# Research Progress On Video Monitoring Based On Machine Learning

Tingfeng Liu[1, a], Jianjun Wu[1, b], Weifeng Zhu[2, c]

[1]College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China;

[2]College of Science, Henan University of Technology, Zhengzhou 450001, China;

[a]3137118581@qq.com, [b]13939003632@163.com, [c]395288252@qq.com

## Abstract

**The analysis of video surveillance has received more and more attention and is widely used. Aiming at the shortcomings and shortcomings of video surveillance algorithm research and application, it is proposed to apply machine learning algorithm to video surveillance with good effect. The basic knowledge of machine learning and the application status of machine learning in video surveillance are expounded, and the specific application of machine learning in video surveillance and monitoring is summarized. The performance of video surveillance algorithms needs to be improved..**

## Keywords

**Video surveillance,machine learning,target tracking,motion recognition,video retrieval,shadow removal.**

## 1. Introduction

Video is dynamic monitoring information formed by a large number of images in a continuous time series. Video surveillance is widely used in various fields, such as production line supervision, urban transportation, public space supervision, distance education, medical and health care, and living communities. The traditional video surveillance detection accuracy is poor, the identification is inaccurate, and it seriously affects the progress of the work. Video surveillance has evolved from analog surveillance and digital surveillance to today's intelligent surveillance, largely overcoming the drawbacks of traditional video surveillance.

The core of intelligent video surveillance is to implement automatic detection, tracking and behavior analysis on the target information in the obtained video information, sort the target critical degree, actively filter out the information with lower criticality, and timely feedback the high criticality information to Monitor personnel. There are still some problems with intelligent video surveillance. At present, there are many algorithms combined with intelligent video surveillance, but the overall performance is still affected by target segmentation, real-time tracking, and occlusion problems. Since the machine learning algorithm is applied to the video surveillance system, a new generation of intelligent video surveillance system is born.

## 2. Machine Learning

Machine learning is an important research area of artificial intelligence. Through the research of a large number of researchers, machine learning has developed to the present day, its main task is to study the data analysis technology that can be effectively realized by computer, and integrate various learning methods and various forms of integrated learning system, the application range of various learning methods. growing. In terms of scope. Machine learning and pattern recognition, statistical learning. Data mining is similar. At the same time, machine learning is combined with processing techniques in other fields. Interdisciplinary subjects such as computer vision, speech recognition, and natural language processing have been formed. The core of machine learning is to use algorithms to parse data, learn from it, and then make decisions or predictions about something in the world. This means that instead of explicitly writing a program to perform certain tasks, it is better to teach the computer how to develop an algorithm to accomplish the task. There are three main types of machine learning: supervised learning, unsupervised learning, and reinforcement learning.

## 3. Research Status of Machine Learning in Video Surveillance

Carolina Redondo-Cabrera et al. proposed two concatenated structures for convolutional neural networks and their corresponding new loss functions to learn unlabeled video, which together exploit local temporal coherence between successive frames and for separation The global discriminant boundary of the representation of different videos. First, we show how learning functions can be used to discover actions and scenes in a video collection. Secondly, we show the benefits of doing this unsupervised learning, only from unlabeled video, which can be used directly as an a priori monitoring task for the recognition of motions and objects in the image. Experimental results show that their unsupervised model can More than the pre-training supervision model.[1]Gül Varol et al. proposed a novel and effective large-scale motion recognition system based on real video clips. The improved dense trajec features are combined with Fisher vector coding and learned and classified using an extreme learning machine classifier. The system is a fast and accurate method that can replace traditional motion classification methods such as word packs and support vector machines. The results show that the extreme learning machine can effectively obtain higher precision without using a well-trained and computationally intensive deep neural network[19].JyothiV K et al. proposed a flower video retrieval system based on deep learning method. There are three ways to network training: using keyframes, segmentation flowers, and flower gradients as inputs. For a given query video, the system uses a multi-class Support Vector Machine (MSVM) to retrieve similar video from the database. We have conducted extensive experiments on one of our own relatively large floral video datasets, which contain more than 2,600 floral videos of 30 different types of flowers. The experimental results show that the complexity of the system can be greatly reduced by using the gradient of the flower to train the DCNN without affecting the performance of the system. However, among the three methods, the retrieval system trained by the split key frame shows the best retrieval accuracy. The experimental results show that the DCNN-based retrieval method is superior to the traditional retrieval method[22].Xiaohui Yuan et al. proposed a learning-based shadow removal method. Our approach suppresses light shadows by dynamically calculating thresholds and uses online learning strategies to remove shadows, which is fine-tuned with examples of automatic recognition in new videos. Experiments show that this method has good adaptability to video and can effectively remove shadows[22].Afsar et al. proposed an intelligent prediction system for predicting the final destination area of pedestrian free activity in a real environment. The system uses passive video capture to directly process raw video data, extracting motion features from automatically detected human bones (including body mass, head, hand and leg positions) for predicting pedestrian target position (location) , speed, acceleration. It mainly consists of three modules: human spot detection based on background subtraction; stellar skeleton detection - including shadow removal and contour peak detection; based on preprocessing (dimension reduction and balanced sampling methods) and logistic regression, neural networks, random forest (RF) and Final target region prediction for four classification methods of support vector machine (SVM). Under the sampling and RF models, all inputs get the best results[7].

## 4. Typical Research Applications of Machine Learning in Video Surveillance

### 4.1 Target Tracking

Target tracking in the field of computer research has always been a hot issue, and target tracking requires high stability, accuracy and real-time. However, the robustness of these performances is not high in the target tracking shown. The main problems are: (1) in the real environment, due to the obstruction of obstacles, the intensity of light, the change of weather, the change of background, it will interfere with the tracking target; (2) the deformation of the target, such as the underarm or collapse of the person makes the monitoring unrecognizable aims.

Commonly used target tracking algorithms include (1) Mean Shift-based target tracking algorithm (2) particle swarm optimization based tracking algorithm (3) template matching based tracking algorithm (4) TLD target tracking algorithm (5) model-based tracking. The existing method based on target tracking generally uses the connected domain analysis or template matching detection method after obtaining the foreground region through background modeling. In recent years, the research shows that the accuracy of the target detection method based on machine learning is much stronger than the connected domain. Analysis or template matching method. Jia Huixing et al. based on the Histograms of Oriented Gradients (HOG) and Support Vector Machine (SVM) target detection methods for accurate detection of human head and shoulders in video sequences, combined with target tracking And the module such as trajectory analysis realizes the automatic number of people in the video system. (Automatic number counting based on machine learning in intelligent video surveillance, Jia Huixing, Zhang Yujin, 2009). Lu Yunxiang and others used the HOG and SVM classifiers to identify pedestrians from the surveillance video frame by frame. According to SSIM, different pedestrians are classified, and the facial image of the pedestrian is intercepted by using the Haar-Like feature and the cascade classifier. Finally, the target pedestrian is identified in the surveillance video according to the eigenface algorithm, and is tracked according to the video position and time information[8].

### 4.2 Motion Recognition

Human motion recognition in video can be divided into two main steps: human motion extraction and representation and human motion classification and recognition. Human motion extraction and representation is a method of extracting and representing different information contained in various human actions. The main task of human motion classification is to use the information extracted from the video sequence that can reflect the human motion (such as model parameters, shape information, motion information, and direction information) to match the human body motion, determine the category of the current human motion, and classify it. The mainstream human motion classification methods are: template-based classification method, state transition-based classification method, and machine learning-based classification method.

The machine learning-based classification method mainly refers to the use of a classifier to train the sample data so that it automatically classifies the test data. Commonly used classifiers include the NN (Nearest Neighbor) classifier, which is a classifier that is directly and without training. It only needs to give a measure to measure the distance between samples; SVM (Support Vector Machine) classifier, The classifier uses the kernel function to map the training samples to the high-dimensional feature space, and then finds a hyperplane to segment and mark the training samples into different categories. The Boosting classifier can improve the accuracy of the weak classifier, first utilizing The weak classifier classifies the samples and generates a base classifier, and then uses the classification result to improve the accuracy of other weak classifiers. Thus, iteratively, a strong classifier that is weighted by multiple weak classifiers is finally obtained. Pontil and Verri use SVM for 3D object recognition. This method does not perform pose estimation and processes the image points in high dimensional space. Schuldt et al. implement motion recognition through SVM combined with local representation of spatiotemporal feature points. Haoran Wang proposed a motion recognition supervised classification method based on improved sparse model. In order to obtain spatial information of adjacent points of interest, a composite motion and appearance feature of low-level points of interest is proposed. At the upper level, a continuous motion segment descriptor is proposed to combine the chronological information of the motion. Secondly, an improved sparse model is proposed, which combines similarity constraints and dictionary non-coherent terms for classification. In addition, a classification loss function is proposed to optimize the classification dictionary. Experimental results show that the framework has comparable performance to existing frameworks[9].

### 4.3 Video Retrieval

In the video retrieval, the amount of video data is large and the dimension is high, which requires a large amount of memory and search time in the retrieval process. The key frame of the video represents the most prominent feature of each shot of the video, so accurately extracting the key frames of each shot can effectively reduce the processing time of the search and improve the accuracy of the search. Liang Jiansheng et al. proposed a video key frame extraction and video retrieval scheme based on deep learning, and designed three kinds of retraining modules: unsupervised, semi-supervised and supervised, which can effectively improve the feature extraction effect and video of convolutional neural networks. Retrieve accuracy[25]. Liu Song et al. explored the target fast retrieval algorithm in video, and used hash image matching algorithm and color layout and edge histogram descriptors. The matching algorithm is combined for target detection in video. It has a good effect on detection accuracy, and has a much better performance than the traditional image matching algorithm[12]. Hou Yanming et al proposed a multi-feature fusion video retrieval technology, first extracting video keyframes and their features, then calculating the similarity of keyframe feature sequences and matching them. Similar videos. Experimental results show that this method can achieve better video retrieval accuracy[13].

### 4.4 Video Shadow Removal

In scenes with strong lighting conditions, there are shadows in the video, and the shadows produced by the moving objects move along with the moving objects. When detecting targets, we tend to mark shadows as foreground targets, which can cause distortion of moving targets and distortion of contours of real moving objects. Affected by the complexity of shadow detection and recognition, the accuracy and efficiency of algorithms such as video segmentation and object detection recognition, image or video eigendecomposition are greatly reduced. In addition, the extracted shadows can be used for image/video scene editing to produce more vivid image/video effects. Therefore, the detection and removal of shadows is also an important research topic.

GUO R et al. proposed a single image shadow removal method based on paired regions, using custom low-level features to detect shadows, and then removing shadows based on physical illumination models, but processing high deng clear images is inefficient[14]; A method based on convolutional neural network is proposed to construct a multi-context structure from three levels of global positioning, appearance model and semantic model. Removes shadows from a single image in an end-to-end format[15]. Liao Bin et al. propose a region-based guided light-propagation shadow removal method, resulting in no shadow The video results are less time-spaced and the image error after restoration is smaller, and the root mean square error of the pixel is reduced; and the operation efficiency is higher[26]. Shuang Luo et al. proposed a shadow removal method based on separate illumination correction, in which shadow removal is performed only on the shadow-related illumination. An adaptive weighted total variation model is established for a space to obtain shadow-dependent illuminance and unshaded reflectivity. The objects in the shadow are detected based on the reflectivity, and then object-oriented illumination correction is implemented to compensate for the shadow area. Combined with the corrected illumination, shadow removal and reflection can be obtained. The results show that the method has a good visual effect[27].

## 5. Conclusion

This paper summarizes the research status and typical application research and methods of video monitoring based on machine learning. Target tracking, motion recognition, video retrieval, video shadow removal, etc. in video surveillance are all specific studies on a certain aspect of video surveillance to obtain the valuable information needed. From the existing research, the preprocessing of video dataset needs to be improved. In addition, the neural network algorithm can improve the accuracy of prediction. Applying the deep convolutional network to feature extraction of input images can also improve the classification to a certain extent. accuracy. Applying machine learning to video surveillance research has improved the accuracy of video processing, and further improved algorithms for further research in the future.

## Acknowledgements

This work was supported by Multi-parameter integrated sensing technology for grain conditions of China(31120043).


## References


[1] Carolina R C,Roberto L S:Unsupervised learning from videos using temporal coherency deep networks[J]. Computer Vision and Image Understanding, 2018:S1077314218301772-.

[2] Bacardit J,Llorà,Xavier:Large-scale data mining using genetics-based machine learning[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2013, 3(1):37-61.

[3] Hong S,Im W,Yang H S:Deep Learning for Content-Based,Cross-Modal Retrieval of Videos and Music[J]. 2017.

[4] Teng K,Wang J,Xu M,et al:Mask Assisted Object Coding with Deep Learning for Object Retrieval in Surveillance Videos[J].2014.

[5] Lee H H,Hong K S:Automatic recognition of flower species in the natural environment[J]. Image and Vision Computing, 2017, 61:98-114.

[6] Mohapatra D:Automatic removal of complex shadows from indoor videos[J]. Unt Theses & Dissertations, 2015.

[7] Afsar P,Cortez P ,Santos H:Automatic human trajectory destination prediction from video[J]. Expert Systems with Applications, 2018:S0957417418301805.

[8] Lv Yunxiang,Ma Lianxi, Xiong Hanxi, et al:Design and implementation of monitoring and video tracking and tracking system based on machine learning[J]. Industry and Information Education, 2016(11).

[9] Wang H,Yuan C,Hu W,et al:Supervised class-specific dictionary learning for sparse modeling in action recognition[J]. Pattern Recognition, 2012, 45(11):0-0.

[10] Li Xiang:Research on Video Sensitive Information Retrieval Based on Deep Learning[J]. Electronic Design Engineering, 2017(21).

[11] Ren Huifeng:Research Progress of Content-Based Video Retrieval[J]. Smart Factory, 2018(10):69-71.

[12] Liu Song,Wang Ziyiwen,Wei Dong, et al:Fast Target Detection Algorithm in Video[J]. Criminal Technology, 2018, 43(06): 15-18.

[13] Hou Yanming, Li Feifei, Chen Wei:Video Retrieval Algorithm Based on Multi-feature Fusion[J]. Electronic Science & Technology, 2019(5).

[14] Guo R,Dai Q,Hoiem D:Single-image shadow detection and removal using paired regions[J].2011.

[15] Qu L,Tian J,He S,et al: DeshadowNet: A Multi-context Embedding Deep Network for Shadow Removal[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2017.

[16] Fu Ping,Fang Shuai,Xu Xinhe,et al: Shadow Removal Method for Moving Target Detection in Video Surveillance System[J].Computer Engineering, 2007,33(10): 22-24.

[17] Xiaole S,Qingquan L,Yingjie T, et al: An Uneven Illumination Correction Algorithm for Optical Remote Sensing Images Covered with Thin Clouds[J]. Remote Sensing, 2015, 7(9):11848-11862.

[18] Luus,Bergh V D,Maharaj:The effects of shadow removal on across-date settlement type classification of QuickBird images[C]//Geoscience & Remote Sensing Symposium. IEEE, 2012.

[19] GülVarol,Albert AliSalah:Efficient large-scale action recognition in videos using extreme learning machines[J].Expert Systems with Applications,2015.

[20] Varol G , Salah A A . Efficient large-scale action recognition in videos using extreme learning machines[J]. Expert Systems with Applications, 2015.

[21] Hang L, Lu Z. Deep Learning for Information Retrieval[J]. 2016.

[22] V K, Jyothi & Guru, Devanur & Kumar,Y H.(2018):Deep Learning for Retrieval of Natural Flower Videos.Procedia Computer Science.132.1533-1542. 10.1016/j.procs.2018.05.117.

[23] Yuan X,Li D, Mohapatra D,et al:Automatic removal of complex shadows from indoor videos using transfer learning and dynamic thresholding[J]. Computers & Electrical Engineering, 2017:S0045790617329270.

[24] Fan OU,Chong L,Ou Fan,et al:Shadow removal from face image based on illumination compensation Face image occlusion shadow elimination based on illuminance compensation*[J]. Journal of Computer Applications, 2010, 27(1): 399- 400.

[25] Liang Jiansheng,Wen Heping:Video Key Frame Extraction and Video Retrieval Based on Deep Learning[J].Control Engineering,2019,26(05):965-970.

[26] Liao Bin,Wu Wen:A method for video shadow removal based on regional matching guidance[J].Computer Applications,2019,39(02):556-563.

[27] Shuang Luo,Huanfeng Shen,Huifang Li,Yumin Chen: Shadow removal based on separated illumination correction for urban aerial remote sensing images[J]. Signal Processing,2019.

[28] Wang H,Ye M,Yang S:Shadow compensation and illumination normalization of face image[J]. Machine Vision and Applications, 2013, 24(6):1121-1131.

[29] Shi, Lin, Li, Zhigang, Song, Dingli:A Flower Auto-Recognition System Based on Deep Learning[J]. IOP Conference Series: Earth and Environmental Science, 234.

[30] Guan H,Xue X Y,An Z Y: Advances on application of deep learning for video object tracking[J]. Acta Automatica Sinica, 2016.

[31] Song Y, Xu M, Qiao M, et al: Predicting Head Movement in Panoramic Video: A Deep Reinforcement Learning Approach[J]. 2017.