A Face Recognition System Based on Cloud Computing and AI Edge for IOT

Junjie Zeng[1,2(✉)], Cheng Li[1,2], and Liang-Jie Zhang[1,2]

[1] National Engineering Research Center for Supporting Software of Enterprise Internet Services,
Chengdu, China
junjiezeng87@gmail.com
[2] Kingdee Research, Kingdee International Software Group Company Limited, Shenzhen, China

Abstract. With the demand for interconnection of all things, more and more kinds of sensors are connected to the Internet of Things. Different from traditional sensors, such as low transmission frequency and small data volume, visual sensors have the characteristics of high transmission rate and large data volume. Vision sensors are widely used in security, health care and other face recognition. This paper proposes a combination of edge-based artificial intelligence and cloud computing that is suitable for areas such as face recognition and security that require a large number of visual sensors and image processing and analysis. In order to verify the effectiveness of the technical framework proposed in this paper, a complete demonstration system was built at the end of the paper based on the rk3288 and cloud server to prove the excellence of the system described in this paper.

Keywords: Face recognition system · AI Edge · IOT · Cloud computing

1 Introduction

At present, the Internet of Things is growing faster and faster. More and more types and quantities of IoT devices will be connected to the network. IoT can be viewed as a global infrastructure for the information society [1, 2]. Because of the numerous opportunities that IoT provides, the number of connected devices is increasing rapidly, and International Data Corporation (IDC) predicted that number to reach 29 billion by 2020 [3–5].

Among these growing IoT devices, in addition to traditional low-speed, small-data-rate sensor devices, some new types of IoT devices are gradually increasing, and vision sensor devices are among the more and more concerned types. The vision sensor is a general term for a series of image input devices, and the data it acquires are mostly still photos or dynamic videos. Benefiting from the development since they have abilities to recognize a person in the incorrect area and at the false hour because this person may be a bad person for the environment [12].

The system based on vision sensor is divided into two kinds of systems based on cloud processing and local processing according to the different ways of image processing. In a cloud-based system, the vision sensor collects the image data and

uploads it to the cloud service center for processing. In the local processing system, the image data obtained by the vision sensor will be processed in the local hardware system.

As shown in Fig. 1, the cloud-based face recognition system, the vision sensor uploads the obtained image data and waits for the cloud center to perform comparative analysis according to the image data. Its advantages are a simple structure, low cost on the equipment side, and can support different numbers of vision sensor devices according to the capabilities of the cloud computing center. However, because the image data needs to be uploaded through the network, although the image data can be reduced in size by a compression technique, it still suffers from a great deal of interference from the network state. When the network state is not ideal, the delay is very serious. And limited by the ability of the cloud computing center, when the number of visual sensors is huge, the computational expense of concurrent processing will be enormous of artificial intelligence technology, vision sensors are widely used in fields such as autopilot, security protection, and health care that require visual image information [6–11]. Computer Vision fusions can present more security system in an IoT platform for smart homes.
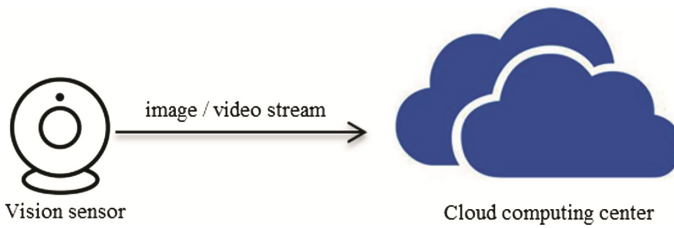
Fig. 1. Cloud-based face recognition system

The local-based face recognition system of Fig. 2, after the vision sensor acquires the image data again, the image data is transmitted to the local computing device through direct connection for comparison and analysis. The advantage is that it will not be disturbed by the network conditions and it will be processed quickly. The disadvantage is that the local processing hardware system based on artificial intelligence technology is expensive and the installation is inconvenient.
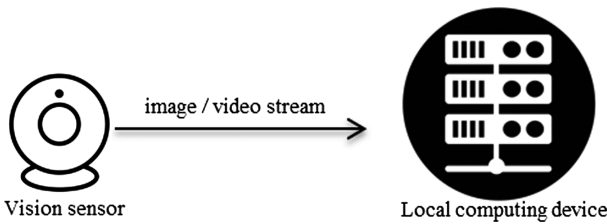
Fig. 2. Local-based face recognition system

Based on these two recognition systems, this paper proposes a face recognition system based on AI Edge and cloud computing (AE-FRS), which not only has lower

network delay time than the traditional cloud-based system, but also lowering the cost and easier to install than local face recognition system.

The following framework of this paper is as follows. Section 2 will describe the problems solved in this paper. Section 3 will introduce the face recognition scheme based on edge AI and cloud computing in detail. Section 4 will prove the effectiveness of the proposed scheme through experiments. Section 5 is the conclusion of this paper. Work with in future.

2   Problem Formations

The purpose of this article is to significantly reduce the response delay of cloud-based face recognition solutions. The response delay in the face recognition system is shown in Formula 1, t_{delay} represents the delay time until the visual sensor obtains the image until the recognition result returns, and the size is determined by the time spent in each step of the recognition process.

t_{delay} is mainly composed of three parts, the first part is the time delay t_{trans} transmitted from the visual sensor to the identification processing center, the second part is the processing center wait time t_{wait} (mainly due to the queue when the processing volume is large), and the third part is Recognition time t_{rec}, the size of which is determined by the algorithm's time complexity and hardware calculation speed.

$$t_{delay} = t_{trans} + t_{wait} + t_{rec} \tag{1}$$

In Sect. 3, we will describe how to reduce the system's response latency through edge AI. Since the edge AI's capability is sufficient to meet the requirements, and after a real test (see Sect. 4 for details), we will ignore the processing time of the edge AI.

3   AE-FRS

Compared with the old cloud-based face recognition system, AE-FRS will complete a part of the image processing that needs to be processed in the cloud due to the powerful computing power of the edge AI, which will greatly reduce the response delay. At the same time because of the edge AI and the need for complex identification calculations, will not increase the excessive local equipment overhead. We can see from Fig. 3 that compared with cloud-based systems, this system is more than happy local preprocessing. The following will explain in detail the work performed by the preprocessing and how to reduce the response delay of the system.
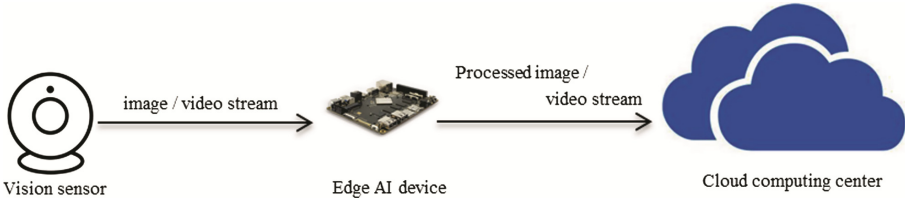
Fig. 3. AE-FRS

As shown in Fig. 4, the preprocessing is mainly divided into the following four steps: (1) background information processing (2) removal of background interference (3) removal of interference repeatedly identified by the same person (4) facial region reduction. Each step is described in detail below.
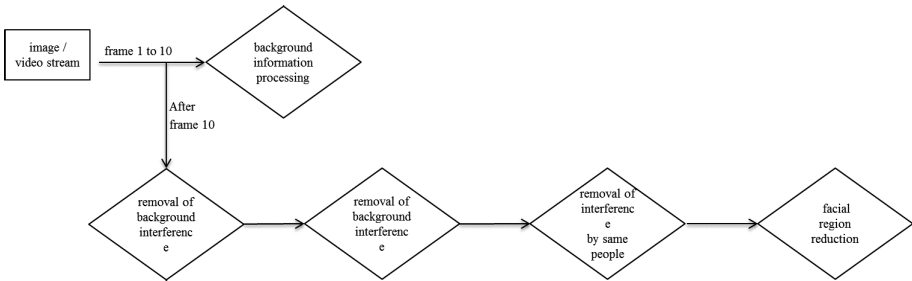

Fig. 4. Pretreatment

Step 1: background information processing, the main purpose is to eliminate the background of the human face interference, such as posters, billboards and other still face images. In real life, the scene that needs to be recognized may have the existence of interfered portrait, which not only brings extra recognition overhead but also takes up the file content to waste network traffic. Because of the calculation ability based on edge AI, this system uses the model based on Uniform Pattern LBP eigenvalue to detect the position of the face in the photo.

After the boot, the system checks a frame of an image every second, and the set of rectangular regions of a face detected in a frame I image is $S^i = \{s_1^i, s_2^i, s_3^i \ldots \ldots s_k^i\}$, s_k^i represents the location of the k-face detected in frame i, $s_k^i = (x_k^i, y_k^i, w_k^i, h_k^i)$. As shown in Fig. 5, x_k^i. denotes the distance between the face rectangle image and the left boundary of the k-frame image, y_k^i denotes the distance between the face rectangle image and the boundary on the k-frame image, w_k^i. denotes the width of the face rectangle image, and h_k^i denotes the height of the face rectangle image is indicated.

If a face rectangle appears in more than six frames, it is stored in the background interference region set S_{back}. Because the local area of the face in ten frames is rarely fully overlapped, we regard the rectangular area with an overlap ratio of more than 90% as a coincidence, and the other reclosing areas mentioned in this paper are also defined

Fig. 5. Face rectangular region

in the same way. A method of calculating the overlap ratio r of the k face rectangle of frame i and the l face rectangle region of frame j such as formula 2.

$$r = \frac{\left(s_k^i \cap s_l^j\right)}{\left(s_k^i \cup s_l^j\right)} \tag{2}$$

Step 2: remove background interference. After the previous step to determine the background interference S_{back}, the system entered normal working mode. After obtaining a frame of image from the visual sensor, the rectangular region of the face in the current frame is obtained by using the eigenvalue model based on Uniform Pattern LBP. If there is a rectangular region of the face, it should be compared with the rectangular region in the background interference set. Remove overlapping rectangular areas of the face.

Step 3: remove repeated interference from the same person. In the traditional image recognition system, when a continuous frame is recognized, if a person is still in the image, it will cause additional network overhead and recognition overhead for multiple recognition of that person, resulting in the delay of normal face recognition. Therefore, by comparing the face rectangular region set S_{k-1} of the previous frame with the face rectangle region of the frame, the final face region set S_{final}^i of the first frame is obtained by removing the overlapped region.

Step 4: make facial cuts. In face recognition, the traditional cloud system directly pushes the complete image or video stream, in which the redundant background information such as scenery, objects and so on occupying a large amount of image content, resulting in additional network overhead. The cloud recognition center needs information about the face and its vicinity. The final facial region set S_{final}^i obtained by the pre-recording step will be clipped according to the position of the facial region and the cut

image will be transmitted to the cloud for processing and recognition. In order to avoid the lack of image information near the face area, we can magnify the rectangular area of the face by a certain multiple (the default 1.2 times) and upload it.

The image processing recognition section of the cloud will be described in a separate article. Because of the edge AI based processing, the number of images uploaded and the size of the image are greatly reduced, thus reducing the t_{trans} and t_{wait}, thus reducing the t_{delay}.

4   Performances and Evaluation

In order to verify the effectiveness of this system, this paper uses rk3288 development platform, webcam, remote server background to build a set of identification system to prove the effectiveness of the system (Fig. 6).

Fig. 6.   System hardware equipment

The parameters of system are shown in Table 1.

Table 1.   System parameters

	Parameters
Pixel	2048(H) × 1536(V)
Video frame rate	30 fps
Focal length	2.8 mm
Main frequency	Four core A17, main frequency 1.8 Ghz
Memory	2 Ghz
Display card	Mali T760
Uplink bandwidth	500 k/s
Downlink bandwidth	2 M/s
Identification frequency	1 fps

In order to verify the processing speed of edge AI, the images containing 100, 500, 1000, 2000 and 5000 human images were transferred to rk3288 to obtain the time delay

from the input image to the final rectangular region S_{final}^i. You can see that it's stable at around 0.3 s (Fig. 7).
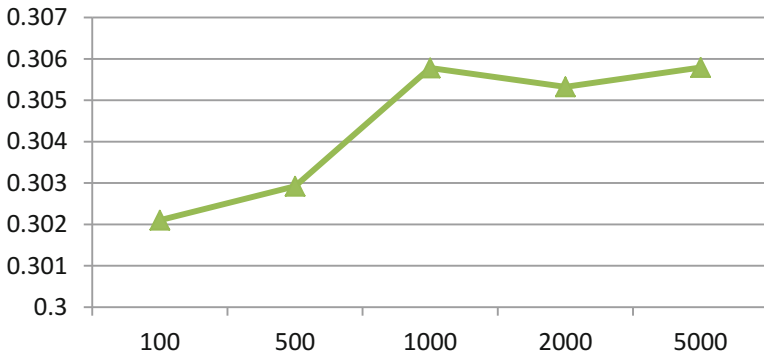

Fig. 7. Processing delay of edge AI

In order to verify the delay comparison between this paper's scheme and the cloud-based face recognition scheme, the rk3288 operation will be applied to the image set containing 100, 500, 1000, 2000, and 5000 portraits respectively, and the time from the image input to the server to return the verification result will be obtained. From the Fig. 8, we can see that the delay of the traditional cloud scheme of blue-folded modern (cyan line) watches has an average value of more than 5 s. With the increase in the number of photos, the delay is significantly increased. The increase in the number of images is 2,000. The cloud center handles the congestion. The purple broken line represents the edge-based AI-based cloud processing system. Its average delay is more than 2 s, and its delay does not increase or decrease significantly, because it eliminates photos that will bring extra recognition costs and reduces the size of the image required for recognition.
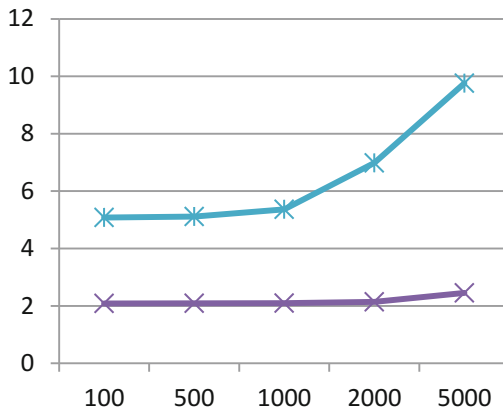

Fig. 8. Compare with traditional cloud-based scheme

5   Conclusions and Future Works

Experiments show that compared with the traditional face recognition cloud system, the face recognition system based on edge AI and cloud computing center proposed in this paper can effectively reduce the delay.

In the future work, we will continue to modify the relevant recognition algorithms in the cloud computing center in order to further reducing the delay.

References

1. Medina, C.A., Perez, M.R., Trujillo, L.C.: IoT paradigm into the smart city vision: a survey. In: 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 21–23 June 2017

2. Borgia, E.: The internet of things vision: key features, applications and open issues. Comput. Commun. **54**, 1–31 (2014)

3. Ericsson: More than 50 billion connected devices. Ericsson white paper, pp. 1–12 (2011)

4. Internet of things (IoT) 2013 to 2020 market analysis: Billions of things, trillions of dollars. International Data Corporation, Technical Report (2013)

5. Yigitoglu, E., Mohamed, M., Liu, L., Ludwig, H.: Foggy: a framework for continuous automated IoT application deployment in fog computing. In: 2017 IEEE 6th International Conference on AI and Mobile Services, June 2017

6. Munaro, M., Basso, F., Menegatti, E.: Tracking people within groups with RGB-D data. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2101–2107, October 2012

7. Cherubini, A., Chaumette, F.: Visual navigation of a mobile robot with laser-based collision avoidance. Int. J. Robot. Res. **32**(2), 189–205 (2013)

8. Ess, A., Leibe, B., Schindler, K., Van Gool, L.: A mobile vision system for robust multi-person tracking. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, pp. 1–8. IEEE (2008)

9. Piatkowska, E., Belbachir, A., Schraml, S., Gelautz, M.: Spatiotemporal multiple persons tracking using dynamic vision sensor. In: 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 35–40, June 2012

10. Reverter Valeiras, D., Orchard, G., Ieng, S.H., Benosman, R.B.: Neuromorphic event-based 3D pose estimation. Front. Neurosci. **9**(522) (2015)

11. Ni, Z., Ieng, S.-H., Posch, C., Regnier, S., Benosman, R.: Visual tracking using neuromorphic asynchronous event-based cameras. Neural Comput. **27**(4), 925–953 (2015)

12. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 711–720 (2007)