


# Using Technology and Algorithms for Face Detection and Recognition Using Digital Image Processing and Relying on a Computer Vision Sensor

Rasha Basim Yousif Al-khafaji  <sup>1</sup>\*

<sup>1</sup> Lecturer, Department of Computer Science, College of Computer Science and Mathematics, University of Thi-Qar, Nasiriyah, Iraq

\* **Corresponding Author:** [rasha.b.alkhafaji@utq.edu.iq](mailto:rasha.b.alkhafaji@utq.edu.iq)

**Citation:** Al-khafaji, R. B. Y. (2024). Using Technology and Algorithms for Face Detection and Recognition Using Digital Image Processing and Relying on a Computer Vision Sensor. *Journal of Information Systems Engineering and Management*, 9(1), 25018. <https://doi.org/10.55267/iadt.07.14328>

## ARTICLE INFO

Received: 24 Nov 2023

Accepted: 25 Jan 2024

## ABSTRACT

Advancing in variable scopes in network technology, many new technologies were developed. Security issues were important, especially in the detection and recognition of people using variable methods like face details. Sensors have been used widely in recent days to support security systems. Sensors are devices used to convert any type of signals into electrical signals that are recorded to be processed later. These signals can be viewed by the user in several ways. Sensors increased in the development stage that can be integrated with operating systems, data storage systems, processing units, communication units, and any other function units. Detection and recognition systems were developed into a new level of technology. Some systems like figure print and palm lines face many problems because the possible change of the skin structure can be faced in time. So, these methods faced a certain problem and limitations that caused them to search for other methods more accurate. This search aims to create a new method for face detection and recognition depending on sensors. Most of the methods used for face recognition depend on OpenCV libraries that give good accuracy and time recovery availability. On the other hand, practical applications were developed to increase the accuracy of these systems like SeetaFace and YouTu methods. Three methods of detection were important to be detected too to increase the accuracy of the whole system which are the side face detection, the occlusion detection, and the face expressions. Then these data were compared to create the whole accuracy result of the system.

**Keywords:** HAAR Filter, Face Recognition, OpenCV.

## INTRODUCTION

Face detection and recognition technologies have been around for decades, but they have recently seen a surge in popularity due to advances in artificial intelligence and machine learning.

This technology is used in a variety of applications, from security monitoring to crowd detection. Face recognition is the process of recognizing a person's face in an image, while facial recognition is the process of recognizing faces from previously stored images. With the help of advanced algorithms, this technology can match a face in an image with a known face in a database—individuals can be identified in many applications, such as surveillance, passport authentication, law enforcement, etc. Advances in deep learning algorithms can also identify facial expressions by facial recognition. It can provide a more efficient, accurate, and comfortable approach.

For example, Yong and Wu (2017) built a face recognition system based on skin color, the accuracy of face recognition depends on the detection of skin color. For example, face recognition algorithms use skin color as a feature some can identify people with darker skin tones. This is because the dark skin is difficult to distinguish from other features and features in the image. This is especially true in low light conditions. To address this issue, researchers have developed systems tailored to different skin tones and capable of recognizing faces of all skin colors with high accuracy. Using a variable face recognition algorithm, it can for accurate facial recognition regardless of skin tone. The experts use additional YCbCr and HIS colour spaces to tackle this problem. The SNoW categorization filter was then applied to these samples.

Deep learning based on neural networks was used by G. Liang and Zeng (2017). Intelligent technology known as deep learning has completely changed modern facial recognition. This allows faster and more accurate facial recognition than ever before, resulting in better and more accurate recognition. Deep learning algorithms are used to recognize faces in digital images or movies. He can recognize individuals and recognize features such as faces such as lips, nose, and eyes. Additionally, deep learning algorithms can learn from data and adjust their search strategy accordingly.

This means that more data can be accurately matched, resulting in more accurate facial recognition. Access, security, and surveillance are just a few applications of deep learning in facial recognition. In addition, it allowed the application of facial recognition technology in medicine, including diagnosis and treatment of various diseases and automated video as one method of education deep is playing a role in the entertainment industry. Deep learning has improved the accuracy and reliability of face recognition and is expected to revolutionize future face recognition experiments.

Template matching technique, an important tool in facial recognition technology, was used by L. Liang, Ai, and He (1999). It allows us to identify a person by comparing faces in two pictures. First, the template is extracted from the face image and compared with the face in another image. To perform this comparison, the difference between each pixel in the template and its corresponding pixel in the other image is compared.

Two angles are considered congruent if the contrast is within a certain tolerance. Besides being fast—comparing two images takes only a few seconds—pattern matching is a robust and reliable way to recognize faces while capturing the facial characteristics of each person like the shape of his face, eyes, and nose.

To accomplish the orientation task in the search and detection condition, Mi, Chen, and Ji (2017) presented. Different algorithms like AdaBoost and Haar solved this complex problem. AdaBoost, also known as Adaptive Boosting, is a system learning rules program that optimizes the facial recognition system.

By combining more than one weak classifier, AdaBoost constructs a sturdy classifier that surpasses the accuracy of character classifiers. It achieves this using assigning weights to the vulnerable classifiers based on their accuracy, giving better weights to extra-accurate classifiers. This approach efficaciously reduces the mistake fee of the gadget and enhances the precision of face recognition. AdaBoost has been considerably employed in face popularity systems and has validated superior accuracy compared to other strategies like Support Vector Machines and Neural Networks. Besides its software face popularity, AdaBoost has also observed utility in numerous different domains, which include object detection and photo category. By leveraging AdaBoost, face recognition structures can achieve heightened accuracy and reliability, thereby improving safety and precision throughout numerous packages.

Necklace-like characters aid in facial recognition because of their ability to capture facial features important for individual recognition. Simple mathematical calculations on the matrices of pixels representing the image are the basis for crazy-looking results. The process of estimating features involves passing a narrow window, also called a kernel, over the image to calculate the sum of the differences between the pixels inside and outside the window at each iteration at each iteration. Thus this difference is used to create a histogram and used as an identification characteristic. Compared to other methods, these extraction methods are quite efficient and can handle much larger image sizes. In addition, compared to other methods, necklace-shaped features are it is more sensitive to light, scale, and style adjustments, making it suitable for facial recognition in practical applications. Furthermore, the attributes can be efficiently implemented by modern hardware such as GPUs because they are based on a simple mathematical algorithm. Tests show excellent turning and rear detection speed and latitude. It works best when used on an embedded platform.

The main objective of this research was the issue of training load initiation and slow training speed, which Wu and Zhan (2017) addressed using a different type of conventional neural network.

By using the Gabor optimizer to extract features, the network was constructed as a layered structure network and accurate generalization results were obtained for the entire system. Wang, Luo, Zhong, and Li (2018) increase the overall accuracy of the system by using a multi-faceted approach. Multi-feature clustering is a successful

method for extracting complex features from multiple sources. Combines parts from different sources by taking the treble part of one source and the bass part of another source. Compared with traditional feature extraction methods, the advantage of this method is that it can extract a larger set of features, resulting in higher performance in various applications. Other multidimensional cluster techniques can also be employed to recognize small patterns in data, such as the images of medical diseases. These methods can be obtained by combining low-level data from multiple sources. This is difficult to overcome. This is often helpful in diagnosing medical conditions since small data samples can provide insight into a patient's condition.

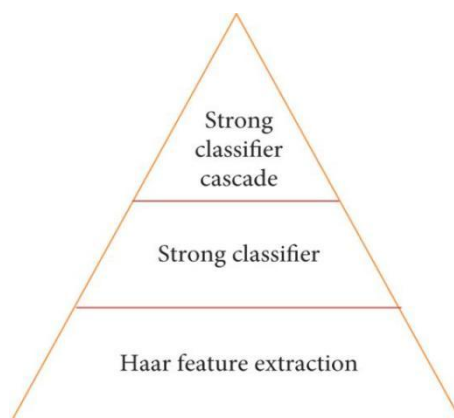
Using this technique can reduce the amount of data required for a specific task. The amount of data to be processed can be reduced by integrating features from different sources, which can save computing costs and increase speed. The LBP approach was used by Di, Li, Ma, and Gao (2017) to detect faces more accurately and with comprehensive results. This research aims to map between linear and non-linear features of the trained faces, where the feature extraction was done using the SVM algorithm (Qu, Xu, Zhao, & Zhang, 2021). These methods were good for preventing illumination effects from decreasing the prediction accuracy.

Cheng (2017) used another method that aims to decrease the effect of the illumination too, where face images could be taken in different environments. The MSR algorithm was used for this purpose to isolate different areas of object illumination. Then other methods like GF and INPS were used to extract illumination. For fusion states classification, linear discriminant analyses were used to verify the different luminance degrees of the face colors. Jing et al. (2018) and Q. Y. Li, Jiang, and Qi (2017) use a deep neural network DNN for face recognition tasks where the method gives a high-accuracy result.

## OPENCV METHOD

OpenCV is an open-source computer vision library that has been widely used in a number of applications. It provides multiple modes of detection, video analysis, and image recognition. Face recognition is one of the most widely used functions of OpenCV. Taking a source image or video frame extracting its facial features and comparing it with a library of hidden faces is an OpenCV face recognition technique comparison. SVM uses several methods such as Principal Component Analysis (PCA), and support vector machines. Local Binary Pattern (LBP) and Histogram of Oriented Gradient (HOG). To optimize image data, OpenCV face recognition uses preprocessing and feature extraction to extract facial features, followed by a classification phase to classify the image data. The captured facial features are sent to a classification system and compared against a stored facial database. Once a match is found, the OpenCV facial recognition system displays the person's name in the database. OpenCV's face recognition method can detect subtle differences in facial features with high accuracy.

In summary, OpenCV's facial recognition techniques allow you to identify people from digital images and videos with high accuracy and confidence. It is an effective tool that can be used in a variety of computer vision applications. The OpenCV library is suitable for facial recognition projects due to its many features. **Figure 1** shows how the facial recognition model works. The necklace classifier is a popular signal extraction tool. In this feature extractor, facial symbols are represented by multiple rectangles of different sizes. The pixels in the image define the values of the necklace distribution, and these squares are it Painted in various shades of black and white.



**Figure 1.** Human Face Detection System Steps

The formula defines the rectangles used for the features is defined as (1):

$$\pi_{[w,h]}^m = \left( \binom{m}{w} + \binom{m-1}{w} + \dots + \binom{w+1}{w} + 1 \right) \cdot \left( \binom{m}{h} + \binom{m-1}{h} + \dots + \binom{h+1}{h} + 1 \right) \quad (1)$$

where  $w$  and  $h$  mean the width and height of the rectangle, respectively. Due to the many features derived from this equation, OpenCV uses a specialized approach to streamline transactions. For example, the value of each pixel in an image  $x$  and its integral version can be determined using the following formula (2).

$$g(x, y) = \sum_{x' \leq x, y' \leq y} f(x', y') \quad (2)$$

And this pixel is calculated using the equation (3) (4):

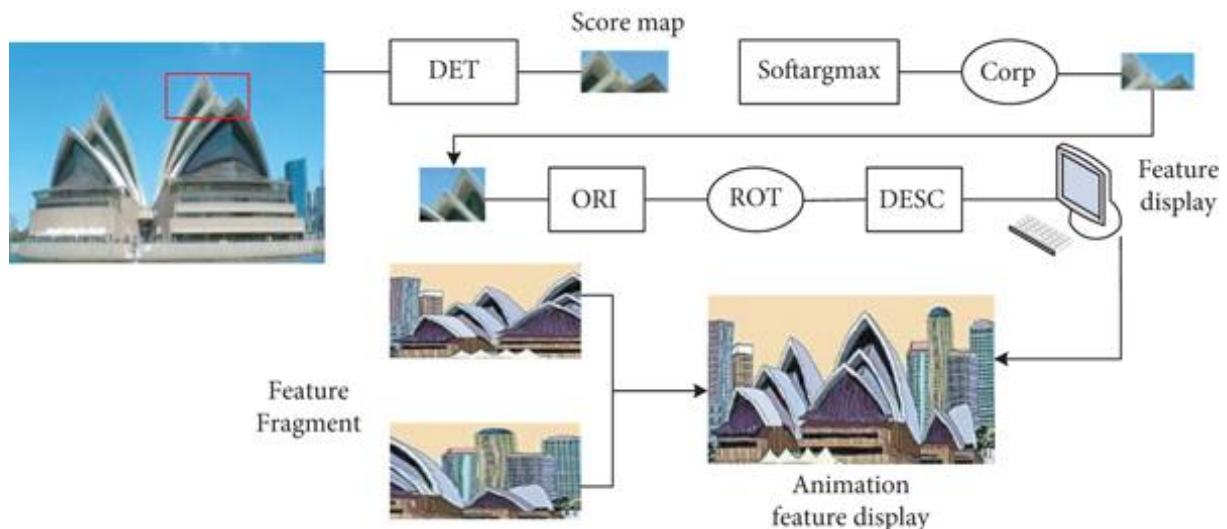
$$s(x, y) = s(x, y - 1) + f(x, y) \quad (3)$$

$$g(x, y) = g(x - 1, y) + s(x, y) \quad (4)$$

The values resulting from a cumulative operation for each row in which the initial value is = to zero are denoted by the symbols  $s(x, y)$ .

## SEETAFACE METHOD

A vital advance in facial recognition technology is the SeetaFace method. Deep learning is a computer vision technique used to recognize faces in images and videos. A deep changing neural network trained to recognize facial features and distinguish between different people is the basis of the technology with 99.8% accuracy. The SeetaFace method can identify a person from a single image, proving more accurate than traditional face recognition techniques. It can also identify individuals. Even with the ability to recognize faces in real time, the delay is less than a second. This technology can be employed to identify individuals in multiple ways, including security, monitoring, access control, and more. **Figure 2** illustrates the primary processes.

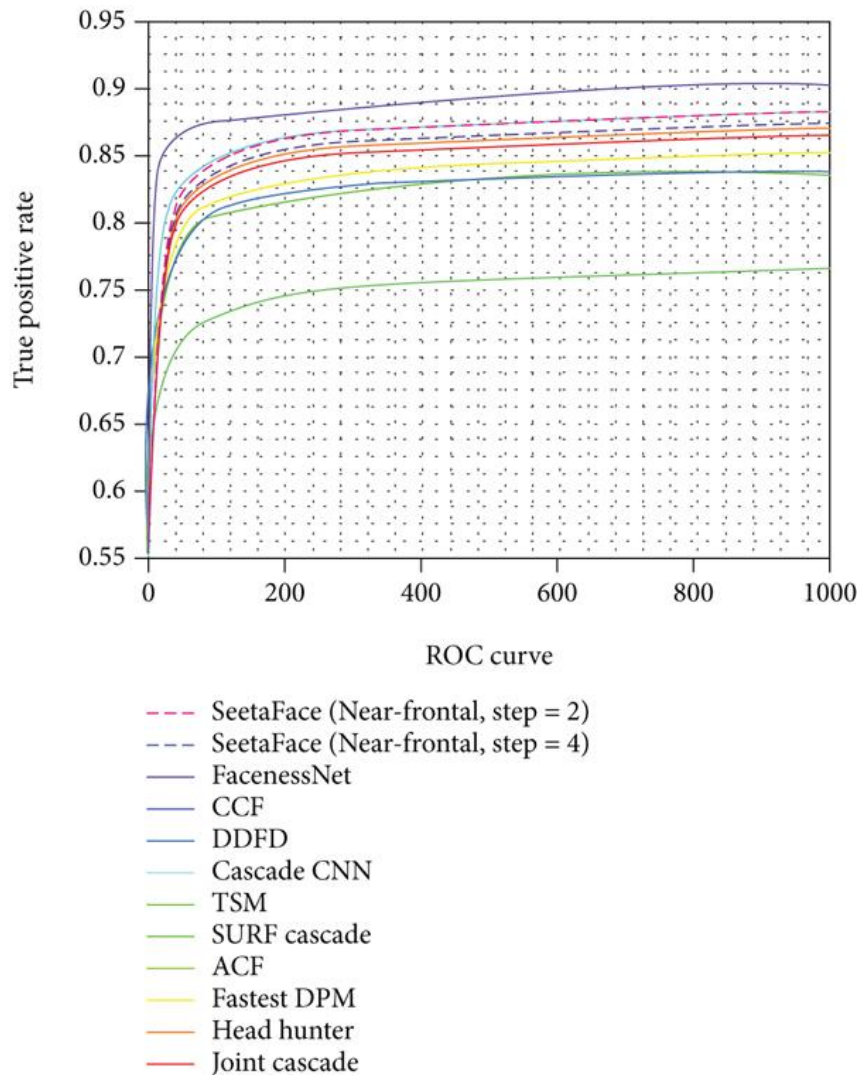


**Figure 2.** The Method of SeetaFace's Structure

The SeetaFace path algorithm is a powerful form of facial recognition that employs deep learning and convolutional neural network (CNN) to differentiate and categorize faces in digital images and video. The process involves three steps. Facial traits are based on existing faces to start this procedure. The 2D segment involves using deep learning to pick out faces from a present database of previously diagnosed faces. A category approach is in the long run used to classify the discovered face characteristics inside the 0.33 layer. SeetaFace Roadways' structure is tremendous for facial recognition programs inclusive of security cameras, add-ons, and video recording. Moreover, it miles used in a software program that identifies emotions to perform responsibilities like gender and age prediction. This process is critical for the most unique and dependable face popularity machine currently to be had. The SeetaFace era can discover faces in low-light conditions and is flexible in its use. This gadget is robust and proof against being overturned or challenged. The ROC curve for the SeetaFace system showcases its accuracy in detecting by illustrating the false alarm charge and real alarm fee. The ROC curve is

constructed by using plotting the True Positive Rate (TPR) at the Y-axis in opposition to the False Positive Rate (FPR) at the X-axis. TPR is calculated by dividing the whole quantity of actual wonderful occurrences by way of the entire number of faces in the dataset. Both the Fake Positive Rate (FPR) and False Negative Rate (FNR) of the entire model are taken into consideration.

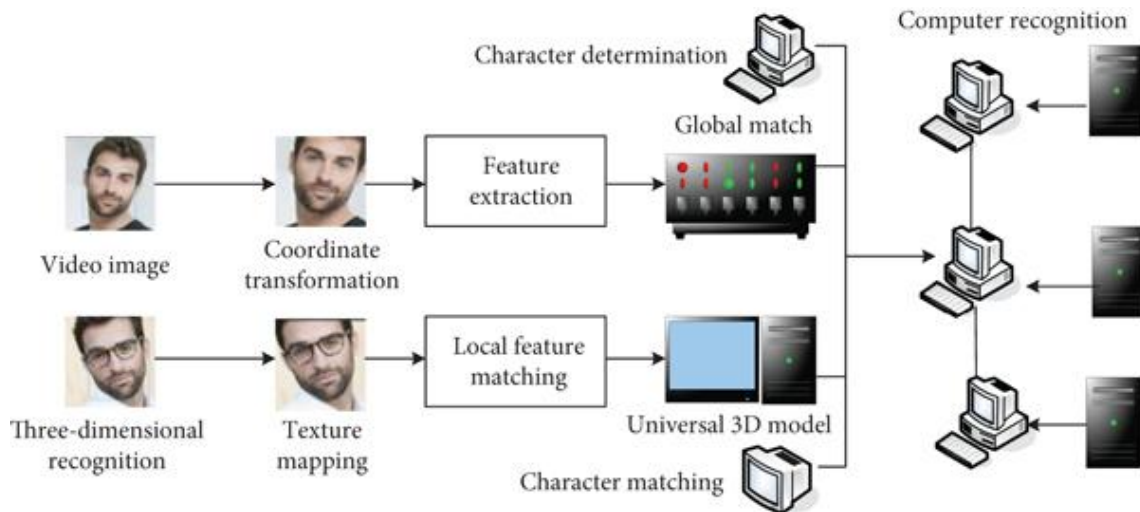
The technique's usual efficacy is quantified via the region underneath the ROC curve, also known as AUC. An AUC of 1.0 shows an ideal model, at the same time as an AUC of 0.5 indicates a version that is no higher than a random measurement. The SeetaFace detection machine has an AUC of 0.988, making it one of the most accurate face recognition systems. **Figure 3** displays the ROC curve for numerous facial reputation structures.



**Figure 3.** SeetaFace Method ROC Curve Compared with Other Methods

The surface extraction method from images provides the premise for the Scale-Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF) capabilities used by SeetaFace. SeetaFace uses this SURF attribute to fast and appropriately recognize faces in photographs. The programme employs a method referred to as "face alignment" to assure the precision of facial popularity. The programme recognizes facial traits in the picture and matches them with the reference image during the face insertion procedure. This enhances the precision of the facial recognition technology. Additionally, SeetaFace recognizes facial markings in a photo that can be used to confirm identity. This helps with the safety of the face recognition. Furthermore, SeetaFace offers sufficient facial recognition capability to recognize faces in different lighting conditions, from different perspectives, even in low light conditions thus contributing to the safety and reliability of the facial recognition system. **Figure 4** illustrates the basic steps of face recognition and recognition.



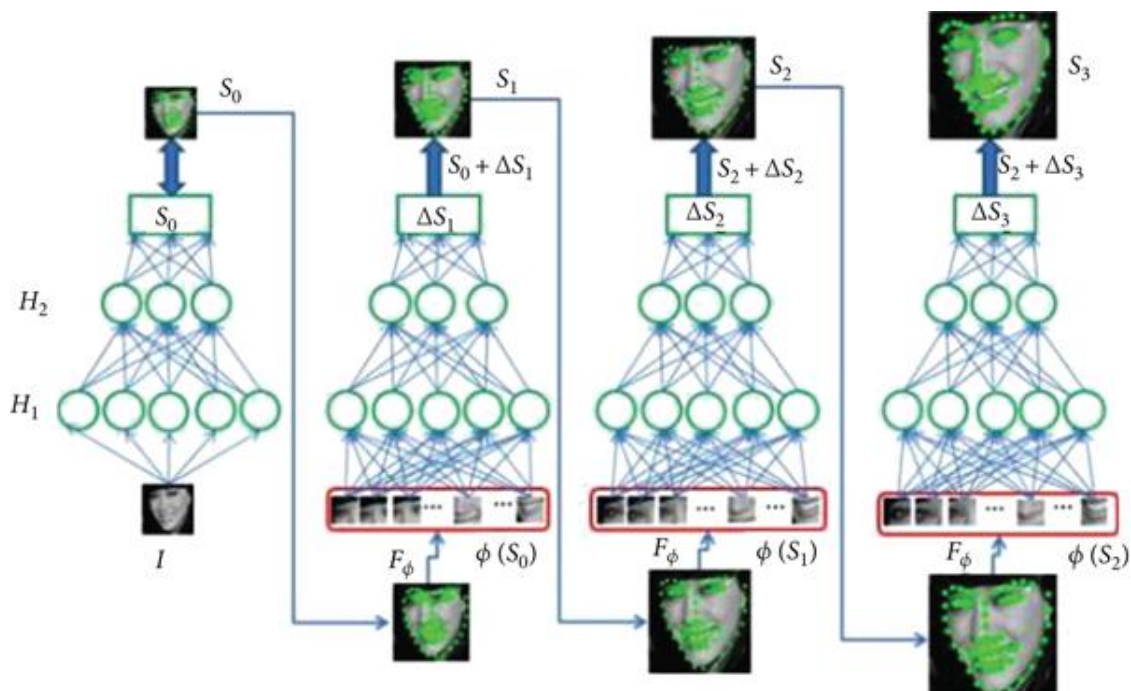


**Figure 4.** SeetaFace Method for Face Detection

Convolutional feature-attention network (CFAN) are the method SeetaTech utilises to recognize faces in digital photos. A deep learning method called CFAN is based on convolutional neural network. Using a sequence of convolutional layers, the technique finds edges, forms, textures, and other elements in a picture. After the convolutional layers and the final CFAN model are used to assess whether a face is present in the image, there is a concept-based convolutional layer that is intended to identify and concentrate on significant features in the image.

The purpose of the CFAN properties is to endure variations in facial features like lighting, expression, and posture. It can also identify faces in photos with varying qualities, from low to high definition. The CFAN model's processing efficiency makes it suitable for embedded systems and mobile devices. Lastly, because the CFAN model can identify faces in photos with notable orientation changes, it is useful for applications like video surveillance.

**Figure 5** shows the multi-step process neural networks use for the detection and recognition of faces.



**Figure 5.** Neural Networks Method for Face Detection and Recognition

**Figure 6** lists the 68 significant points that are employed for facial recognition.



**Figure 6.** The 68 Main Landmarks for Face Detection and Recognition

## YOUTU METHOD

The latest technological breakthrough that completely varies how people perceive and recognize faces in photos and videos is the YouTu platform for facial recognition. It is a deep learning-based algorithm developed by YouTu Lab, Beijing, a Chinese subsidiary of YouTu Technologies. CNNs are the tools used in the YouTu method to accurately detect and recognize faces in images and movies. It is able to recognize facial features with far greater accuracy and efficiency compared to traditional methods. The YouTu method is suitable for a wide range of applications because it can recognize faces from different angles and lighting conditions. Face recognition, face authentication, facial recognition, facial recognition, emotion recognition, and 3D face reconstruction are just a few of the functions you have implemented. Several security screening applications including facial recognition in busy environments such as train stations and airports have also used the YouTu channel. The YouTu channel is a powerful tool for facial recognition and recognition because of its proven accuracy greater than traditional methods. The whole process is described by the following function (5).

$$f(x) = \sum_{t=1}^T \gamma_t w_t(x) \quad (5)$$

Where  $\gamma$  represents the correlation coefficient for every group member. Throughout the boosting process,  $\mu$  and  $w$  are learned. Equation (6) can be used in this method to determine the posterior probability of a certain hidden topic E given the current parameters:

$$E_{r \rightarrow G} = \min_{G \in G} |R - G| + \beta \quad (6)$$

Equation (7) can be used to define an estimated hidden topic R, from which we can maximize the full likelihood function.

$$L(d) = \frac{1}{|M|} \sum_{r \in R} [E_{r \rightarrow G} < D] \quad (7)$$

We can use the Lagrangian coefficient function, which is defined by the equation (8), to calculate the highest likely parameters.

$$T_{r \rightarrow G} = \min_{r \in R} |M - Y| \quad (8)$$

## IMAGE SENSOR

When it comes to the facial recognition approach, image recognition is important because it allows the system to recognize faces in different environments. Image sensors can be used to detect faces in movies or still images.

Typically, a face detection system combines software algorithms with hardware, like an image sensor, to identify faces in a scene. After the scene is captured by the image sensor, the image is processed by a software algorithm to determine whether a face is present. CMOS or CCD image sensors can be used, depending on the requirements of the application.

The software programme determines whether a face is present in the pixel data by searching for specific features such as the existence of eyes, noses, and mouths. The image sensor records the scene as an array of pixels. After that, the software system determines the face's identity by contrasting its features with a database of recognized faces. The system can then do additional actions, such as granting entry to a secure location or starting a response, after identifying the face.

1. This project seeks to meet the needs of two primary markets: before a surveillance camera is deployed in an actual setting, its image sensors can be used to estimate the camera's accuracy in face recognition from a distance.
2. Cost may be the primary consideration when local governments, businesses, etc. decide to install a series of cameras in various locations, thus we must weigh the pros and cons of the cameras.

## PRACTICAL TASK

This study employed several techniques for face detection and recognition, including:

### HAAR Cascade

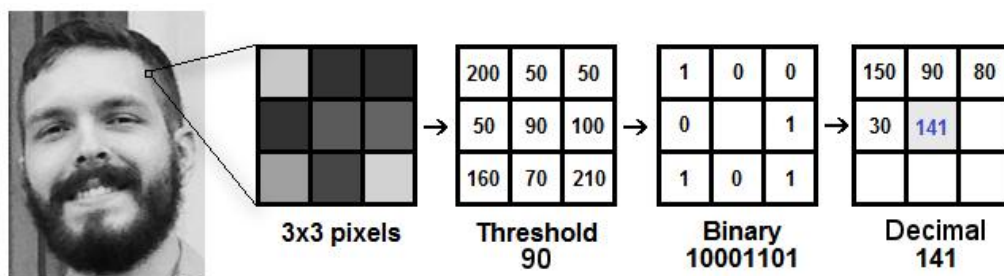
In computer vision, the machine learning method Haar Cascade is used to recognize objects. It was initially put forth by Paul Viola and Michael Jones in 2001, and its foundation is the idea of a "cascade" of classifiers that are classified as "weak". This shows that identifying objects in images is a decision made at multiple levels. The method starts with the input of an image, and it terminates with the collection of objects used to create the "Haar Cascade Classifier". The classifier is used to categorize objects inside the photograph by distinguishing between the item type and the classifier type.

Segmenting droplets allows you to alter the characteristics of the detected items, resulting in an extra unique outcome. The Haar cascade is a broadly used approach for item identification that serves several capabilities, inclusive of figuring out faces and analyzing films. The constancy of this approach is similar to different advanced strategies, making it suitable for detecting gadgets in problematic photos. Moreover, the technique is characterized with the aid of a modest computational time value, making it appropriate for a wide variety of eventualities. The software generates a classifier for each consecutive window of movement because it progresses via the photographs. Success is finished when the recognizer detects an object within the window. Otherwise, the outcome can be unfavorable.

The algorithm repeats this system till it reaches the subsequent window. Once the scan finishes, the algorithm gives a listing of items detected within the photograph. The Haar Cascade set of rules' entire technique is as follows:

1. Feature extraction entails identifying and keeping apart certain traits such as borders, stripes, and forms from pics.
2. Training: Utilizing the waterfall technique to educate susceptible people through visuals labeled as positive or poor.
3. Detection involves identifying gadgets in snap shots and the usage of a sequence of weak classifiers that result in a cascade.
4. Post-processing: Enhance the detection machine's findings by minimizing fake positives and false negatives.
5. Returns: A listing of things which can be recognized inside the photograph applications.

**Figure 7** illustrates the Local Binary Pattern Histogram (LPBH) technique of spotting faces.

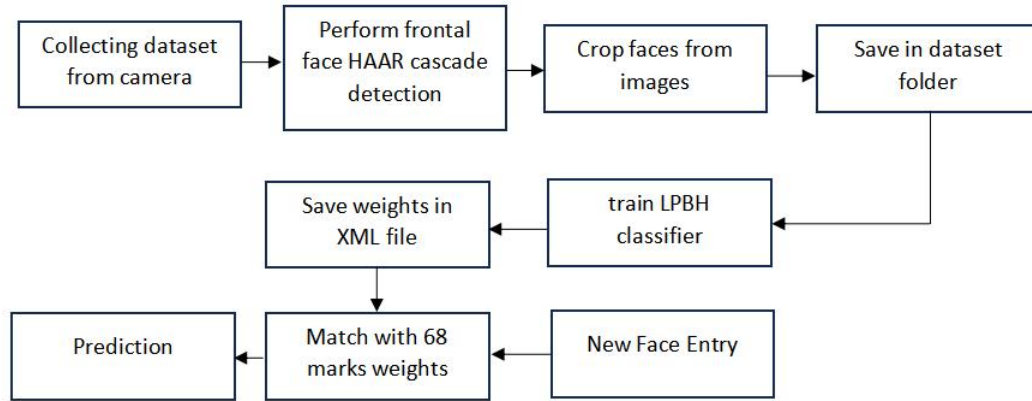


**Figure 7.** LPBH Technique of Spotting Faces

The technique is as outlined: The algorithm first identifies the faces in the image by way of an evaluation of the photo. This suggests that each face inside the photograph has been acknowledged. Subsequently, generate a nearby binary sample desk using the segments. Face detection is finished by the usage of a function set called



Local Binary Pattern (LBP). It is executed by evaluating the luminance of person pixels with that of neighbouring pixels. Once the Local Binary Pattern Histogram (LBPH) is created, the algorithm compares it with a preexisting database of LBPH. If a match is found, the algorithm will provide the identification of the individual in the picture. If the database is empty, the algorithm will repeatedly generate a new entry in the database containing the LBPs histogram of the new face. Finally, the algorithm will store the new LBPH in the database. This will allow the algorithm to identify the person in the image the next time it is encountered. **Figure 8** shows a flowchart for the proposed method:



**Figure 8.** Flow Chart of Proposed System

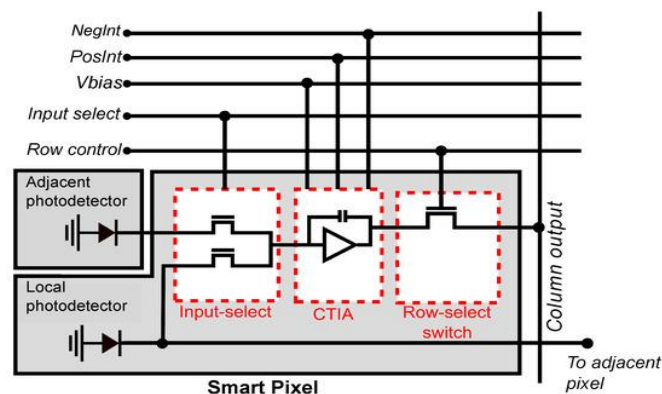
The system was trained on 15 persons faces each of them with 200 several positions images. The accuracy of the system prediction was 97% which was high compared with neural network methods that require more processing power and more training time.

This method shows the powerful performance of OpenCV libraries and their ability to work in recognition and prediction for face surveillance systems. The system has advanced to the point where it can identify the gender and emotion of the person in the picture.

### Smart Pixels in Image Sensors

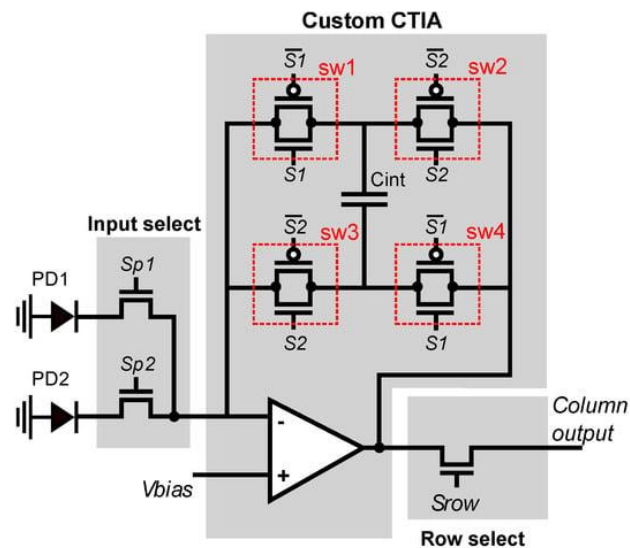
The newest technological advancement that is completely changing how we see our surroundings is called smart pixels. These tiny image sensors, which can swiftly and precisely capture and interpret images, are finding their way into an expanding number of devices, including drones and smartphones. A range of tiny sensors, such as light and image sensors, are used by smart pixels to sense and interpret their surroundings. Because of this, they can take precise and detailed pictures even in the darkest settings. Smart pixels are extremely helpful for security and surveillance applications since they can recognize faces, objects, and motions.

They are also being utilized in a wide range of artistic endeavours, including interactive art installations and virtual reality experiences. The way we interact with the world is being altered by smart pixels, which will only proliferate in the next years. The basic architecture of a smart pixel in an image sensor is depicted in **Figure 9**.



**Figure 9.** Structure of Smart Pixel

Input selection works as a switch to select the input, while the CTIA works as an enhancement for low light conditions as in IR cameras for example, compared with normal CCD circuits this circuit can perform more amplification for the voltage in a wide range. It works as an op-amp for the input signal. The row-select switch defines the method of image scanning. **Figure 10** shows the structure of the CTIA circuit.



**Figure 10.** Structure of CTIA Circuit

Although it occupies a larger area compared to alternative pixel circuits, the utilization of a CTIA for photocurrent integration is a favored technique in low-light environments and IR cameras. This preference stems from its low input impedance, which enables efficient injection with weak photodiode currents. Specifically, our interest lies in employing the smart pixel for facial recognition in thermal IR video. Furthermore, in comparison to other pixel circuits, a CTIA offers a broad linear output voltage range, minimal frame-to-frame lag, and reduced noise through enhanced control of the photodiode bias.

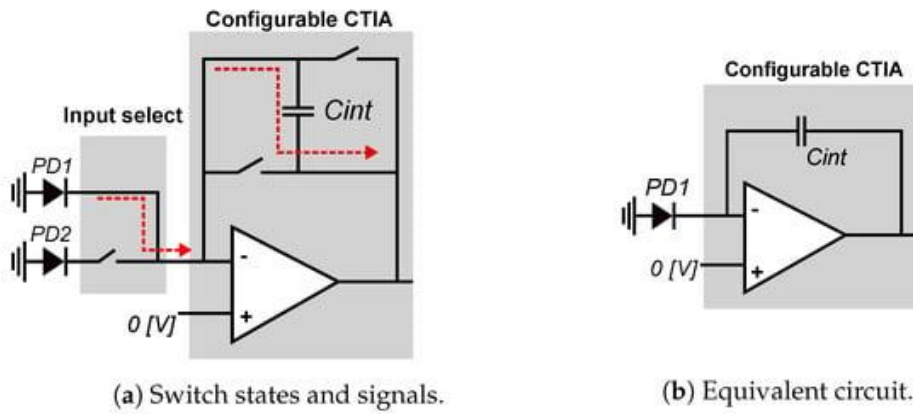
It integrates the input current to generate an output voltage, and a set of four switches, implemented as conventional CMOS transmission gates, can manipulate the orientation of the integration capacitor. The input current originates from the photodetectors in the local or neighboring pixel.

**Figure 10** depicts the CTIA functioning in its conventional mode. Throughout the integration period, the input-control switch establishes a connection between the CTIA input and the local photodetector PD1. The CTIA is set up in direct mode, with sw1 and sw4 in a closed state, and sw2 and sw3 in an open state. The equivalent circuit illustrated in **Figure 10** demonstrates the CTIA operating as a conventional integrator. Equation (10) outlines the computation of the output voltage, which signifies the local pixel value.

$$V = I \cdot \Delta t / C_{int} \quad (10)$$

The solution can be derived by considering the output voltage ( $V$ ), the input current from photodetector PD1 ( $I$ ), the integration time ( $\Delta t$ ), and the capacitance value ( $C_{int}$ ).

**Figure 11** illustrates the smart pixel's configuration for computing local horizontal gradients. The CTIA's global bias input is adjusted to the midpoint between the rails, which corresponds to 1.65 V for a 3.3 V supply voltage. The integration time is divided into two equal phases: direct and inverse. In the direct phase, depicted in **Figure 11 (a)**, the CTIA operates conventionally by integrating the current from the local PD1 detector, starting from 1.65 V. Conversely, in the negative phase [**Figure 11 (b)**], the input switches select the current from the neighboring pixel PD2. Specifically, sw1 and sw4 are open, while sw2 and sw3 are closed. Consequently, during the inverse phase, the CTIA integrates the negative current value of the PD2 photodetector. The output voltage at the conclusion of the integration period is calculated using Equation (11).



**Figure 11.** Smart Pixel's Configuration

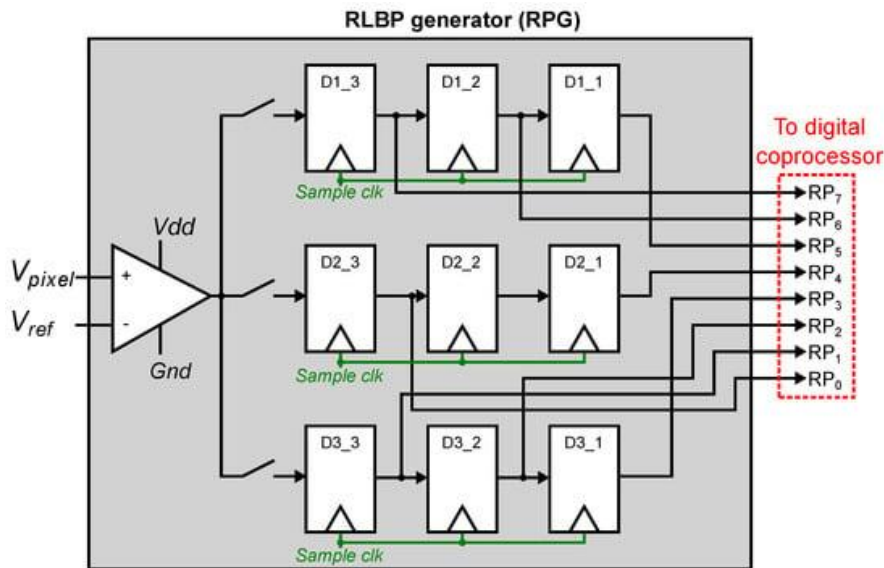
In the conventional mode of smart pixel operation, the current integration process is facilitated by the closure of input-select switches PD1, sw1, and sw4. Conversely, switches sw2 and sw3 remain open during this process.

$$V = (I_1\Delta t_s + I_2\Delta t_s)/2C_{ini} \tag{11}$$

The output voltage, V, is determined by the input current from the local detector PD1, I<sub>1</sub>, and the current from the adjacent detector PD2, I<sub>2</sub>. Additionally, the integration time, Δt<sub>s</sub>, and the capacitance, C<sub>int</sub>, also play a role in determining the output voltage.

**RLBP Generator**

The topology of the RPG circuit is illustrated in **Figure 12**. An input opamp is utilized to compare the readout value V<sub>pixel</sub>, representing the difference between two adjacent pixels, with a global reference voltage V<sub>ref</sub>. When V<sub>pixel</sub> > V<sub>ref</sub>, the comparator's digital output is 1, while in other cases, it is 0. The RLBP is then produced by writing the comparator's output into a 3×3 array of flip-flops set up as three shift registers.



**Figure 12.** Topology of the RPG Circuit

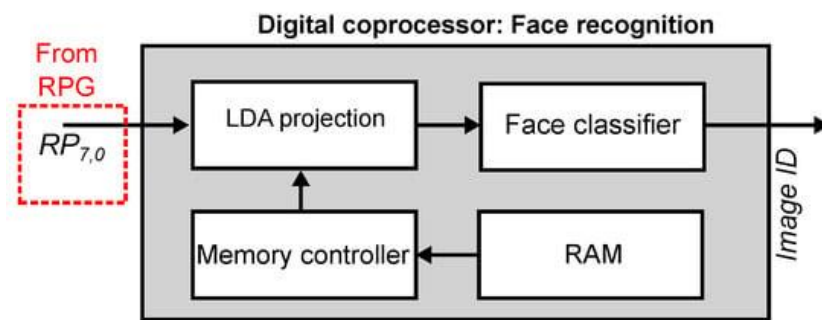
The architecture of the RPG includes an input comparator that evaluates each pixel's local gradient value in relation to a reference voltage. Three shift registers are created by arranging an array of 3×3 flip-flops to store the outputs of the digital comparator in a sequential manner. All in all, the RPG generates an 8-bit RLBP, excluding the flip-flop in the middle.

The RPG uses a row-by-row read of the FPA to calculate the RLBP in each zone. Each pixel value is read by the RPG together with its two vertically adjacent neighbours. The comparator result is then fed into flip-flops (D1\_3, D2\_3, and D3\_3). The next three pixels are read from the FPA after the register array shifts to the right.

The RLPB for the central pixel is stored in the array after nine readings, and the digital coprocessor uses it to compute the histogram. The RLPB computation for adjacent pixels overlaps due to the  $3 \times 3$ -pixel windows used, and the next RLPB is completed after 3 reads. This process continues until all pixel values in the region have been read, and the RPG moves to the next region in the image. The RLPB computation requires only a  $3 \times 3$ -bit array because the FPA directly outputs the local pixel differences, eliminating the need for large line buffers. Although each RLPB requires 3 reads from the FPA, these reads complete significantly faster than in conventional mode because they are only used for a 1-bit comparison instead of a complete analog-to-digital conversion.

### Digital Coprocessor

The architecture of the face recognition coprocessor, as depicted in **Figure 13**, involves the digital coprocessor responsible for computing the histograms of RLPBs from the image, normalizing and centering the data, projecting the resulting histogram vector using LDA, computing the Euclidean distance between the projected vector and a stored database of known faces, and selecting a label for the input image using the nearest neighbor criterion. The coprocessor receives the 8-bit RLPB vector  $RP_{7,0}$  from the RPG module as input. The LDA coefficients are read from external RAM by the memory controller and sent to the LDA projection module. The module computes the histogram vector and projects it using the LDA coefficients, which are fused into a single step to save memory and arithmetic resources. The output of the LDA module is the feature vector of the input image projected onto the LDA subspace. The face recognition module computes the Euclidean distance between this vector and a set of stored vectors that represent the known faces. The module selects the minimal distance and compares it against a chosen threshold. When the distance is smaller than the threshold, the module outputs the ID of the selected known face. Otherwise, it outputs a null value.



**Figure 13.** Architecture of the Face Recognition Coprocessor

The digital coprocessor's architecture involves the reception of a continuous stream of RLPBs from the RPG, while simultaneously constructing the histogram vector and projecting it through LDA. The LDA coefficients are retrieved from RAM by a memory controller. The nearest neighbor criterion is utilized for classification by computing the Euclidean distance between the projected vector and the contents of the database of stored faces.

## RESULTS

The system was built on an FPGA Altera DE2, in our study we try to rebuild the sensor structure. Using databases of images for the recognition task. A poly1-poly2 capacitor with a capacitance per area of  $845 \text{ aF}/\mu\text{m}^2$  was utilized for integration. The pixel necessitates a  $74 \text{ fF}$  integration capacitor measuring  $6.6 \mu\text{m} \times 3.6 \mu\text{m}$ , assuming an integration time of  $30 \mu\text{s}$  and a maximum photodetector current of  $9 \text{ nA}$ . The overall dimensions of the circuit, encompassing all passive and active elements, are  $20 \mu\text{m} \times 16.5 \mu\text{m}$ . With reference to a standard  $32 \mu\text{m} \times 32 \mu\text{m}$  pixel, the circuit attains a fill factor of 41%. The inclusion of additional transistors to compute local gradients augments the circuit's area by 30%. In the absence of the switches employed for smart mode operation, the fill factor amounts to 56.8%.

In **Table 1**, a comparison is made between our smart pixels and other designs. Despite the fact that the smart pixel described in Lee, S. Park, S. Y. Park, Cho, and Yoon (2017) utilize a CTIA only at the column level to compute local differences for edge detection, it has a fill factor of only 19%, which is significantly lower than the 56% achieved by our solution in a similar CMOS process. The other designs presented in the table employ a simpler 2-transistor integrator, which is only suitable for capturing images in the visible spectrum. Nonetheless, our design outperforms all other designs in terms of fill factor.

**Table 1.** Face Recognition Using Image Sensor Compare with Previous Studies

SIS	Pixel Pitch $\mu\text{m} \times \mu\text{m}$	Fill Factor	Tested Spectrum	Type of Integrator
Proposed Face Recognition	32×32	56%	Visible Thermal IR NIR	CTIA
Edge Detection (Lee et al., 2017)	31×31	19%	Visible	2T-Integrator, CTIA at column level
LBP Edge Detection (Zhong, Yu, Bermak, Tsui, & Law, 2018)	7.9×7.9	55%	Visible	2T-Integrator
Spatial Contrast LBP (Berkovich, Lecca, Gasparini, Abshire, & Gottardi, 2015)	26×26	23%	Visible	2T-Integrator
4-Neighbour LBP (Gottardi & Lecca, 2019)	64×64	15%	Visible	2T-Integrator

**Table 1** shows the total accuracy of the system proposed using ALTERA FPGA with a MATLAB HDL code structure for Yale Face Database which contains 38 persons each with 64 images of the size 192X168.

The accuracy of the system in the classification of images reaches 94% using the image sensors on the ALTERA FPGA where the images were split into 80% for training and 20% for testing tasks. Results show the lower parts are used the more pixel detection and recognition could be performed.

## DISCUSSION AND CONCLUSION

Face detection systems are becoming increasingly popular in our society. They are used for a variety of purposes, such as security, surveillance, and even facial recognition for identification and authentication. Face detection systems are able to detect faces in digital images or videos, and in some cases, they can even identify the person. The technology works by analyzing the features of a face, such as the eyes, nose, and mouth, and then comparing those features to a database of known faces. The system then produces a score that can be used to determine whether the face belongs to a particular person.

The system proposed exhibits low power consumption and area utilization, rendering it appropriate for portable systems and mobile devices. Despite the CTIA integrator's larger size in the smart pixel compared to alternative readout circuits, it is still suitable for low-light imagers and IR. By computing local differences during photocurrent integration, the circuit area and fill factor's impact is minimized, even though halving the integration time may reduce the image sensor's signal-to-noise ratio in face recognition mode. Our algorithm outperforms other methods in the literature when classifying faces using different databases, except when there are significant variations in illumination between the training and test data sets. These variations are considerably smaller in IR images, for which our smart pixel is designed. The results also indicate that our proposed RLBP, replacing conventional LBP, still captures sufficient texture information to perform face classification with a minor decrease in accuracy.



## REFERENCES

- Berkovich, A., Lecca, M., Gasparini, L., Abshire, P. A., & Gottardi, M. (2015). A 30  $\mu\text{w}$  30 fps 110 $\times$  110 pixels vision sensor embedding local binary patterns. *IEEE Journal of Solid-State Circuits*, 50(9), 2138-2148.
- Cheng, Y. (2017). Complex illumination face recognition based on multi-feature fusion. *Computer Engineering and Applications*, 53(14), 39-44.
- Di, S., Li, Y., Ma, M., & Gao, S. (2017). Lens detection based on LBP and SVM. *Computer Technology and Development*, (9), 44-47.
- Gottardi, M., & Lecca, M. (2019). A 64 x 64 pixel vision sensor for local binary pattern computation. *IEEE Transactions on Circuits and Systems I, Regular Papers*, 66(5), 1831-1839.
- Jing, C., Song, T., Zhuang, L., Liu, G., Wang, L., & Liu, K. (2018). Overview of face recognition technology based on deep convolutional neural network. *Computer Applications and Software*, 35, 223-231.
- Lee, K., Park, S., Park, S. Y., Cho, J., & Yoon, E. (2017, June). A 272.49 pJ/pixel CMOS image sensor with embedded object detection and bio-inspired 2D optic flow generation for nano-air-vehicle navigation. In *2017 Symposium on VLSI Circuits* (pp. C294-C295). Piscataway, NJ: IEEE.
- Li, Q. Y., Jiang, J. G., & Qi, M. B. (2017). Face recognition algorithm based on improved deep networks. *Chinese Journal of Electronics*, 45(3), 619-625.
- Liang, G., & Zeng, H. (2017). Design of intelligent video surveillance face detection system based on ARM. *Journal of Computer Applications*, 37(a02), 301-305.
- Liang, L., Ai, H., & He, K. (1999). Multi-template-matching based single face detection. *Journal of Image and Graphics*, (10), 825-830.
- Mi, Y., Chen, D., & Ji, P. (2017). Face detection algorithm based on geometric features and new Haar features. *Transducer and Microsystem Technologies*, (2), 154-157.
- Qu, S., Xu, W., Zhao, J., & Zhang, H. (2021). Design and implementation of a fast sliding-mode speed controller with disturbance compensation for SPMSM system. *IEEE Transactions on Transportation Electrification*, 7(4), 2611-2622.
- Wang, C., Luo, Z., Zhong, Z., & Li, S. (2018). A face detection method with multi-layer feature fusion. *CAAI Transactions on Intelligent Systems*, (1), 138-146.
- Wu, S., & Zhan, Y. (2017). Face detection based on selective search and convolutional neural networks. *Application Research of Computers*, (9), 2854-2857.
- Yong, D., & Wu, Y. (2017). Face detection method based on double skin model and improved SNoW algorithm. *Computer Applications and Software*, 34(5), 135-140.
- Zhong, X., Yu, Q., Bermak, A., Tsui, C. Y., & Law, M. K. (2018, June). A 2pJ/pixel/direction MIMO processing based CMOS image sensor for omnidirectional local binary pattern extraction and edge detection. In *2018 IEEE Symposium on VLSI Circuits* (pp. 247-248). Piscataway, NJ: IEEE.