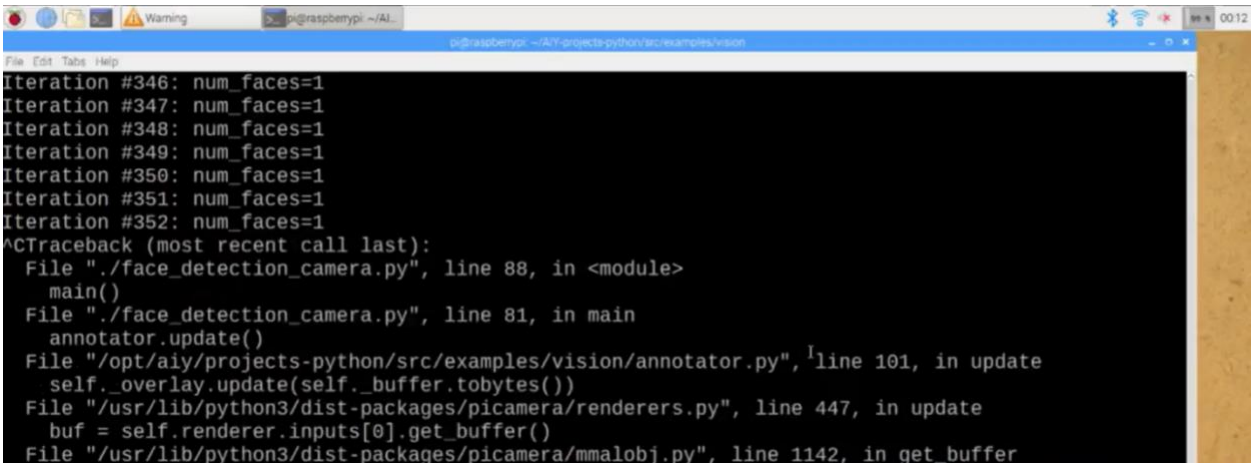

Fig. 16. Terminal outputs correct number of faces detected.

All the lines start with the word iteration, which means that the face is recorded for every frame that the camera captures. The three-digit number next to the iteration is the frame number. After that is displayed the number of faces that were detected. In other words, the intelligent camera does not record a video of what is happening, but rather it captures photos and distinguishes them from one another by iterating through each frame for any changes or abnormalities between them. This is quite effective because it pays great attention to detail and minor things that might go unnoticed.

And lastly, some of the source files were opened in Visual Studio Code, an Integrated Development Environment (IDE) for all programs. While it is difficult to fully understand why and what each command line means, the general purpose for opening it is to see how the software is divided into sections and how they connect. The vision kit is created using Python, which is known for being a versatile and user-friendly programing language. A screenshot of the code is provided below.

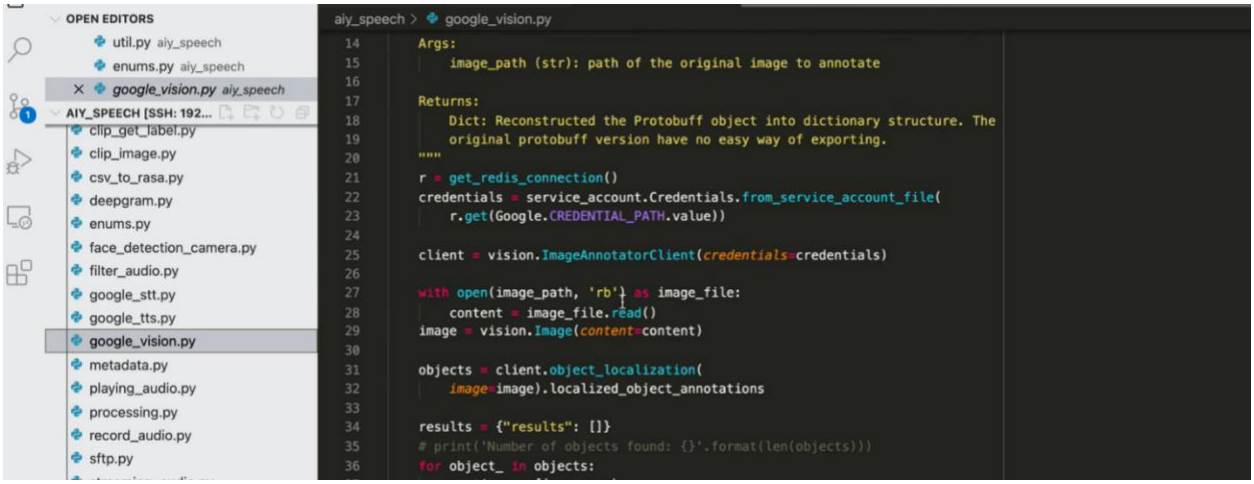

Fig. 17. The source files of the intelligent camera are displayed.

On the left side, all the files comprise the software used in the kit. All the processing and 'intelligent' work is done here. This is what differentiates this camera from a regular plain camera; it contains a lot of power and potential. Inside the dark background, the code of the 'google_vision.py' file is displayed. If one looks at the top of the photo, one notices that this file is located inside the 'aiy_speech' folder. Explained in simple terms, this nesting of files inside another is done to efficiently apply software in any given situation.

Conclusion

Although not widely used, intelligent cameras will soon be almost everywhere due to their incredible performance and the features that they bring along. Casual cameras that are installed on rooftops of banks or traffic lights record footage, but if something noteworthy were to happen, one would need to filter through all that footage and find the accident scene or the crime committed. In comparison, an intelligent camera would be able to filter and delete all the unnecessary footage, separate the noteworthy event, and make it stand out. Artificial intelligence is so good at learning that many of the world's brightest minds fear the future of technology and the way artificial intelligence will evolve. In the same way, all the magic in the kit is happening between the Raspberry Pi Zero WH and the Vision Bonnet. Those components collect, store, separate, analyze, and portray data. They obtain the capacity and power to analyze and distinguish data in an organized and understandable manner. This is what today's technology can achieve; it is astonishing, to say the least, and the reason why intelligent cameras will be everywhere.

Acknowledgments. This material is based upon work supported by the Innovation Lab at SPC.

Disclosures. The authors declare no conflicts of interest.

References

[1] S. Genser, S. Muckenhuber, S. Solmaz and J. Reckenzaun, "Development and Experimental Validation of an Intelligent Camera Model for Automated Driving," *Sensors,* vol. 21, p. 22, 2021.

[2] F. Khodadin and S. Pudaruth, "UOB Journals," 01 November 2020. [Online]. Available: https://journal.uob.edu.bh/handle/123456789/4046. [Accessed 18 November 2022].

[3] K. Spirina and A. Zharovskikh, "InData Labs," 9 June 2020. [Online]. Available: https://indatalabs.com/blog/how-does-computer-vision-work. [Accessed 1 December 2022].

[4] A. Projects, "Vision Kit," 10 January 2020. [Online]. Available: https://aiyprojects.withgoogle.com/vision. [Accessed 1 December 2022].

[5] C. Hart, "webrtchacks," webrtchacks, 06 March 2018. [Online]. Available: https://webrtchacks.com/aiy-vision-kit-tensorflow-uv4l-webrtc/. [Accessed 19 November 2022].