Camera Anomaly Detection based on Morphological Analysis and Deep Learning

Lingping Dong, Yongliang Zhang, Conglin Wen
School of Computer Science
Zhejiang University of Technology
Hangzhou, China
titanzhang@zjut.edu.cn

Hongtao Wu
School of Computer Science &Engineering
Hebei University of Technology
Tianjin, China

Abstract—**Recently, camera anomaly detection has attracted increasing interest in order to generate real-time alerts of camera malfunction for video surveillance systems. The existing camera anomaly detection methods still haven't enough ability to detect comprehensive types of anomaly, and lack the self-improvement ability in the case of miscarriage of justice by self-learning. So, this paper proposes a morphological analysis and deep learning based camera anomaly detection method to detect comprehensive types of anomaly. Morphological analysis is used to detect simple camera anomalies to accelerate the processing speed, and deep learning is utilized to detect complicated camera anomalies to improve the accuracy. The experimental results show that the detection accuracy of the proposed method achieves more than 95%.**

Keywords-video surveillance system; camera anomaly detection; morphological analysis; convolution neural networks

I. Introduction

Video surveillance systems (VSS) are widely used in the field of modern security. Millions of cameras are installed in VSS so that keeping the camera long time proper functioning is a fundamental requirement for all the technologies used in VSS. However, there are a large number of camera anomalies resulting in low quality videos or useless videos which lower the performance of VSS. Therefore, detecting camera anomaly to generate a real-time alert of camera malfunction becomes critical for VSS.

Camera anomaly means any sustained event which thoroughly alters the image captured by a video camera in VSS [1]. Various camera anomalies are usually caused by: (1) intentional sabotages, such as camera motion and occlusion; (2) natural conditions, such as image blurring due to fogging, leaf occlusion; (3) abnormal disturbances, such as screen shaking, defocus, color cast, and screen flickering [1]. Although anomaly detection is benefit to VSS for preventing from crimes by generating real-time smart alerts, it is not a trivial task considering complex surveillance situations in real world [2,3].

Camera anomaly detection is still a new emerging field although it has attracted increasing interest in both academic and industry. Most of previous research has paid more attention to the first type of camera anomaly for finding a camera tamper attack when the normal camera capture process is disturbed on malicious purposes. Ribnick et al.[4] attempted to identify camera tamper by detecting large differences between older frames of video and more recent frames. Aksay et al. [5] presented two wavelet domain methods to detect obscured camera view and reduced visibility. Sağlam and Temizel [6] proposed adaptive methods to detect abnormalities when the camera lens was defocused, moved, or covered. Lin and Wu [7] tried to detect occlusion, defocus, and motion, by detecting large edge differences and grayscale histogram comparisons between current and previous frames. Tung et al. [8] trained an adaptive background codebook model to detect displacement and obstruction tamper. Shih et al. [9] combined an edge intensity analysis and an illumination change detection to handle the dramatic illumination change or large crowds which passed through the scene. Yin et al. [10] utilized SIFT feature to capture the change between un-tampered and tampered video image when camera was moved or covered. Tsesmelis et al.[11] detected tamper events including defocusing, occlusion and displacement by comparing the incoming frames to a background model and extracting features relevant to each particular tamper type. Lee et al. [12] proposed an edge information based low-complexity algorithm for camera tamper detection.

Besides intentional sabotage, other abnormalities such as screen shaking, fogging, color cast, and screen flickering are also important, but have received less attention [1]. WANG et al. [2] devised a reduced-reference method to detect camera self-defocusing, spray-paint, covering, and redirecting. Wang et al. [3] devised a two-stage classifier by modeling image quality and video dynamics with probabilistic state transition. Huang et al. [1] proposed an automated method for rapidly detecting camera tamper and various abnormalities, such as screen shaking, fogging, defocus, color cast, and screen flickering, based on brightness, edge details, histogram distribution, and high-frequency information. Yuan et al. [13] used Support Vector Machine to detect leaf occlusion. Liu et

The project was supported by the Research Foundation of Education Bureau of Zhejiang Province (No.Y201329342), the Science and Technology Planning Project of Hebei Province, (No.15210124), and the Science and Technology Research Project of Higher School in Hebei Province (No. Z2015105)

978-1-5090-4165-7/16/$31.00 ©2016 IEEE

al. [14] presented a spatiotemporal domain characteristics-based fast blind detection algorithm for stripe disturbance based on color, shape, and movability information. Ji and Zhou [15] classified tamper into five categories, namely covered camera, defocused, intensity error, color error, and noise of video, and then extracted different features for each category to detect tamper and recognize its type.

To deal with more categories of camera tamper, this paper proposes a morphological analysis and deep learning based camera anomaly detection method. In the proposed method, morphological analysis is used to detect simple camera anomalies to accelerate the processing speed, and deep learning is utilized to detect complicated camera anomalies to improve the accuracy.

The reminder of this paper is organized as follows. In Section II, the proposed method is given in detail. Experimental results are provided in Section III. Finally, the paper is concluded in Section IV.

II. Proposed Method

By summarizing previous research, all camera anomalies are classified into seven categories as shown in Table I in this paper. The seven camera anomaly categories are brightness fault, freeze abnormal, disturbance abnormal, color cast, lose abnormal, shake abnormal, and clearness abnormal. Some samples of camera anomalies are shown in Fig. 1.

Since morphological analysis is very effective for image analysis, it has often been used in previous methods. However, morphological analysis is not efficient enough to deal with some complicated anomalies, such as blurring, camera occlusion, and stripe disturbance. In order to improve the robustness of morphological analysis based anomaly detection, deep learning is introduced to detect the complicated anomalies in this paper. In other words, the proposed method includes two-stage detectors to satisfy the requirement of real application with high efficiency and low false alarm rate. First, morphological analysis is carried out on video image sequence captured by VSS to detect simple camera anomalies. Then a deep learning based camera anomaly detector is applied to detect complicated anomalies. At last, the detection results of two-stage detectors are combined to determine whether a camera anomaly appears. The flowchart of the proposed method is given in Fig.2.

A. Anomaly Detection based on Morphological Analysis

Based on a large number of experiments, morphological analysis based anomaly detection has sufficiently good performance to detect five camera anomalies listed as follows.

(1) too bright or too dark

To detect anomalies that camera is too bright or too dark, a brightness threshold *L* and a darkness threshold *D* are set. Then, all grey level of pixels in a video image is compared with *L* and *D*. If the number of pixels with grey level great than *L* is greater than threshold *NL*, the video image is identified as too-bright-anomaly. Similarly, if the number of pixels with grey levels less than *D* is greater than threshold *ND*, the video image is identified as too-dark- anomaly.

(2) gain disorder

Selecting a suitable color space is a key to measure image gain disorder [16]. YUV color space is utilized to detect gain disorder in this paper.

$$\begin{aligned} Y &= 0.114 * B + 0.587 * G + 0.299 * R \\ U &= 0.436 * B - 0.147 * R - 0.289 * G \\ V &= 0.615 * R - 0.515 * G + 0.100 * B \end{aligned} \qquad （1）$$

where Y stands for the luminance component, and U and V are the chrominance components. As shown in Fig.3, the origin of coordinate means image grey. The greater the distance from the origin, the deeper the color is. Video image is segmented by setting deep color pixel threshold *dcp*. When the number of deep pixels is greater than 1/2 of the number of all pixels, and each partition of coordinate system as shown in Fig.3 and each image partition have uniform number of deep pixels, the current video image is an image with gain disorder.

(3) freeze abnormal

When a certain number of consecutive video images are identical, freeze abnormal appears.

(4) lose abnormal

TABLE I Camera anomaly category

category	description
brightness fault	too bright, too dark, gain disorder.
freeze abnormal	frozen image video
interference abnormal	strip,ribbon, corrugated and snowflake interferences
color cast	too much red, too much green or too much blue
lose abnormal	black screen, blue screen
scrolling abnormal	screen shaking, screen scrolling
clearness abnormal	camera occlusion, image blurring

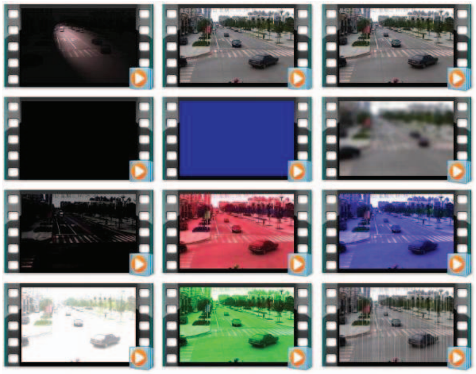


Figure 1 Camera anomaly samples

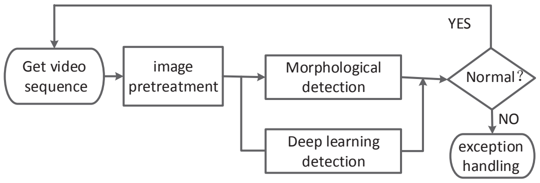


Figure 2 Flowchart of the proposed method

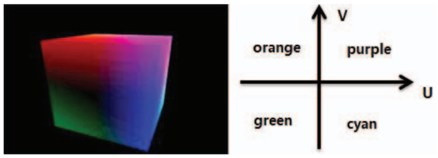


Figure 3 YUV color space and U, V-dimensional coordinate diagram

When a screen picture of a monitoring video is missing, blue or black screen presents. Therefore, lose abnormal detection is implemented by counting the number of blue screen frames or black screen frames.

(5) scrolling abnormal

In scrolling abnormal, the content change of adjacent image frames is little, but the locations of their corresponding contents are shaking. If the adjacent image frames are aligned, ghosting phenomenon will occur and their contents will increase. The detection steps are as follows based on above observations:

First, a gradient map of current image is calculated based on Eq. (2).

$$\begin{aligned} f(x,y) &= d_x(x,y) + d_y(x,y) \\ d_x(x,y) &= I(x+1,y) - I(x,y) \\ d_y(x,y) &= I(x,y+1) - I(x,y) \end{aligned} \qquad (2)$$

where I (x,y) is the pixel value of a pixel (x,y) in current image, and $f(x,y)$ is its gradient value.

If $\| f_k(x,y)| - |f_{k+2}(x,y)\| > T_k$, pixel (x,y) is identified as a contour point of the scrolling image region. Here, T_k is a threshed. For current image, the number N_1 of contour points and the number N_2 of the edge points are calculated. If $| N_2 - N_1 |$ is greater than $1/10$ of the area of current image, the current image frame is identified as scrolling image.

(6) Color cast

RGB color space is not suitable for color cast detection in some scenarios. For example, if green leaves account for a large proportion in an image, it is easy to identify the normal image as abnormal image with too much green. So the RGB color space is converted to Lab space. Lab color model is a color-opponent space with dimension *L* for lightness and *a* and *b* for the color-opponent dimensions, based on nonlinearly compressed coordinates.

There are no simple formulas for conversion between RGB values and Lab because the RGB color models are device-dependent. So, the RGB values first are transformed to the XYZ color space according to Eq.(3), and then the XYZ values are transformed to the Lab color models according o Eq.(4).

$$\begin{aligned} X &= 0.4124 \times R + 0.3576 \times G + 0.1805 \times B \\ Y &= 0.2126 \times R + 0.7152 \times G + 0.0722 \times B \\ Z &= 0.0193 \times \mathrm{R} + 0.1192 \times \mathrm{G} + 0.9505 \times \mathrm{B} \end{aligned} \qquad (3)$$

$$\begin{aligned} \mathrm{L} &= 116 f(Y/Y_n) - 16 \\ \mathrm{a} &= 500[f(X/X_n) - f(Y/Y_n)] \\ \mathrm{b} &= 200[f(Y/Y_n) - f(Z/Z_n)] \end{aligned} \qquad (4)$$

where

$$f(t) = \begin{cases} t^{1/3} & \text{if } t > (\frac{6}{29})^3 \\ \frac{1}{3}(\frac{29}{6})^2 t + \frac{4}{29} & \text{otherwise} \end{cases} \qquad (5)$$

X_n , Y_n and Z_n are the CIE XYZ tristimulus values of the reference white point. Here, the values of X_n , Y_n and Z_n are 95.047, 100 and 108.883,respectively.

After Lab values are calculated, image average chrominance *D*, central moment of image chrominance *M*, and color cast factor $K = D / M$ are computed based on Eq.(6). Then, color cast is determined according to color cast factor calculation criteria as shown in Fig.4.

$$\begin{aligned} d_{(a/b)} &= \frac{\sum_{i=0}^{H} \sum_{j=0}^{W} (a/b)}{MH} \\ M_{(a/b)} &= \frac{\sum_{i=0}^{H} \sum_{j=0}^{W} ((a/b) - d_{(a/b)})^2}{MH} \\ D &= \sqrt{d_a{}^2 + d_b{}^2}, M = \sqrt{M_a{}^2 + M_b{}^2} \end{aligned} \qquad (6)$$

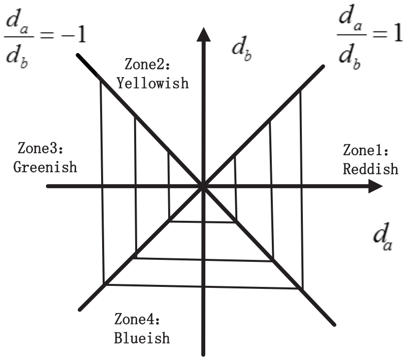


Figure 4 Color cast factor calculation criteria

B. *Anomaly Detection based on Deep Learning*

Besides some simple camera anomalies described above, there are some complicated anomalies, such as strip interference, camera occlusion and image blurring, which are difficult to be detected by morphological analysis based detection methods. Deep learning has strong learning ability and high efficient feature representation, which is extracted from the pixel level raw data to the abstract semantic concepts. This makes it a prominent advantage in extracting the global features of the image and the context information [17]. So this paper introduces Convolutional Neural Networks (CNN) to detect these complicated anomalies because CNN can generate stronger feature vectors that are more invariant to image distortion and position. The architecture of our CNN is shown in Fig.5. It consists of 4 convolutional layers, 3 pooling layers, 1 fully-connected layer. Softmax loss is used to train a classifier. Stochastic gradient descent with back propagation is utilized to optimize our deep learning based anomaly detector.

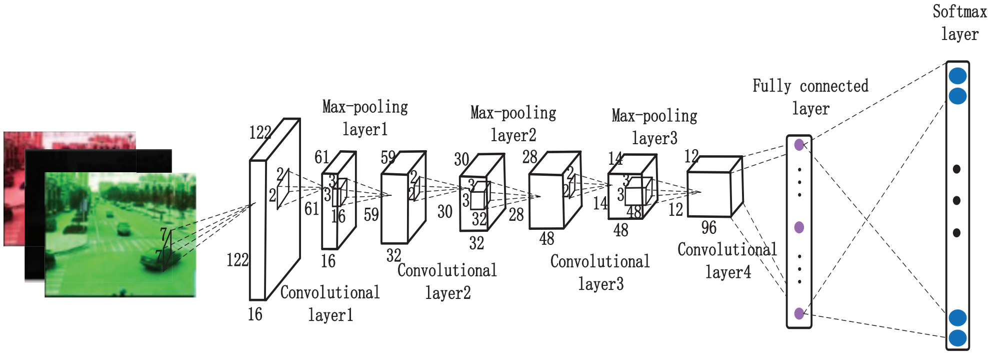


Figure 5 The architecture of our convolutional neural network

TABLE II accuracy-time relation table for residence community test

Detection duration(day)	2	4	8	10	12	16	24	26	28	30
Accuracy (%)	96.25	96.25	95	99.375	98.25	100	98.75	100	99.375	99.375
Average detection time (second/round)	252	248	264	252	254	246	248	258	248	246

TABLE III accuracy-time relation table for school test

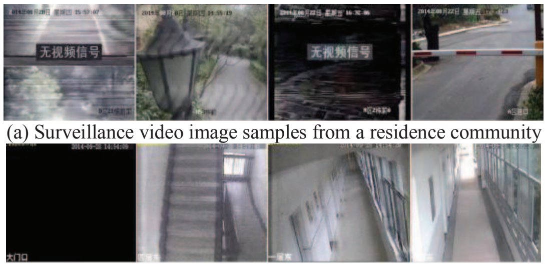
Detection duration(day)	2	4	8	10	12	16	24	26	28	30
Accuracy (%)	98.125	99.375	100	100	99.375	100	99.375	98.75	100	100
Average detection time (second/round)	236	248	234	252	252	246	238	236	242	236

III. EXPERIMENTAL RESULTS

As there are no public datasets available for camera anomaly detection, we collect a certain amount of surveillance video from two real application scenarios to evaluate the performance of our proposed method. One scenario is a residence community where the VSS includes 10 16-channel Hikvision DVR surveillance devices and 160 cameras. Another is a school where the VSS includes 14 16-channel Hikvision DVR surveillance devices and 194 cameras. The normal and abnormal samples from two scenarios are given in Fig. 6(a) and Fig. 6(b).

The test results for accuracy and average detection time in various different detection durations are shown in Table II and Table III. The accuracy of the proposed method in two application scenarios exceeds 95% and indicates that the proposed method can satisfy the requirements of real application. The false alarm rates in various different times of a day are shown in Fig. 7. Test results in Fig.7 suggest that false alarms occur when light condition changes relatively greatly. Because false alarms mostly happen in anomaly detections of camera occlusion and image blurring, we apply “detection-train-detection” tactics in which video images leading to false alarms are used as training samples to train a new classifier for the following camera anomaly detection. Experimental results demonstrate that the tactics enhances the robustness of our anomaly detection method and prevents the recurrence of the same false alarm.

Finally, we compare our proposed deep learning based detection method with the proposed method in [13] for strip interference and the proposed method in [15] for camera occlusion. The comparison results shown in Fig.8 demonstrate that our deep learning based detection method has more excellent accuracy for complicated anomalies. As time goes by, when enough samples leading to false alarm are added as training samples, the accuracy of our deep learning based detection method can be further improved.



(a) Surveillance video image samples from a residence community

(b) Surveillance video samples from a school

Figure 6 Video image samples

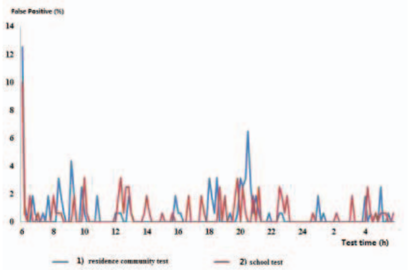


Figure 7 False alarm rates

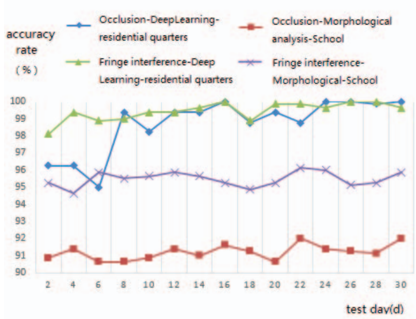


Figure 8 Deep learning-based method vs. morphological analysis based method for strip interference and camera occlusion

IV. Conclusion

A solution combing morphological analysis and deep learning to detect more comprehensive camera anomalies has been presented in this paper. The proposed method utilizes morphological analysis based detection method to deal with simple camera anomalies to fasten the process of anomaly detection. Due to its strong learning ability and high efficient feature representation, deep learning is applied to handle complicated anomalies to improve the accuracy of anomaly detection. Experiments have shown that the proposed method could produce high precision and low false alarm rate in two different scenarios. In future work, we try to investigate more sophisticated camera anomalies, such as camera redirecting detection.

References

[1] D. Y. Huang, C. H. Chen, T. Y. Chen, W. C. Hu, and B. C. Chen, “Rapid detection of camera tampering and abnormal disturbance for video surveillance system,”J.Visual Communication and Image Representation, vol.25, no.8, pp.1865-1877, 2014.

[2] Y. K. Wang, C. T. Fan, K. Y. Cheng, and S. Deng, “Real-time camera anomaly detection for real-world video surveillance," in Proc. International Conference on Machine Learning and Cybernetics, Piscataway, NJ:IEEE Press, 2011,pp.1520-1525

[3] Y. K. Wang, C. T. Fan, and J. F. Chen, “Traffic camera anomaly detection,” in Proc. 22nd International Conference on Pattern Recognition, Piscataway, NJ:IEEE Press, 2014, pp. 4642-4647.

[4] E. Ribnick, S. Atev, O. Masoud, N. Papanikolopoulos, and R. Voyles, “Real-Time Detection of Camera Tampering,” in Proc. IEEE Conference on Advanced Video and Signal Based Surveillance, Piscataway, NJ:IEEE Press,2006, pp. 1-6.

[5] A. Aksay, A. Temizel, and A. E. Cetin, “Camera Tamper Detection Using Wavelet Analysis for Video Surveillance,” in Proc. IEEE Conference on Advanced Video and Signal Based Surveillance, Piscataway, NJ:IEEE Press,2007, pp. 558-562.

[6] A. Sağlam, and A. Temizel,“Real-time Adaptive Camera Tamper Detection for Video Surveillance,”in Proc. IEEE Conference on Advanced Video and Signal Based Surveillance, Piscataway, NJ:IEEE Press,2009, pp.430-435.

[7] D. T. Lin, and C. H. Wu, “Real-time Active Tampering Detection of Surveillance Camera and Implementation on Digital Signal Processor,” in Proc. Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Piscataway, NJ:IEEE Press,2012,pp. 383-386.

[8] C. L. Tung, P. L. Tung, and C. W. Kuo, “Camera Tamper Detection Using Codebook Model for Video Surveillance,” in Proc. International Conference on Machine Learning and Cybernetics, Piscataway, NJ:IEEE Press, 2012, pp. 1760-1763.

[9] C. C. Shih, S. C. Chen, C. F. Hung, K. W. Chen, S. Y. Lin, C. W. Lin, and Y. P. Hung, “Real-time camera tampering detection using two-stage scene matching,” in Proc. 2013 IEEE International Conference on Multimedia and Expo, Piscataway, NJ:IEEE Press,2013,pp.1-6

[10] H. Yin, X. Jiao, X. Luo and C. Yi, “Sift-based camera tamper detection for video surveillance”, in Proc. IEEE 25th Chinese Control and Decision Conference, Piscataway, NJ:IEEE Press, 2013, pp.665-668.

[11] T. Tsesmelis, L. Christensen, P. Fihl, and T. B. Moeslund, “Tamper detection for active surveillance systems,”in Proc. 10th IEEE International Conference on Advanced Video and Signal Based Surveillance, Piscataway, NJ:IEEE Press, 2013,pp.57-62.

[12] G. Lee, Y. Shin, J. Park, and M. Lee, “Low-Complexity Camera Tam per Detection based on Edge Information,” in Proc. IEEE International Conference on Consumer Electronics-Taiwan, Piscataway, NJ:IEEE Press, 2014, pp. 155-156

[13] Y. Yuan, S. Ding, X. Xu,and C. Chen, “Support vector machine based approach for leaf occlusion detection in security surveillance video,” Journal of Computer Applications, vol.34, no.7, pp.2023-2027, 2014 (in Chinese).

[14] J. X. Liu, L. Chen, and K. H. Zhou, “A fast blind detection algorithm for horizontal stripes disturbance in surveillance video,” Computer applications and software,vol.32,no.12,pp.175-178 ,2014 (in Chinese).

[15] G. Ji, and L. Zhou, “Tampering detection and classification of intelligent video surveillance system,” Journal of Data Acquisition & Processing, vol. 28, no. 2, pp.231-238, 2013(in Chinese).

[16] E. A. Styles, Attention, perception, and memory an integrated introduction. New York :Taylor& Francis Routledge, 2005.

[17] Y. Bengio, I. Goodfellow, A. Courville, Deep Learning. Cambridge, MA: MIT Press,unpublished.