ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20171229024444
Original scientific paper

Applying the Convolutional Neural Network Deep Learning Technology to Behavioural Recognition in Intelligent Video

Lele QIN, Naiwen YU, Donghui ZHAO

Abstract: In order to improve the accuracy and real-time performance of abnormal behaviour identification in massive video monitoring data, the authors design intelligent video technology based on convolutional neural network deep learning and apply it to the smart city on the basis of summarizing video development technology. First, the technical framework of intelligent video monitoring algorithm is divided into bottom (object detection), middle (object identification) and high (behaviour analysis) layers. The object detection based on background modelling is applied to routine real-time detection and early warning. The object detection based on object modelling is applied to after-event data query and retrieval. The related optical flow algorithms are used to achieve the identification and detection of abnormal behaviours. In order to improve the accuracy, effectiveness and intelligence of identification, the deep learning technology based on convolutional neural network is applied to enhance the learning and identification ability of learning machine and realize the real-time upgrade of intelligence video's "brain". This research has a good popularization value in the application field of intelligent video technology.

Keywords: convolutional neural network; deep learning technology; intelligent video; optical flow method

1 INTRODUCTION

In various chemical parks, logistics parks and other similar parks, increased density of buildings, abundant products as well as people from other areas, of course, are inevitably a series of hidden security problems. With more strict requirements on safety and post-disaster rescue, the number of camera and video capture systems is increasing day by day, and the monitoring coverage is also getting wider and wider. In general, the traditional video management system only provides the functions of video capture, video data preservation and video playback. Otherwise, for the real-time detection of anomalous data and abnormal behaviour and to take prompt and effective measures, we need to keep watch on the monitors. However, according to the result of a laboratory in the United States, once common people stare at the video screen for more than 20 minutes, eyes will be extremely tired. That means almost 100 % of screen becomes "blind". In other words, the traditional video management system is not very qualified for the task of early warning and alarm. Therefore, we need to introduce intelligent video technology to provide guarantee for managers' decisions. At the moment, video capture and processing technology broadly is divided into 3 generations: analog, digital and intelligence [1].

The first generation (analog video monitoring system): In this stage, the video signal collected by the analog camera is transmitted through coaxial cable, displayed by analogue TV, and its data is stored by tape recorder. Whereas it is cheap, easy for installation, and suitable for small-scale security systems, the video monitoring technology still has problems on limited transmission distance and monitoring capability, overloaded video data capacity as well as low video quality.

The second generation (Digital Video Surveillance System): At this stage, we use network video sampling equipment and digital video server to work. The technology is available for networked IP, so it can accommodate more video capturing equipment to access to the Internet. Besides, firm and redundant digital storage can be used to protect video data permanently. But we always need our human eyes to find the problems in videos.

The third generation (Smart Video Surveillance System): The core of technology in this stage is based on video content understanding and analysis techniques of computer vision, and mainly on a series of algorithm analyses including background modelling, object detection, object identification and object tracking. And then it derives the dependent event of monitored objects to answer people's questions, to judge and predict regulators' behaviours and to give the alarm when abnormal behaviour occurs, according to human thoughts and default safety regulations.

Mr. Huang Kaiqi and other scholars have reviewed the development history of video management and related technologies, and summarized the intelligent video algorithm [1]. Qin Lele et al. have taken the video monitoring as an integral part of the Internet of things and applied it to the chemical industry park. Due to the restriction of current technology, intelligent video algorithm was not used [2, 3]. L. Sumi, v. Rang et al. applied the sensor and other technologies to smart city, but they did not supply more information about the intelligent video technology [4]. In Computer Vision and Pattern Identification Conference, R. Girshick, J. Donahue et al. mentioned the application of feature hierarchy algorithm of object detection and semantic segmentation in the intelligent video field [5]. By referring to relevant data, many scholars have begun an all-round research to the intelligent video technology within a certain range, but the research on applying optical flow method and convolutional neural network deep learning technology to intelligent video is just on its way. Therefore, there is a larger study space and application value.

The article presents an algorithm framework of bottom, middle and top level smart video monitoring, proposes object detections based on background modelling and object modelling. Moreover, detections can be applied to different scenarios. It also designs several methods to detect and recognize the abnormal behaviours of optical flows, and devises an identification combined deep approach of learning with video behaviours on account of convolutional neural networks, which aims at enhancing success rate and intelligence of it. This technology is able to replace "human eyes" to discover various anomalies in video pictures step by step, and can be widely used in security management system and emergency rescue system, and so on.

2 ALGORITHM FRAME FOR INTELLIGENT VIDEO SURVEILLANCE

The monitoring and calculation of intelligent video surveillance mainly studies how to extract semantic comprehension in line with human cognition from video images captured by cameras, that is to say, to make computer emulate human thinking, even to understand and analyze video data like humans. As a rule, the processing of video data and video images by intelligent video surveillance could be generalized into bottom layers, middle layers and high layers, as indicated in Fig. 1 [1].

2.1 Bottom Layer

The bottom layer mainly acquires object of interest from video data collection terminals, detects and tracks the object to pave the way for subsequent, further processing and analysis. The core mission of this layer is resolving "where is the object". Object detection could be grouped into object modelling and background modelling according to different modelling types.

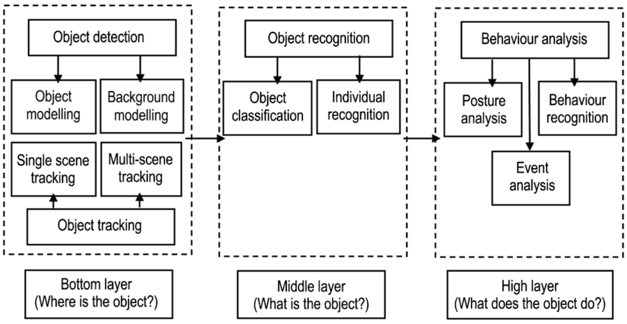


Figure 1 Algorithm flow chart of intelligent video surveillance

2.2 Middle Layer

The mission of middle layer is to determine and analyse various data of moving objects provided by bottom layer, including object identification, i.e. by classifying object and with resources of learning library, it further identifies the object. Its core is to sort out the problem "what the object is".

2.3 High Layer

The mission of high layer is to comprehend and analyse object behaviour with resources of learning library. The semantics of high layer includes specific semantic scene, usually closely related to specific application and specific scene. High layer is grouped into behaviour identification, posture identification and event analysis, whose goal is to sort out "what does the object do".

3 OBJECT DETECTION OF INTELLIGENT VIDEO SURVEILLANCE TECHNOLOGY

The aim of object detection is to extract an object of interest from a frame of video data, and identify the size, location and basic feature of the object. As for the technology today, object detection could be divided into detection methods of object modelling and background modelling according to different objects they process and their varied features.

The principle of background modelling detection method is to keep the background unchanged, while the object of interest is moving; by this principle, extract the object from the background and detect its size, velocity and location and so on. But when the background changes, this method would mistakenly identify the changing background as moving object; similarly, the moving object that has stopped for a while would be taken as background. So this method is not much suitable for mobile image acquisition unit. With these detection features of background modelling, an intelligent city would lay reasonable quantity of fixed cameras to ensure stability of background and improve detection performance [2, 3].

Object modelling detection method is free from restrictions on application scenes in the course of extraction, capable of handling data from mobile image acquisition units and performing object detection on videos acquired by fixed cameras. But due to huge amount of scanning windows, this method would cause low detection rate and poor efficiency, and it would be difficult to popularize it in those practical applications that have high real-time requirements. In applications of intelligent city, this method would be used to perform query of unstructured video data, i.e. use it for query and search after the event [4-7].

Table 1 Application of the Two Methods in Intelligent Cities

	Detection method based on background modelling	Detection method based on object modelling
Source data	Video	Source data video/image
Object	Moving object	Moving/static
Background requirement	Background must be fixed	Fixed/mobile
Algorithm speed	Fast	Slow
Susceptibility to occlusion	Very little	Great
Acquisition unit	Fixed image acquisition unit	Fixed image acquisition unit or mobile image acquisition unit
Usage	Daily real-time detection and early warning	Data query and search after the event

The detection comparison of the two methods and their applications in intelligent city is indicated as Tab. 1.

3.1 Object Detection Method Based on Background Modelling

Object detection method based on background modelling, first, structures background model according to background feature of the video image, then clips the moving object from the background and analyses and detects its size, location and shape and so on, and at the same time, changes background models with time according to a certain algorithm. The key of its implementation is how to structure a robust background model. In the video detection models of intelligent city, Gaussian Mixture Model (GMM) is adopted.

First, the features of various pixels in certain frame of image of video data are represented by *K* (*K*=3-5) Gaussian Models of Gaussian Mixture Model, and update model after acquiring a new frame of image, then use Gaussian Mixture Model to contrast and match various pixels of current image as necessary. If they match, the point could be identified as background point, otherwise, foreground point. In a comprehensive view, Gaussian Mixture Model is mainly determined by two parameters, mean value and variance. As for learning these parameters, different learning mechanisms will extend major influence over the stability, accuracy and convergence of the model. Since the system is modelled on background extraction of moving objects, the mean values and variance of model need to be updated in real-time. By improving update of these two parameters and adopting different learning mechanisms, learning capability of the model could be greatly enhanced. In order to improve the detection performance on big and slow, moving objects in busy scenes, the system introduces the mean of the weight into algorithm, and updates in real-time by setting up background image, then combined with weight, mean of the weight and background image, classifies pixels into foreground and background, and in this way, the problem could be resolved effectively [8-10]. The detection results are indicated as Fig. 2.

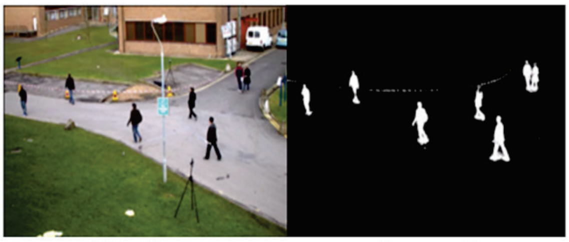
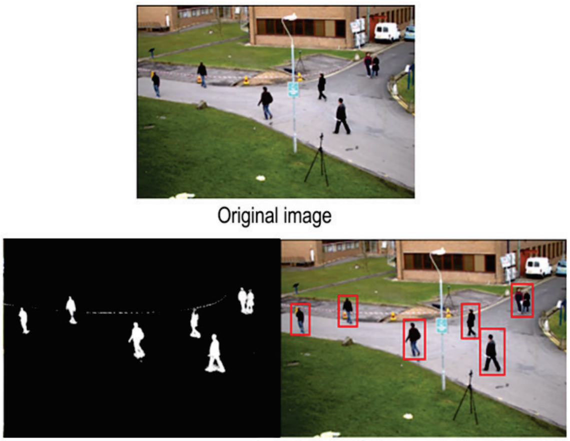

Original image Foreground

Figure 2 Detection Results of GMM Method


Original image

Background modelling object detection method

Object modelling object detection method

Figure 3 Contrast of object modelling and background modelling object detection methods

3.2 Object Detection Based on Object Modelling

The object detection based on object modelling requires simulation and learning on massive training objects, and by this means, trains the learner. The image will undergo sliding window scan in different measures so as to determine the scanned window is the background or object and to acquire the size, location and shape of the object in the image. Being different from the detection method based on background modelling, the extracted object by this method is a circling frame, not simply an outline as indicated in Fig. 3. This detection method is almost free from restrictions on scenes, capable of handling video data acquired by mobile image acquisition units to analyse and segment detection objects [11-15].

There are many object detection algorithms based on object modelling, but the key to the stability and accuracy of the whole intelligent video surveillance system lies in how to extract descriptive features of the image of interest from original image of background, therefore the core to solve this problem is how to build robust, scientific and accurate object model and classifier. Usually, the detection strategy based on sliding window adopts sliding window approach, and according to different methods, they may be grouped into detection model based on components, rigid global template detection, models based on learning and so on. In the latest research there are also bio-inspired model and grammar model and so on. The general technical architecture of object detection system based on sliding window is indicated as Fig. 4 [16-19].

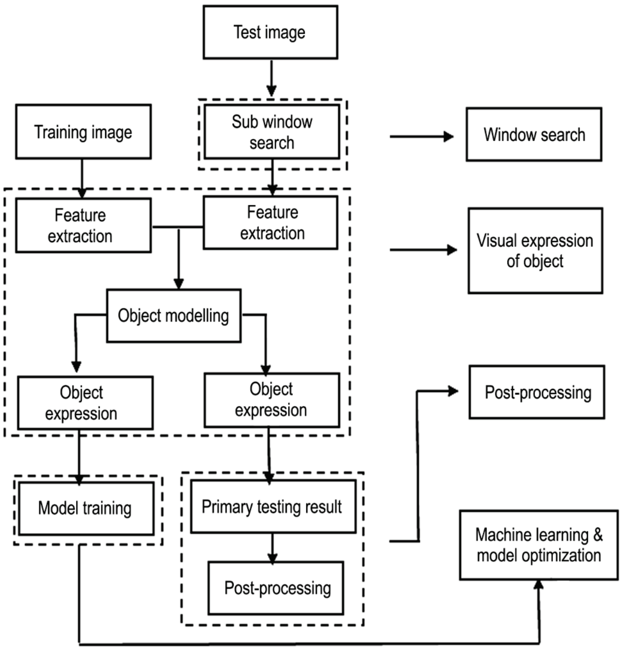


Figure 4 Architecture diagram of object detection system based on sliding windows

4 ABNORMAL BEHAVIOUR IDENTIFICATION AND DETECTION

In order to enhance the intelligent video surveillance system's capability of recognizing and detecting abnormal behaviours, it adopts an algorithm based on analysis of optical flow and deep learning. The algorithm first retrieves each frame of video images, uses optical flow method to establish image constraint equation, and determines whether there is abnormal behaviour(s) by the threshold value of velocity vector of behaviour. At the same time, by extracting properties of abnormal behaviour, it analyses and calculates layer by layer, uses convolutional neural networks technology to carry out deep learning and adjusts related weighted parameters by back propagation of convolutional difference to accomplish the learning process. The process is indicated as Fig. 5 [20].

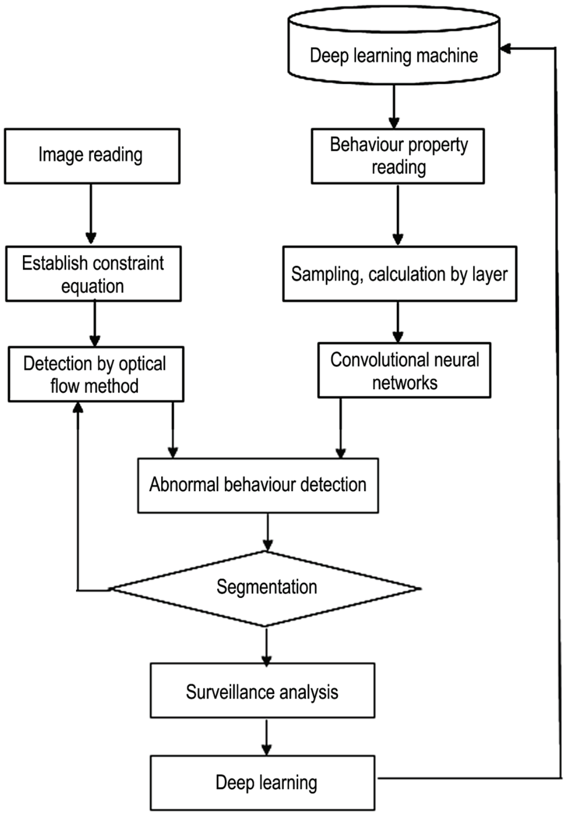

Deep learning machine
Image reading
Behaviour property reading
Establish constraint equation
Sampling, calculation by layer
Detection by optical flow method
Convolutional neural networks
Abnormal behaviour detection
Segmentation
Surveillance analysis
Deep learning

Figure 5 Technical architecture of abnormal behaviour identification and detection

As a rule, intelligent video surveillance systems adopt a basic process of "object detection, behaviour analysis and behaviour determination", i.e. they need to detect by background modelling whether there is object entering the scene, then conduct constraint equation computation over image, and combined with convolutional neural networks, determine whether it is abnormal behaviour, or to raise alarm. The system adopts an optical flow method, a frame difference, optical flow estimation method. Its principle: the method will assess whether there is deformation between two images. At first, it assumes that the image pixels and voxel are constant and there is no obvious or major change between two frames of contiguous images, and as the thinking goes, the image constraint equation is obtained and formulated as below [20]:

$$\boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{t})=\boldsymbol{I}(\boldsymbol{x}+\Delta\boldsymbol{x},\boldsymbol{y}+\Delta\boldsymbol{y},\boldsymbol{t}+\Delta\boldsymbol{t}), \tag{1}$$

Eq. (1) is a constraint equation as indicated above, which calculates the pixel moving status of two frames between t and Δt. As it is Taylor series based on image signals, this method may be called difference, i.e. partial derivative applied to time coordinate and space, and $f(x, y, t)$ is the pixel located at (x, y). If the displacement is fairly small, apply Taylor's formula to image constraint equation and obtain formula (2) [20]:

$$\begin{aligned}&\boldsymbol{I}(\boldsymbol{x}+\Delta\boldsymbol{x},\boldsymbol{y}+\Delta\boldsymbol{y},\boldsymbol{t}+\Delta\boldsymbol{t})=\\&=\boldsymbol{I}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{t})+\frac{\partial\boldsymbol{I}}{\partial\boldsymbol{x}}\Delta\boldsymbol{x}+\frac{\partial\boldsymbol{I}}{\partial\boldsymbol{y}}\Delta\boldsymbol{y}+\frac{\partial\boldsymbol{I}}{\partial\boldsymbol{t}}\Delta\boldsymbol{t}\end{aligned} \tag{2}$$

According to the equivalent relationship in formula (1), simplify and deduce formula (2) to obtain formula (3):

$$\frac{\partial\boldsymbol{I}}{\partial\boldsymbol{x}}\frac{\Delta\boldsymbol{x}}{\Delta\boldsymbol{t}}+\frac{\partial\boldsymbol{I}}{\partial\boldsymbol{y}}\frac{\Delta\boldsymbol{y}}{\Delta\boldsymbol{t}}+\frac{\partial\boldsymbol{I}}{\partial\boldsymbol{t}}\frac{\Delta\boldsymbol{t}}{\Delta\boldsymbol{t}} \tag{3}$$

in formula (3) $\frac{\Delta\boldsymbol{x}}{\Delta\boldsymbol{t}}$, $\frac{\Delta\boldsymbol{y}}{\Delta\boldsymbol{t}}$ are comprised of $\boldsymbol{x}$, $\boldsymbol{y}$ vectors of optical flow of $\boldsymbol{I}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{t})$ respectively, where $\frac{\partial\boldsymbol{I}}{\partial\boldsymbol{x}}$, $\frac{\partial\boldsymbol{I}}{\partial\boldsymbol{y}}$, $\frac{\partial\boldsymbol{I}}{\partial\boldsymbol{t}}$ are the corresponding direction differences of the image at point (x, y, t) as indicated in formula (4):

$$I_x\boldsymbol{V_x}+I_y\boldsymbol{V_y}=-I_t \tag{4}$$

In formula (4), optical vectors $\boldsymbol{V}_x$, $\boldsymbol{V}_y$ are comprised of x, y directions, respectively, where I_x, I_y, I_t are the corresponding direction differences of the image at point (x, y, t). This formula includes five unknown numbers, so the flow $(\boldsymbol{V}_x, \boldsymbol{V}_y)$ is a constant within a 3×3 small window. Therefore equation set (5) can be obtained [20]:

$$\begin{cases}I_{x1}\boldsymbol{V}_x+I_{y1}\boldsymbol{V}_x=-I_{t1}\\I_{x2}\boldsymbol{V}_x+I_{y2}\boldsymbol{V}_x=-I_{t2}\\\qquad\vdots\\I_{x9}\boldsymbol{V}_x+I_{y9}\boldsymbol{V}_x=-I_{t9}\end{cases} \tag{5}$$

Transform formula (5) into linear determinants as indicated in formula (6):

$$\begin{bmatrix}I_{x1}I_{x1}\\I_{x2}I_{x2}\\\vdots\\I_{x9}I_{x9}\end{bmatrix}\begin{bmatrix}\boldsymbol{V}_x\\\boldsymbol{V}_y\end{bmatrix}=\begin{bmatrix}-I_{t1}\\-I_{t2}\\\vdots\\-I_{t9}\end{bmatrix} \tag{6}$$

Transform formula (6) into formula (7):

$$A\vec{V}=-b \tag{7}$$

Introduce ordinary least square techniques into formula (7), solve it and obtain formulas (8) and (9):

$$A^{\mathrm{T}}A\vec{V}=A^{\mathrm{T}}(-b) \tag{8}$$

$$\vec{V} = (A^{\mathrm{T}} A) A^{\mathrm{T}} (-b) \tag{9}$$

From formula (8) and (9), we can obtain formula (10):

$$\begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} \sum I_{xi}^2 \sum I_{xi} I_{yi} \\ \sum I_{yi}^2 \sum I_{xi} I_{yi} \end{bmatrix}^{-1} \begin{bmatrix} -\sum I_{xi} I_{ti} \\ -\sum I_{yi} I_{ti} \end{bmatrix} \tag{10}$$

From V_x, V_y in formula (8) and (9), we may obtain the change rules, retrieve behaviour property based on the query model in learning machine and calculate layer by layer and we may determine the abnormal condition at the point of interest, foresee behaviours and raise alarm to administrator for final confirmation. The confirmed data will enter learning machine through convolutional neural network learning to compose a powerful model library so that the accuracies of intelligent video system to determine abnormal behaviour and behaviour prognosis are further enhanced.

5 EXTENSION AND UPGRADE OF INTELLIGENT SYSTEM BASED ON DEEP LEARNING

The intelligent video surveillance system in the city uses object detection model based on deep learning, and adopts convolutional neural network technology. Its principle is to use multi-layer connected network to simulate the visual processing mechanism and multi-layer abstract information processing mechanism of human brain, and by layers, conduct abstract processing on the image to acquire distinguishing features and to realize abstract classification, relevant description and abstract analysis of the image [21].

5.1 Basic Technical Architecture Based on Deep Learning Network

The system adopts convolutional neural network to construct its extended function based on deep learning, whose advantage is capable of avoiding parameter extraction of dominant character and where weighting parameters of each neuron are shared in each layer of the network. Neuron S- involves two important parameters-receptive field and threshold value, the former's function is to determine the number of input connection features and the latter's function is to control the extent of reaction to feature subsequence. The system is designed with 6 layers of convolutional neural networks, and parameters (the hyperlink weight between layers) of each layer need to acquire through training after initialization, whose functions go as follows [22]:

(1) The first layer (input layer): data acquisition, i.e. after certain pre-processing, input the acquired image data into the model;

(2) The second layer (the first convolutional layer): extract multi-view features of the input image;

(3) The third layer (sub-sampling layer): conduct feature extraction and finish sampling compression;

(4) The fourth layer (the second convolutional layer): extract feature map as partial detection image;

(5) The fifth layer (hidden layer): grade these partial detection images;

(6) The sixth layer (output layer): obtain the final label y.

In the course of training, all parameters are optimized through BP algorithm.

5.2 Pre-processing Image Data

First, set the training data into an 84×28 pixels window. When the image is input as detection image, first use this window to complete scanning work, and process 84×28 image data, and as input data, there are following channels:

(1) Shift the original image from RGB model to HSV (Hue Saturation Value) model, so the input channel becomes HSV channel;

(2) Adjust the 3 channels of HSV model into 14×43 and re-connect the 3 channels, and after zero supplement, they become the second channel;

(3) Use Sobel image edge detection to compute the horizontal and vertical edge grades of HSV image so that 3 sets of edge maps are obtained, take edge map of maximum order of magnitude to produce another edge map, then reconnect these four images and they become the third channel.

By unit variance and zero-mean process through various channels, the illumination change could be resolved effectively.

5.3 Feature Extraction

Under the combined influence of the first convolutional layer and sub-sampling layers of convolutional neural network, the feature extraction layer could be formed. The input layer of the first layer of convolutional neural network takes 84×28 image as input for the first layer, and 64 filters constitute the first layer. Input each 9×9 domain data of data to act upon parameters of filters, acting as a basic element of an input characteristic pattern, and under the convolutional influence of the 64 filters in the first convolutional layer, 64 frames of 67×20 characteristic maps could be generated.

Next, export to sub-sampling layer for filtering, and conduct average operation on input image data of each 4×4 domain, and 64 frames of 5×19 compressed feature extraction images could be obtained.

5.4 Generation of Partial Characteristic Detection Image

The second convolutional layer is the partial feature detection image obtained through further processing of the above mentioned feature image. As the sizes of various parts or components of interest object (such as people, vehicle and so on) vary, the detection operation should take this into full consideration. The algorithm is designed with three levels, and dozens of occlusion filters act as a component model. Several lower level components constitute component of a higher level, that is to say, as the level goes higher, the images become more complete. Parts of the important components learned by deep learning model are indicated in Fig. 6, which are the partial detection images of the second convolutional layer, such that sufficient visual clues could be acquired from the input image [23-24].

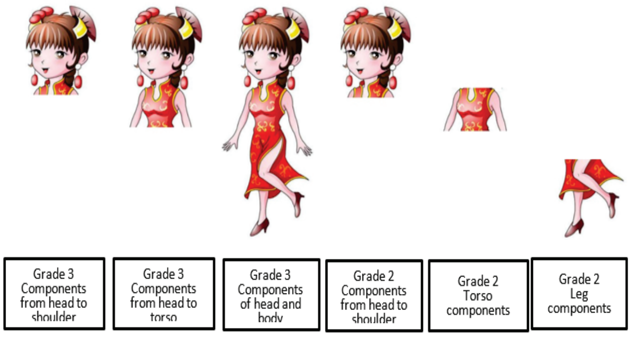

Grade 3 Components from head to shoulder
Grade 3 Components from head to torso
Grade 3 Components of head and body
Grade 2 Components from head to shoulder
Grade 2 Torso components
Grade 2 Leg components

Figure 6 Partial detection image of the second convolutional layer

5.5 Deformed Layer

In order to learn deformation parameters of various parts and components of the interest object, the system introduces deformation layer into convolutional neural network for necessary processing, and this deformation layer takes as input several partial detection images exported by the second convolutional neural network, and calculates grades of these partial detection images.

This layer weight-sums detection image and corresponding deformation image of each part to obtain sum image, as indicated in formula (11):

$$T_p = L_p + \sum_{n=1}^{4} C_{n,p} D_{n,p} \tag{11}$$

T_p – the final obtained sum image
L_p – the p^{th} partial detection image
$D_{n,p}$ – the n^{th} deformation image corresponding to the p^{th} partial detection image
$C_{n,p}$ – weight of corresponding $D_{n,p}$.

Then conduct global most prominent feature process on sum image, and partial detection grade is obtained as indicated in formula (12):

$$S_p = \max_{(x,y)} b_p^{(x,y)} \tag{12}$$

S_p – grade of the p^{th} partial detection image;
$b_p^{(x,y)}$ – the $(x, y)^{\text{th}}$ element of the T_p.

Coordinates of detection elements could infer formula (13) from sum image:

$$(x,y)_p = \arg\max_{(x,y)} b_p^{(x,y)} \tag{13}$$

In formula (13), $C_{n,p}$, $D_{n,p}$ are the key parameters for designing the whole deformation layer while these two parameters are determined through learning [22].

5.6 Classification Layer

The classification layer first estimates the visibility of the lowest level components, and upon this visibility, deduces the visibility of the second level component, and with the visibility estimation of the second level component, deduces visibility of the third level component, eventually, uses the obtained third level components to deduce the y label of the final result, as indicated in formula (14):

$$\tilde{v}_j^l = \delta(b_j^l + g_j^l s_j^l) \tag{14a}$$

$$\tilde{v}_j^{l+1} = \delta(\tilde{v}_j^{lT} W_{*,j}^l + b_j^{l+1} + g_j^{l+1} s_j^{l+1}),\ l = 1,2 \tag{14b}$$

$$\tilde{y} = \delta(\tilde{v}_j^{3T} w^{cls} + o) \tag{14c}$$

$\delta(t) = \left(1+\exp(-t)\right)^{-1}$ is Sigmoid function, g_j^l is the weighting parameter of s_j^l, b_j^l is its partial item, $\tilde{v}_j^l v^l$ is the visibility of the j^{th} of the l^{th} level components, W^l represents relevant coefficient of and v^{l+1}, $W_{*,j}^l$ represents collection of the j^{th} column elements of W^l, w^{cls} is the linear classifier of hidden unit $\tilde{v}^3$, o is offset parameters, eventually, $\tilde{y}$ estimation tag is obtained, where g_j^l, b_j^l, W^l, w^{cls}, o are the parameters determined through learning.

The system constructs a scientific, deep learning system, and through massive training of depth model, can effectively detect, recognize and determine objects of interest in actual scenes. At the same time, to overcome the shortcoming of the model relying on massive training in actual application, the actual behaviour properties and classifications, under human supervision or non-human supervision, are automatically added to the system library. The system could be upgraded online, and as time goes, the learning capacity and identification capability of the system will be greatly improved.

6 TEST FOR THE INTELLIGENT VIDEO SYSTEM OF THE TRIAL NETWORK OF INTELLIGENT CITY INFORMATION PLATFORM

6.1 Warning Test of Crowded People and Abnormal Behaviours

The system was carried out at the designated area of the city with none, small number of people (10 people), and many people (20 people), respectively. When the numbers of people were 0 and 10 people, the system gave no cues; when the number of people reached 20 people, the system sent out sound-light alarm, indicating regional address, physical parameters of environment and so on.

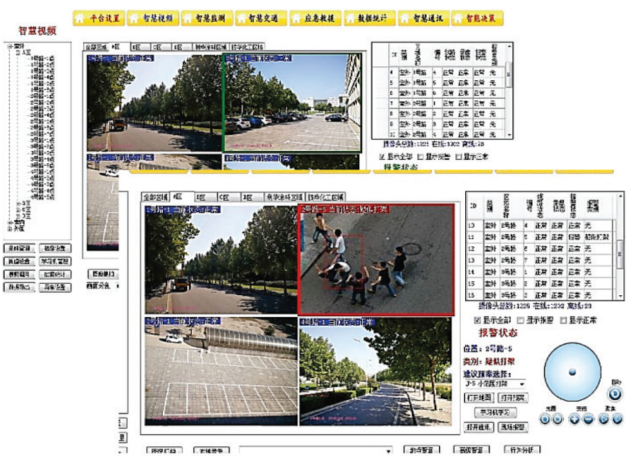
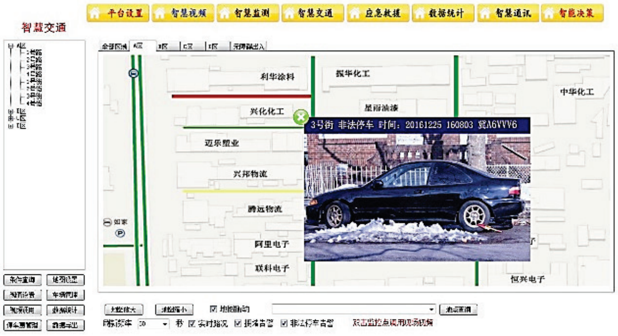


Figure 7 Intelligent video system-interfaces before and after emulation of fighting and brawling

In addition, behaviours such as long stay objects and emulation of fighting and brawling were also tested. The interfaces before and after the alarm are indicated as Fig. 7. The test has shown that the intelligent video system for intelligent city is successful.

6.2 Warning Test of Long Stay of Abnormal Object

It uses the related intelligent video monitoring functions to conduct object detection and identification of abnormal objects, realize "intelligent behaviour analysis" to the key zones in the park and give a warning to the abnormal behaviours (e.g. the vehicle parks at the prohibited area, and none-operating vehicle enter the controlled area, etc.). Then, the park administrator will make a corresponding control. As shown in Fig. 8, a car parks in the chemical control area and over the threshold setting time. The system immediately indicates the location of this car on the map and displays by monitoring image that the car is parking in the No. 3 street and the car's plate number is J. A6VVV6, the parking time is on December 25, 2016, 16:08:03.



3号街 非法停车 时间：20161225 160803 冀A6VVV6

Figure 8 Warning interface of abnormal vehicle behaviours

Through test, it is verified to be successful for applying optical flow approach to behaviour detection and identification in intelligent video, applying convolutional neural network deep learning technology to the real-time learning of learning machine, and conducting algorithm identification by utilizing learning machine to control the video.

7 CONCLUSION

With rising requirements of cities on security and post disaster relief capability, the quantity of cameras and video capture systems are increasing, and surveillance coverage is expanding further. As traditional video management systems could no longer adapt to technological development, the intelligent cities need to adopt intelligent video technology to help administrators to accomplish various tasks. This article designed an algorithm frame based on bottom layers, middle layers and high layers of intelligent video surveillance, proposed object detection approaches based on background modelling and object modelling for different applications, designed identification and detection methods for abnormal behaviours based on optical flow approach, and in order to improve success rate of identification and intelligent type, it designed a deep learning approach based on convolutional neural networks, combined with video action identification. After simulation test in a city, it has demonstrated good performance and achieved the expected goal.

Acknowledgements

This work is financially supported by Scientific Research Project of Hebei Science and Technology Department, China (No.16214707).

8 REFERENCES

[1] Huang Kaiqi, Chen Xiaotang, & Kang Yunfeng. (2015). Intelligent visual surveillance: a review. *Chinese Journal of Computers, 38*(6), 1093-1118.

[2] Qin Lele & Kang Lihua. (2016). Technical Framework Design of Safety Production Information Management Platform for Chemical Industrial Parks Based on Cloud Computing and the Internet of Things. *International Journal of Grid and Distributed Computing, 9*(6), 299-314. https://doi.org/10.14257/ijgdc.2016.9.6.28

[3] Qin Lele & Zhao Xin. (2012). Design and Realization of Information Service Platform of Logistics Parks Based on Cloud Computing. *Advances in Information Sciences and Service Sciences, 4*(23), 112-120. https://doi.org/10.4156/aiss.vol4.issue23.14

[4] Sumi, L. & Ranga, V. (2016). Sensor enabled internet of things for smart cities. *Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, Waknaghat, India, 295-300. https://doi.org/10.1109/PDGC.2016.7913163

[5] Girshick, R., Donahue, J., Darrell, T., et al. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, USA, 580-587. https://doi.org/10.1109/CVPR.2014.81

[6] Pinho, F., Sorreia, J. H., Sousa, N. J., et al. (2014). Wireless and wearable EEG acquisition platform for ambulatory monitoring. *2014 IEEE 3rd International Conference on Serious Games and Applications for Health (SeGAH)*, 1-7. https://doi.org/10.1109/SeGAH.2014.7067078

[7] Li Xiaobin, Wu Hongqi, &Yuan Zhanjun. (2016). Study on the Key Technologies and Algorithms of Intelligent Video Surveillance System. *Control Engineering of China, 23*(S0), 18-22.

[8] Sundmaeker, H., Guillemin, P., Fries, P., et al. (2016). Vision and challenges for realising the internet of things, 2010. bib0340.

[9] Jung, T., Dieck, M. C., Moorhouse, N., et al. (2017). Tourists'experience of virtual reality applications. *IEEE International Conference on Consumer Electronics, 25*(1), 8-10.

[10] David, C. & Gui, V. (2013). Automatic background subtraction in a sparse representation framework. *Proceedings of the systems, signals and image processing (IWSSIP)*, 63-66. https://doi.org/10.1109/IWSSIP.2013.6623450

[11] Kremic, E., Subasi, A., & Hajdarevi, K. (2012). Face recognition implementation for client server mobile application using PCA. *IEEE Proceedings of the ITI 2012 34th International Conference on Information technology interfaces (ITI)*, 435-440.

[12] Lee, J. & Seo, Y. H. (2014). An efficient head pose determination and its application to face recognition using multipose face DB and SVM. *The 9th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA2014)*. IEEE, 527-531.

[13] Andriyenko, A., Schindler, K., & Roth, S. (2012). Discrete-continuous optimization for multi-target tracking. *In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Providence, RI, USA, 16-21 June 2012, 1926-1933. https://doi.org/10.1109/CVPR.2012.6247893

[14] Sodagar, I. (2011). The MPEG-DASH standard for multimedia streaming over the Internet. *IEEE Multi Media, 18*(4), 62-67. https://doi.org/10.1109/MMUL.2011.71

[15] Wang, Y. & Yuille, A. L. (2013). An approach to pose-based action recognition. *Proceedings of the 26th IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Portland, USA, 915-922. https://doi.org/10.1109/CVPR.2013.123

[16] Idrees, H., Saleemi, I., Seibert, C., et al. (2013). Multi-source multiscale counting in extremely dense crowd images. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Portland, USA, 2547-2554.

[17] Huang Yongzhen, Wu Zifeng, & Wang Liang. (2013). Feature coding in image classification: A comprehensive study. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(3), 493-506. https://doi.org/10.1109/TPAMI.2013.113

[18] Concolato, C., Feuvre, J. L., Denoual, F., et al. (2017). Adaptive streaming of HEVC tiled videos using mpeg-dash. *IEEE transactions on circuits and systems for video technology, 99*, 1-2. https://doi.org/10.1109/TCSVT.2017.2688491

[19] Vasenev, A., Hartmann, T., & Dorée, A. G. (2017). Employing a virtual reality tool to explicate tacit knowledge of machine operators. *ISARC Proceedings*, 1-7.

[20] Liu Yong. (2017). Video Monitoring System Based on Optical Flow Field Analysis and Deep Learning. *Journal of Xiangnan University, 38*(2), 18-22.

[21] Lecun, Y. & Ranzato, M. A. (2013). *Deep learning tutorial. Atlanta, International Conference on Machine Learning (IC ML)*, Technical Report.

[22] Zeng Min & Zhou Yilong. (2015). Simulation of pedestrian detection based on deep learning model. *Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 35*(6), 111-116.

[23] Zhang, D. (2017). High-speed Train Control System Big Data Analysis Based on Fuzzy RDF Model and Uncertain Reasoning. *International Journal of Computers, Communications & Control, 12*(4), 577-591. https://doi.org/10.15837/ijccc.2017.4.2914

[24] Zhang, D., Sui, J., & Gong, Y. (2017). Large scale software test data generation based on collective constraint and weighted combination method. *Technical Gazette, 24*(4), 1041-1049. https://doi.org/10.17559/TV-20170319045945

Contact information:

Lele QIN, Associate Research Fellow (Corresponding author)
School of Economic Management of Hebei University of Science & Technology
Shijiazhuang Hebei 050000 China
E-mail: Mr_qin@163.com

Naiwen YU, Assistant Research Fellow
Polytechnic College of Hebei University of Science & Technology
Shijiazhuang Hebei 050000 China
E-mail: 40093419@QQ.com

Donghui ZHAO, Postgraduate Student
School of Information Science and Engineering of Hebei University of Science & Technology
Shijiazhuang Hebei 050000 China
E-mail: 18010224037@163.com