ISSN: 2508-7894 © 2021 KODISA & KAIA.
KJAI website: http://www.kjai.or.kr
doi: http://dx.doi.org/10.24225/kjai.2021.9.2.23

# An image analysis system Design using Arduino sensor and feature point extraction algorithm to prevent intrusion

**Myung-Jae LIM[1], Dong-Kun JUNG[2], Young-Man KWON[3]**

**Received: October 31, 2021. Revised: November 23, 2021. Accepted: December 05, 2021**.

## Abstract

In this paper, we studied a system that can efficiently build security management for single-person households using Arduino, ESP32-CAM and PIR sensors, and proposed an Android app with an internet connection. The ESP32-CAM is an Arduino compatible board that supports both Wi-Fi, Bluetooth, and cameras using an ESP32-based processor. The PCB on-board antenna may be used independently, and the sensitivity may be expanded by separately connecting the external antenna. This system has implemented an Arduino-based Unauthorized intrusion system that can significantly help prevent crimes in single-person households using the combination of PIR sensors, Arduino devices, and smartphones. unauthorized intrusion system, showing the connection between Arduino Uno and ESP32-CAM and with smartphone applications. Recently, if daily quarantine is underway around us and it is necessary to verify the identity of visitors, it is expected that it will help maintain a safety net if this system is applied for the purpose of facial recognition and restricting some access. This technology is widely used to verify that the characters in the two images entered into the system are the same or to determine who the characters in the images are most similar to among those previously stored in the internal database. There is an advantage that it may be implemented in a low-power, low-cost environment through image recognition, comparison, feature point extraction, and comparison.

**Keywords :** ESP32-CAM, PIR Sensor, Arduino, Image Precessing, Image Analysis

**Major Classification :** Artificial Intelligent, Image Classification, Feature Exdtraction

## 1. Introduction

Various countries and fields. Face recognition is largely divided into three stages: face detection, face matching (face recognition), and recognition result derivation. First, the face image is divided, and which of the features of the face is selected is determined (Yoo, Seo, Kim, & Kim, 2019). After that, when face recognition is achieved, the final recognition result is derived through combination and comparative analysis with other information. Face recognition system refers to a computer-assisted application that automatically identifies each person through a digital image. It is a technology that scans, stores, and recognizes face shapes or thermal images through thermal infrared photography, three-dimensional measurement, and skeletal analysis, and is used to verify identity compared to face images captured by cameras and stored photo databases (Li, Zhang, Gao, Jiang, & Chen, 2018). Face recognition is a biometric method that can be conveniently used without having to memorize or carry, so it is emerging as a key authentication tool for ICT services, and its application area is expanding to various fields not only in Korea but also in major foreign countries. In particular, more and more countries are using face

1 First Author, Professor, Department of Medical IT, Eulji University, Korea, Email: lk04@eulji.ac.kr
2 Co-Corresponding Author, Professor, Department of Medical IT, Eulji University, Korea. Email: tchung@eulji.ac.kr.
3 Corresponding Author, Professor, Department of Medical IT, Eulji University, Korea, Email: ymkwon@eulji.ac.kr

ⓒ Copyright: The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://Creativecommons.org/licenses/by-nc/4.0/) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

recognition technology as a way to simplify authentication procedures at airports, terminals, and bank transactions, but security problems are also coexisting due to concerns over personal information infringement. There are several types of face recognition systems for personal authentication. Among them, it is common to use an image system, computer, and CCTV that capture face images, and in some cases, infrared cameras are used. Face images change face image data obtained according to aging, hair color, facial expression, disguise, and lighting direction, making it difficult to apply them to areas requiring high security. As a result, multimodal methods that combine other biometric methods such as face recognition, iris recognition, and fingerprint are increasing. As such, face recognition technology is becoming more advanced thanks to the development of artificial intelligence technology. Face recognition technology starts with face detection technology. In other words, the core of face recognition technology is to identify characteristic points in the detected face information and match them with individual face information. The development of artificial intelligence, related algorithms such as deep learning, and the increase in computational speed using GPU have dramatically increased the speed of existing artificial intelligence computation, and face recognition technology using artificial intelligence is developing faster. In the field of copyright, research and discussions on the use of face recognition technology are also being actively conducted (Gosselin, Murray, Jégou, & Perronnin, 2014). After recognizing a person's face in a video or image using face recognition technology in various works, a method of detecting whether it is an illegal distribution image or image based on the information is mainly used (Kang, Seo, Lee, & Kang, 2017). Therefore, in this paper, a system for preventing unauthorized intrusion was proposed through interworking with Arduino Uno, ESP32-CAM, and smartphone applications. It detects objects and people through Arduino boards, PIR sensors, and ESP32-CAM, and in particular, deep learning techniques are applied to the process of face recognition to extract features and determine them by comparing them with basic facial features of storage fields in existing database (Kong & Lee, 2017).

## 2. Literature Review

### 2.1. Semi-living Characteristics.

Physical or behavioral features used to recognize a person's identity are called biometric features, and the technology to identify users using them is called Biometrics. Here, physical characteristics refer to characteristics of the human body itself, such as vein patterns such as face, fingerprints, iris, and hands. On the other hand, behavioral characteristics refer to unusual patterns seen in a person's behavior, such as voice, signature, and gait. These biological features should be distinguishable from others as they are used to identify people, and should have characteristics that do not change over time (Guo, Dong, Li, & Gao, 2017).

### 2.2 SIFT (Scale Invariant Feature Transform)

#### 2.2.1 Scale space

The reason for the scale space treatment is the work that must be done to create a scale invariant property. This means that it is a space formed by gathering several scales. As shown in the figure below, gradually blurring an image twice the original size, and then gradually creating a different scale in this way, 1/2 of the original size, 1/2 of the original size, and so on. Here, oct means octave, and 1 octave means that images arranged vertically become one group. In this way, scale space is a work done to allow objects to compare the same feature points at a certain distance.

#### 2.2.2 Dog operation.

Dog (Difference of Gaussian) operation is an operation used to extract edges or corners, that is, borders, from an image. This can be done by subtracting two adjacent blurring images from the same octave obtained in the previous step. If you remove two adjacent blurring images within the same octave as shown in the figure below, you will see the dog image shown in the black picture. The dog images in the middle are not just black pictures, but if you look closely, the edge is extracted.

#### 2.2.3 Finding the key point.

This is the step of finding the approximate location of the maximum and minimum values within the Dog image. When determining the maximum and minimum values in one pixel, three dog images within the same octave are required. A dog image to be checked and a dog image with a scale of one step at a time are required. 8 pixels around the pixels to be checked now and a total of 26 pixels of Dog images above and below scale will be inspected. When the pixel you are checking now is the largest or smallest pixel among 26 pixels, it becomes a key-point.

#### 2.2.4 Key point selection.

Any of the key points obtained from the Key-point Finding in Step 3 should be removed. First, it removes key points with low contrast. Second, remove the key point on the edge. In the method of removing a key point having

a low contrast, the pixel value of the key point from the dog image is If it is less than a specific threshold, remove it. When Dog finds the edge, it can react sensitively with just a little noise to find the edge, so it is necessary to remove the extreme values on the edge. To this end, calculate the horizontal and vertical gradient at the key point. And if there is little change in both directions depending on how the horizontal and vertical pixel values change based on the key point, it can be judged as a flat region. If the pixel change is large in both horizontal and vertical directions, the change is large in both corners and vertical directions, but if the change is small horizontally, it can be judged as an edge (Steela, Birdsong, & Reddy, 2019).

#### 2.2.5 Key point direction allocation

This is a step of having rotation invariance by assigning directions to key points. This is to collect the gradient direction and size of each key point to find the most prominent direction and allocate it to the direction of the key point. Calculate the gradient size and direction of the pixels in the small window in the figure on the left. After that, divide 360 degrees into eight angles and fill the size and direction. Through this, each of the 16 small windows has values of 8 angles, so if you do 16 * 8, you get 128 numbers.
This is the unique number of the Key-point.

#### 2.2.6 Calculate SIFT features.

You must specify a unique number for the selected key points. Each Key-point unique number is represented by 128 numbers. In order to designate a unique number, it is necessary to understand the tendency of the change in the shape of the main surface of the key point, and 16 * 16 windows are set around the key point. This window consists of 16 small 4*4 windows.

### 2.3 Machine learning.

Machine learning consists of several types of machine learning models that apply various algorithmic techniques. Depending on the characteristics of the data and the desired results, one of the four learning models can be applied: guidance, unsupervised, quasi-supervised, and reinforcement. Depending on the dataset you are using and the desired results, you can apply one or more algorithmic techniques within each model. Machine learning algorithms are basically designed to classify objects, discover patterns, predict results, and make information-based decisions.

## 3. Image Analysis System

### 3.1 Recognition algorithm

Face recognition is a robot that embodies the ability of a person to recognize who the other person is by looking at the face. Although all video recognition is the same, in the case of face recognition, in order for robots to recognize people, they must first teach them who they are. In facial recognition, this process is called user registration. Through this registration process, the face information of several people to be recognized is remembered, and when encountering a person, the recognition is made by judging which of the people who remember the face. In other words, face recognition is a technology that evaluates the similarity between registered face information and newly input face information. However, the problem here is that when the face of a person that a robot finds in everyday life is accepted through a camera, it appears in a wide variety of forms. The image accepted through the camera may include various noises, and a background that is not related to the face is also included. It undergoes a preprocessing process to remove these unnecessary parts. By performing various processing on the image accepted by the camera, it is possible to help with recognition by deleting details unrelated to the recognition that originally existed in the image. The core of the recognition algorithm is to proceed with feature extraction based on preprocessed images and find which feature is the most important part of image recognition and which feature will be important for recognition of the target object. The difference in viewpoint is a serious problem and cannot be solved without three-dimensional modeling of the face. Although research has been actively conducted in recent years to obtain 3D information on the face so that even the face can be recognized, it is still difficult to apply it to robots with disadvantageous image acquisition conditions in that it targets very high-resolution clear images.

#### 3.1.1 Pre-processing

Pre-processing algorithms are indispensable algorithms in image processing, and how they are used has benefits such as accuracy, precision, and reduced computational time.

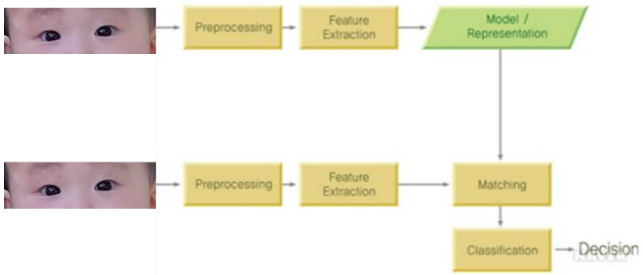


**Figure 1 :** Image Processing Procedure

In Figure 1, Pre-processing algorithms largely include grayscale, binarization, zoom, rotation/transformation, etc. After processing this preprocessing algorithm process, the main algorithm is applied.

#### 3.1.2 Pre-processing Algorithm

The grayscale serves to reduce the width of the data by converting the multi-channel image into a single-channel image. The color image is inevitably a multi-channel image. Finding a shape or shape without classifying it by color can also be detected with black and white images. Therefore, the multi-channel image is changed to a single-channel image, a gray tone image. Most algorithms process sources of images using grayscale images. Multiple channels have three or four channels. Since the grayscale has one channel, the amount of data is reduced to 1/3 or 1/4, but it does not significantly damage the shape of the image. It does not significantly affect the shape of the image or the distribution of pixels, but the amount of data is greatly reduced. Calculating major algorithms with grayscale images can greatly benefit from accuracy and computation volume.

### 3.2 Operating Scenario

The operating scenario of this system is as follows. In Figure 2, when a button for camera control is operated through a smartphone server, a signal is transmittedto the body of the ESP32-CAM through Wi-Fi communication.

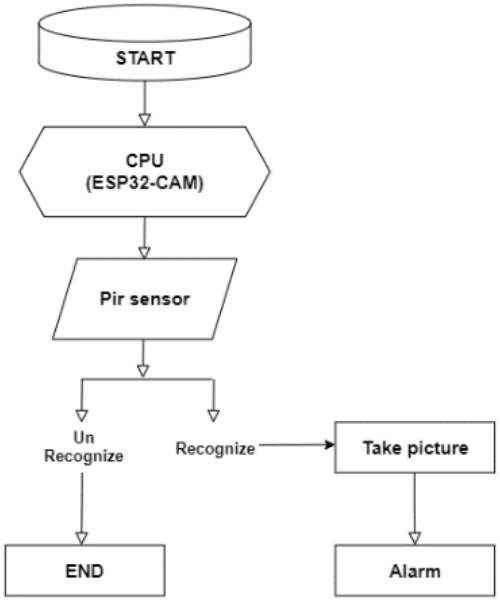


**Figure 2 :** Flow of Operationg Procedure

This signal is transmitted to Arduino to control the camera as desired. The picture taken by a camera built into the ESP32 CAM is transmitted through Wi-Fi communication, to be received in a smartphone server later to be output. Also, LEDs are controlled in the similar manner. Figure 4 displays unauthorized intrusion detection system based on ESP32-CAM, which shuts down the system if a person with a registered face comes forth the face recognition system, and triggers alarm after taking a picture if unauthorized person comes by.

#### 3.2.1 PIR (Passive Infrared Sensor) Sensor

The PIR sensor refers to infrared human body sensor and detects human movement in a certain section at an acute angle of 9 to 12 degrees through a Fresnel lens. Symptoms like masking effect, due to the nature of the sensor, can cause errors in the number of coefficients (about 15% or so), but the adopted statistical algorithm can reduce the range of errors in the long term. It can be operated for a long time without external power with very little current consumption and can be manufactured in a small size, making it easy for waterproof spinning design and operating over a year without special maintenance.

#### 3.2.2 ESP32-CAM

ESP32-CAM is an Arduino-compatible board based on ESP32 processor that supports Wi-Fi, Bluetooth and cameras. It can also be used independently using a PCB on-board antenna, and sensitivity may be expanded by separately connecting the external antenna. Using this, CCTV recorders and video streaming devices can be developed very simply and economically.

### 3.3 Unauthorized Intrusion System Implementation

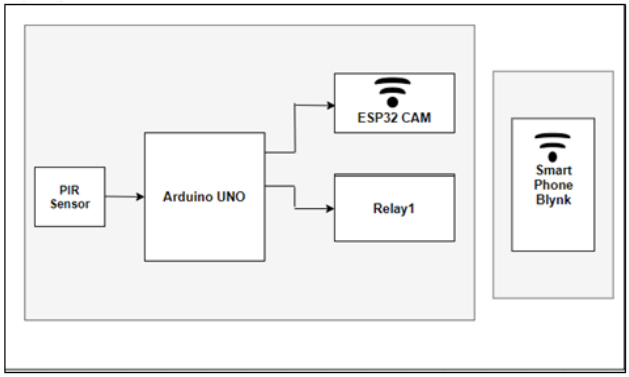


**Figure 3 :** Unauthorized Intrusion System Diagram

This system has implemented an Arduino-based unauthorized intrusion system that can greatly help prevent crimes in single-person households using the combination of PIR sensors, Arduino devices, and smartphones.

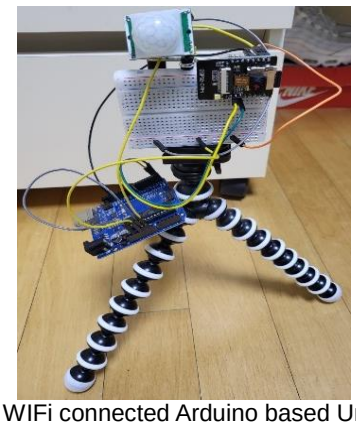

**Figure 4 :** WIFi connected Arduino based Unauthorized Intrusion System Implementation

Figure 4 is the unauthorized intrusion system, showing the connection between Arduino Uno and ESP32-CAM and with smartphone applications. Figure 6 describes Arduino's connection system to implement unauthorized intrusion system. It detects objects/people through Arduino boards, PIR sensors, and ESP32-CAM and can recognize through ESP32.

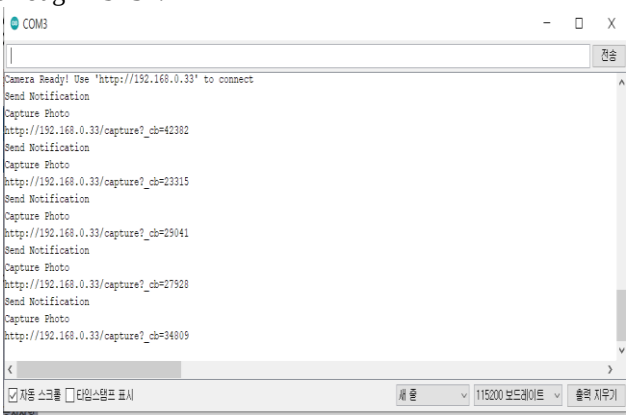


**Figure 5 :** Application Operating Display

Figure 5 shows the application implementation screen to determine whether a photo taken by ESP32-CAM through TAKE PICTURE button located at the center of the application is transmitted to the application. As a result, a control command is transmitted to the ESP32-CAM using Wi-Fi communication. Since it has been confirmed in several testing environment that motion recognition operates smoothly in accordance to the PIR sensor, it is expected to be utilized in various places.

### 3.4 System Operation and Testing

After testing connectivity of wireless connection such as the remote-control environment for ESP32-CAM operation, application is ready to operate when Auth Token is applied as a code as described in picture. At the same time, a PIR sensor starts to detect in real time, taking a picture when object is recognized.

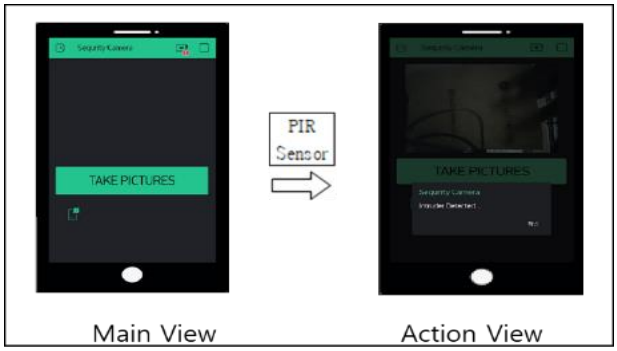


**Figure 6 :** Application Operating Display

Figure 6 is a serial monitor screen output when operating an unauthorized intrusion detection system using a PIR sensor and Arduino. When Wi-Fi based code is uploaded to ESP32-CAM, camera is prepared and provides the connected Ip link. On the serial monitor screen, the phrase ‘photo is taken’ is displayed when the photo is manually taken or systematically taken due to recognition

## 4. Conclusion

This paper studied a system based on ESP32-CAM to efficiently build security management for single-person households using Arduino and PIR sensors. ESP32 is a low-cost, low-power system of chip microcontrollers with integrated Wi-Fi and dual-mode Bluetooth. The ESP32 series uses a Tensillca Xtensa LX6 microprocessor, an Xtensa LX7 dual-core microprocessor or a single-core RISC-V microprocessor, and includes built-in antenna switches, RF balunes, power amplifiers, low noise receiving amplifiers, filters and power management modules. Therefore, by applying the Scale-invariant feature transform, it can play a role in extracting feature points to distinguish people or objects entering a specific area. Machine learning was applied to extract feature points, and through this, it is expected that it can be applied to the process of restricting access or selecting through comparison with collected and stored images. Face recognition has been one of the major research areas of computer vision over the past decades. This technology is widely used to verify whether the person in the two images entered into the system is the same or to identify who the person in the image is most similar to among the people previously stored in the internal database. Recently, if daily quarantine is underway around us and it is necessary to

verify the identity of visitors, it is expected that it will help maintain a safety net if this system is applied for the purpose of facial recognition and restricting some access. This technology is widely used to verify that the characters in the two images entered into the system are the same or to determine who the characters in the images are most similar to among those previously stored in the internal database. There is an advantage that it may be implemented in a low-power, low-cost environment through image recognition, comparison, feature point extraction, and comparison. However, since this paper is designed for the purpose of performing the desired purpose in a limited environment, there are some shortcomings in the rapid processing and accuracy of signals.

Therefore, future research tasks require sufficient data collection and storage based on deep learning, and through this, it is necessary to focus on improving the resolution, rapid processing, and accuracy of images.

## References

Gosselin, P.H., Murray, N., Jégou, H., & Perronnin. F. (2014). Revisiting the fishervector for fine-grained classification. *Pattern recognition letters*, Vol.49, 92-98.

Guo, T., Dong, J., Li, H., & Gao, Y. (2017). Simple convolutional neural network on image classification. *IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, 721-724.

Kang, H. G., Seo, D. S., Lee, B. S., & Kang, M. S. (2017). Applying CEE (CrossEntropyError) to improve performance of Q-Learning algorithm. *Korean Journal of Artificial Intelligence*, 5(1), 1-9.

Kong, Y. H., & Lee, W. C. (2017). Dynamic Obstacle Avoidance and Optimal Path Finding Algorithm for Mobile Robot Using Q-Learning. *Journal of Korean Institute of Information Technology*, 15(9), 57-62.

Li, Y., Zhang, J., Gao, P., Jiang, L., & Chen, M. (2018). Grab Cut Image Segmentation Based on Image Regi. on. *IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, 311-315,

Steela, K., Birdsong, W., & Reddy, B. Y. (2019). Image classification using Tensorflow. *16th International Conference on Information Technology-New Generations (ITNG 2019)*, 485-488.

Yoo, W. S., Seo, J. h., Kim, D. H., & Kim, K. H. (2019). Machine scheduling models based on reinforcement Learning for minimizing due date violation and setup change. *The Journal of Society for e-Business Studies*, 24(3), 19-33.

https://www.seeedstudio.com/32-CAM-Development-Board-with-camer-p-3153.html (accessed 2017)