A New Intelligent Video Surveillance Architecture

Liwei Li¹, Haibin Xie²*

¹ National University of Defense Technology, Changsha, Hunan, 410000, China.
² National University of Defense Technology, Changsha, Hunan, 410000, China.
*Corresponding author’s e-mail: 1853483128@qq.com

Abstract. Aimed at the old mode of traditional video surveillance for manual operation and forensics, and the fact that massive amounts of video data occupy a large amount of storage and transmission resources, an intelligent video surveillance architecture based on network and big data is proposed. The architecture includes three core modules: face real-time detecting, face intelligent coding, and face fast retrieval, which can automatically detect human faces, and then intelligently encode them and compare them with databases quickly. It greatly reduces the space occupied by video data storage and transmission, improves the intelligent level of video surveillance, and transforms “after-the-fact analysis” to “in-the-fact analysis” to meet modern criminal investigation requirements.

1. Introduction

Nowadays, people are increasingly demanding security, and the role of video surveillance systems has become more and more obvious. Traditional video surveillances are mostly used for video capture, storage, and playback, which is difficult to be functioned as an early warning alarm. Real-time monitoring requires continuous monitoring of videos by humans, wasting a lot of manual labor and time. At the same time, with the explosive growth of data volume, the traditional monitoring mode has been unable to meet the storage and transmission of massive amounts of video data.

Intelligent video surveillance came into being. Intelligent video surveillance, also known as video analytics, is to automatically understand and analyze video using artificial intelligence and machine vision. At present, there is an amount of work on intelligent video surveillance technology. Bouwmans et al. [1][2] introduced target detection techniques from the perspective of background modeling. Yilmaz et al. [3], Wang [4], and Wu et al. [5] introduced the target tracking method from single-camera and multi-camera tracking. Huang et al. [6], Andreopoulos et al. [7], Zhang et al. [8] studied target classification and identification. Hu et al. [9], Morris et al. [10], and Aggarwal et al. [11] gave an overview of the behavioral recognition algorithm.

This paper proposes an intelligent video surveillance architecture based on face coding. The architecture includes a front end and a back end, and the front end is composed of six modules including face detection, intelligent coding, and fast retrieval. The back end is a database management module. Front-end and back-end data are transmitted through the cloud network, and a large amount of data can be stored through the cloud network. The face detection module completes the automatic detection of face in the video; the intelligent coding module concisely codes the face, solving the problem of high hardware resource consumption; the fast retrieval module realizes the comparison between coding of the target face and the face coding of the database.

The architecture realizes such functions as automatic face detection, face intelligent coding, and quick face retrieval, which improves the intelligent level of video surveillance, saves the hardware
resources for massive data storage and transmission, completes the upgrading and transformation of the intelligent video architecture, and enables intelligent video surveillance.

2. The Overall Architecture of Intelligent Video Surveillance System

2.1 System Components

As shown in Fig. 1, the intelligent video architecture system consists of six modules: video reception, face detection, face capture, image enhancement, intelligent coding, and fast retrieval.

Among them, the video reception module is responsible for collecting video from the camera; the face detection module quickly detects the faces in the video; the face capture module extracts the detected faces individually; the image enhancement module is responsible for the enhancement, denoising and other processing of face target pictures; the task of intelligent coding module is to intelligently and concisely encode the target picture; the fast search module quickly compares the target code with the database code.

2.2 Function Implementation

The intelligent video architecture function is shown in Figure 2. The entire function is divided into three parts: front-end operation, background operation and cloud operation.

Front-end operation part: The widely distributed camera monitors a large amount of video information in real time, and the received video information is real-time processed in the smart chip. Firstly, face real-time monitoring is performed, and the face frames appearing in the video are extracted as the pictures to be queried; afterwards, the image enhancement processing is performed on the face picture through interpolation, denoising, and other image processing methods; and then the obtained target picture is intelligently coded into a simple, robust and discriminative target code; finally, the target code is transmitted to the cloud through the cloud network.

Background operation part: the face database stores the face images of a large number of objects of interest, and image enhancement intelligent encoding and other processing are performed on all the images in the same way to obtain a coded database, and then all the codes in the entire coding database are transmitted to the cloud through the cloud network.

Cloud operation section: After the cloud receives the encoding of the face database and the target code, the cloud quickly retrieves the most similar object from the encoding database according to the target code, so as to judge whether the target object belongs to a face database. The purpose of warning alarm is achieved.
2.3 Application View
The application view of the intelligent video surveillance architecture is shown in Figure 3.

The intelligent video surveillance architecture proposed in this paper has a wide range of applications. Figure 3 shows four typical application scenarios. In the apprehension of fugitives and apprehension of terror, the front-end monitoring device compares the human face detected by the image processing and the intelligent coding with the background fugitive or face code database, and reports an alarm promptly when finding suspect. In the hotel security and community security applications, the monitoring device automatically sends the detected face code to the cloud at any time, and compares it with the face code registered in the database, and generates an early warning when noticing a foreign person.

3. Key Technologies
3.1 Face Detection
The purpose of face detection is to accurately detect faces in real time. However, for face detection technology, real-time performance and accuracy are a pair of coupled quantities, and it is difficult to achieve a very satisfactory result at the same time. Considering practical applications, real-time performance should be the first consideration. At the same time, the accuracy is easily affected mainly due to the different face posture changes, so it can be considered to detect only the positive face, thus simplifying the detection procedure and obtaining higher accuracy.

Based on the above analysis, VIOLA and JONES's classic algorithms can achieve real-time detection and high accuracy. The algorithm mainly has three contributions. First, a new image representation, called an “integral graph,” is introduced, which allows the detector to quickly calculate features. Second, the Adaboost algorithm is used, leading to the result that a small number of important visual features are selected from a series of a large number of potential features, thereby establishing a simple and effective classifier. Third, the classifiers are combined in a cascaded manner so that the background information is quickly eliminated, and it can be ensured that more calculations are applied to areas that may be like face.

Applying the method to achieve face detection is shown in Figure 4. It can be seen that the detection method can satisfy the task of detecting a positive face.

![Figure 4. Face Detection Effect Chart](image_url)

3.2 Intelligent Coding
BGP has the characteristics of simple calculation, strong discriminability and good robustness, which is very suitable to face recognition. And the BGP feature has the following advantages: First, BGP is defined in the image gradient pattern, and it has a good gradient characteristic, which can effectively deal with changes in light intensity and so on; Second, structural model and multi-spatial resolution are used in BGP. Structured BGP is equivalent to the edge detector, which is the key to accurate identification and concise expression. Meanwhile, the multi-spatial resolution strategy increases the ability of descriptors to cover pixels with different radius in neighborhoods.

At the same time, BGP’s related improvement methods also have great application value. The BGP method based on heuristic information introduces the weights and uses the prior knowledge and heuristic information in human face recognition to improve the level of intelligence in BGP coding. The discriminating power and robustness are strengthened without increasing the complexity of the algorithm. In addition, the feature-fusion cascaded BGP method is used to extract multi-layer features of face images, achieves deep texture information acquisition, and improves the accuracy of BGP face
recognition, but the complexity of the algorithm also increases. Therefore, the trade-off between encoding complexity and encoding robustness is particularly important.

Therefore, the intelligent video surveillance architecture proposed in this paper uses BGP or related improved methods as intelligent coding methods.

3.3 Fast Retrieval

Fast retrieval is based on the coding pyramid method. The coding pyramid is based on the hierarchical BGP method. It obtains depth information through multiple BGP extractions of the image, and uses different layers and blocks to construct the coding pyramid. The method of drawing from fuzzy matching to exact matching in retrieval is based on coarse-to-fine search ideas and is performed in layers according to the coding pyramid.

The purpose of constructing the coding pyramid is to form hierarchical codes with different lengths, laying the foundation for rapid retrieval from coarse to fine. There are two ways to form a hierarchical code. First, different BGP domain radius and block numbers are coded on the original image. Second, on the basis of the coding image of the previous layer, the coding is achieved by changing the radius of the BGP domain and the number of blocks.

Compared with the two methods, the second method has greater advantages. This is because the single-level BGP coding is not enough for the original image feature extraction. At the same time, it is noted that BGP encoded feature images still have rich image information such as gray information, texture information, and gradient information. If we continue to perform BGP encoding on the basis of feature images, we will certainly extract important information such as implicit deep texture information and edge information.

According to the above analysis, if each level of BGP feature coded image is used as the input of a single-level BGP algorithm, multiple sets of codes can be obtained. As the number of levels increases, the length of the BGP code can be shortened by controlling the BGP neighborhood radius or the number of blocks, thereby realizing the construction of the coding pyramid.

The constructed pyramid is shown in Figure 5 (assume that the number of levels is 5).

![Figure 5. BGP Coding Pyramid Diagram](image)

The width of each layer of the pyramid indicates the code length of the layer. As can be seen from the figure, the higher the number of levels is, the smaller the BGP code length is.

In the retrieval process, for the construction of the image encoding pyramid in the database, it is entirely possible to operate offline, and it is enough to perform online for the retrieval image. According to this idea, the retrieval framework can be indicated as shown in Figure 6.
4. Conclusion
This paper proposes a new intelligent video surveillance architecture, which has the advantages of improving the video surveillance intelligence level, saving data storage transmission resources, and so on. The architecture is described in detail in terms of architecture description and key technologies. It is of great inspiration and reference value for the development and application of intelligent video surveillance.

References
[1] Vachon, B., Baf, F. E., & Bouwmans, T. (2008). Background modeling using mixture of gaussians for foreground detection - a survey. Recent Patents on Computer Science, 1(3).
[2] Dollar, P., Wojek, C., Schiele, B., & Perona, P. (2012). Pedestrian detection: an evaluation of the state of the art. IEEE Transactions on Pattern Analysis & Machine Intelligence, 34(4), 743–761.
[3] Yilmaz, A., Javed, O., & Shah, M. (2006). Object tracking: a survey. Acm Computing Surveys, 38(4), 13.
[4] Wang, X. (2013). Intelligent multi-camera video surveillance: a review. Pattern Recognition Letters, 34(1), 3-19.
[5] Wu, Y., Lim, J., & Yang, M. H. (2013). Online Object Tracking: A Benchmark. Computer Vision and Pattern Recognition, 9, 2411–2418.
[6] Huang, K. Q., Ren, W. Q., & Tan, T. N. (2014). A review on image object classification and detection. Chinese Journal of Computers.
[7] Andreopoulos, A., & Tsotsos, J. K. (2013). 50 years of object recognition: directions forward. Computer Vision & Image Understanding, 117(8), 827-891.
[8] Zhang, X., Yang, Y., Han, Z., Wang, H., & Gao, C. (2013) Object class detection. ACM Computing Surveys, 46(1): 1311-1325.
[9] Hu, Q., Qin, L., & Huang, Q. (2013). A survey on visual human action recognition. Chinese Journal of Computers, 36(36), 2512-2524.
[10] Morris, B. T., & Trivedi, M. M. (2008). A survey of vision-based trajectory learning and analysis for surveillance. IEEE Transactions on Circuits & Systems for Video Technology, 18(8), 1114-1127.
[11] Aggarwal, J. K., & Ryoo, M. S. (2011). Human activity analysis: A review. ACM.