Cloud-based Video Surveillance System using EFD-GMM for Object Detection

Ce Li[1], Jianchen Su[1], Baochang Zhang[2, *]

[1] China University of Mining & Technology, Beijing
[2] Beihang University, Beijing
{celi,jcsu}@cumtb.edu.cn; *bczhang@buaa.edu.cn

Abstract. Nowadays, new generation of video surveillance systems integrates lots of heterogeneous cameras to collect, process, and analyze video for detecting the objects of potential security threats. The existing systems tend to reach the limit in terms of scalability, data access anywhere, video processing overhead, and massive storage requirements. A novel cloud computing can provide scalable and powerful techniques for large-scale storage, processing, and dissemination of video data. Furthermore, the integration of cloud computing and video processing technology offers more possibilities for efficient deployment of surveillance systems. This paper deploys the framework of a cloud-based video surveillance system and proposes an EFD-GMM approach for object detection in the overhead video processing. A prototype surveillance system is also designed to validate the proposed approach. It finally shows that the proposed approach is more efficient than GMM in video processing of cloud-based system.

Keywords: Video surveillance, cloud computing, object detection, EFD-GMM

1 Introduction

Modern video surveillance systems are composed of lots of heterogeneous cameras distributed over variety of sites [1]. The systems collect, process, and analyze different video streams to detect objects of potential security threats. Despite of significant benefit, there are important problems concerned in systems which are scalability, resource utilization, ubiquitous access, searching, processing, and storage to support large-scale surveillance. To solve the issues, a novel cloud-based surveillance systems [2] has been possessed and developed with improved processing capability and storage.

The existing work studies design and implement of the cloud-based surveillance system, for example, dependability characteristics [3], resource allocation [4], video recording [5], cloud storage mechanism [2], and cloud computing suitability for video surveillance [6]. However, there are also some significant research challenges to develop a cloud-based video surveillance system. For instance, the strategy for video acquisition and storage over the cloud, the technique for effective processing of video data. Therefore, a whole cloud-based video surveillance framework is in need to address two abovementioned challenges.

adfa, p. 1, 2011.

Some researchers have attempted potential directions for cloud-based video surveillance systems. But cost [7] and security [8] make some organizations difficulty to choose cloud-based solutions. Even some may argue that a cloud approach may seem not needed on account of strong local control in surveillance data acquired [9]. Nevertheless, with the availability of cloud-based video surveillance solutions and strong research on cloud technology, the signs of its potential growth become obvious. In this paper, the framework of a cloud-based video surveillance system is designed and deployed, and a novel approach is proposed for object detection in video processing.

Gaussian Mixture Model (GMM) has been widely applied in background model and video processing [15]. However, it have some weakness of convergence, sensitive to ambient noise and sudden light change, and prone to detect false target. Inspired by the edge information with efficient noise suppression, an improved algorithm based on Edge Frame Difference and GMM (EFD-GMM) is proposed to model the background and detect the moving object. The paper validates it on a deployed prototype surveillance system, and further discuss the detecting performance on two public datasets.

The remainder of this paper is organized as follows. Section 2 describes the related work, and section 3 introduces the framework of a cloud-based video surveillance system. Section 4 explains the proposed method for object detection in video processing. Experimental result and analysis are given in Section 5, and conclusion in Section 6.

2 Related Work

Video surveillance over cloud is an emerging research area. Literature review shows that there is a growing interest in adopting the cloud technology in this new area [10]. A cloud-based video surveillance system is first proposed in [2] with emphasis on storage. The paper analyzed the storage requirements of a cloud-based surveillance system different with the traditional one, and also investigated a secure cloud storage system and a video transmission optimization. Karimaa [3] studied the dependability of video surveillance technologies over cloud, such as the authority, security, maintainability and reliability characteristics of the cloud-based video surveillance solutions.

Recently, Neal [7] explored whether cloud computing is suitable for high-resolution video surveillance management system, and cloud computing is considered a suitable application for video surveillance management system. However, there are issues of cost and other threats to study. Next, Hossain [6] discussed the solutions of cloud-based video surveillance with some reservation to security and privacy aspects. In the paper [5], Lin not only designed a Hadoop distributed file system for recording system, but also provided store backups and monitoring features for video processing tasks. Cucchiara [11] concentrated on the deployment of a software to realize video service platform and object detection in multicamera surveillance system. Hossain proposed a dynamic resource allocation mechanism for service composition in cloud afterwards [4], and suggested that a number of virtual machines need to be optimally utilized for multiple surveillance services. All the above works demonstrate different aspects of system, and some especially focus on the strategy for video acquisition and storage to the cloud.

As for video processing in the cloud-based surveillance system, the results of object detection play important roles in providing information for better video processing. From the multiple cameras, images quality in video are always impacted even being corrupted by noise. There is a challenging work in object detection. Several previous works have been carried out on object detection, in particular like clustering approach, mean shift-based method [12], graph-based method [13], Bayesian-based method [14]. Gaussian Mixture Model (GMM) [15] is well known, but the main drawback of GMM is that the prior distribution does not depend on the pixel index and not on the spatial relationship between the labels of neighboring pixels. Thus, the object detection is extremely noise prone and illumination dependent, even prone to detect false target.

To overcome these disadvantages, Liu employed K-means clustering for GMM initialization to increase the function convergence rate [16], there is still the problems of poor anti-interference ability and easily noise prone when modeling background when coping with videos in complicated environment. Wei introduced three frame difference into GMM to restrain error detection rate of moving object [17]. However, this approach is difficult to adapt the light change and easily get incomplete object because of the three frame difference. Similarly, Mahnood [18] applied the frame difference in edge detection. Although it eliminate the influence of illumination change, the approach has low accuracy of detecting moving object for it is hard to get a complete object in the foreground. Overall, the above works improve the technique of effective video processing from different aspects, so this paper concentrates to develop a distinctive algorithm which is contribute to dealing with the issues of noise prone, illumination dependent and false target in detection.

3 Framework for a Cloud-based Video Surveillance System

In view of a cloud-based video surveillance system, there are several issues to explore for framework deployment. System requirements analysis, system architecture design, core system modules and system prototype deployment are described as follows.

3.1 System Requirements Analysis

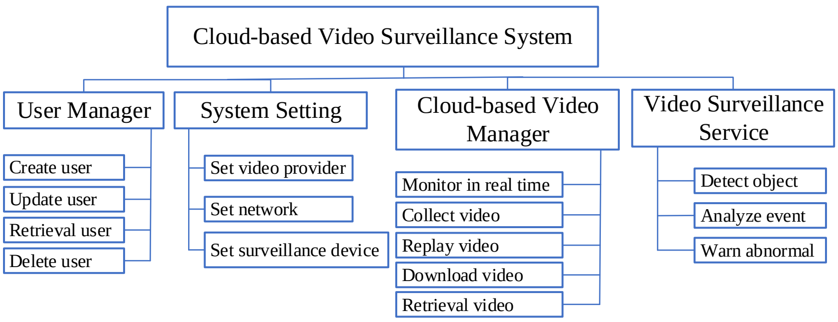


Fig. 1. The chart of system requirements analysis

By analyzing the cloud-based video surveillance system, four main function requirements are listed in Fig.1, user manager, system setting, cloud-based video manager and video surveillance service. User manager module is composed of the operations of creating, updating, retrieval, and deleting system users. The users usually access to the system by heterogeneous devices from anywhere and subscribe the abnormal event happening on heterogeneous video providers, so the system setting module is necessary to include the providers, networks and devices. Cloud-based video manager module is the overall management of cloud-based operations, which bridges between users and video surveillance service. Besides, the functions in video surveillance service include object detection, event analysis and abnormal warning.

3.2 System Architecture Design

Fig.2 shows the architecture of the designed cloud-based video surveillance framework. The system can be deployed on a private cloud in the system used by our single organization at present. Two core system modules are in the cloud, cloud video manager and video surveillance service. Located in the cloud side, notification and sharing mechanism is a vital part of the framework, which facilitates acquiring video streams as well as spreading the events of interest to the appropriate clients. It can provide high scalability for ubiquitous video surveillance service, connect and deliver video streams to various video providers and surveillance users [6]. In this architecture, some types of video providers exists, such as fixed cameras, IP cameras, and PTZ cameras. The video stream from multiple devices is transmitted to the cloud through the notification and sharing mechanism. System users with proper authentication can configure the connected devices and control video capturing and delivering.

3.3 Core System Modules

In the cloud architecture, there are two core system modules in Fig.2, cloud video manager and video surveillance service. The following paragraphs elaborate the modules.

Cloud Video Manager. In the framework, the overall management of the cloud based operations are cloud providers, resource allocation and monitoring statistic.

1. *Cloud Providers.* To encounter high demand for large storage and huge amount of data to process continuous videos [2, 3], cloud providers connect different types of network storage devices to meet specific requirements of system [19].
2. *Resource Allocation*. Cloud resources are computational in the form of virtual machines (VMs) [20]. Various VMs are managed and allocated to run associated services [21] by dynamically configuring capacities following the current workload.
3. *Monitoring Statistic*. Because of the readily available resource usage whenever, the monitoring statistic component is responsible for system monitoring and usage tracking of cloud resources, and provides statics of requests and usage cost.

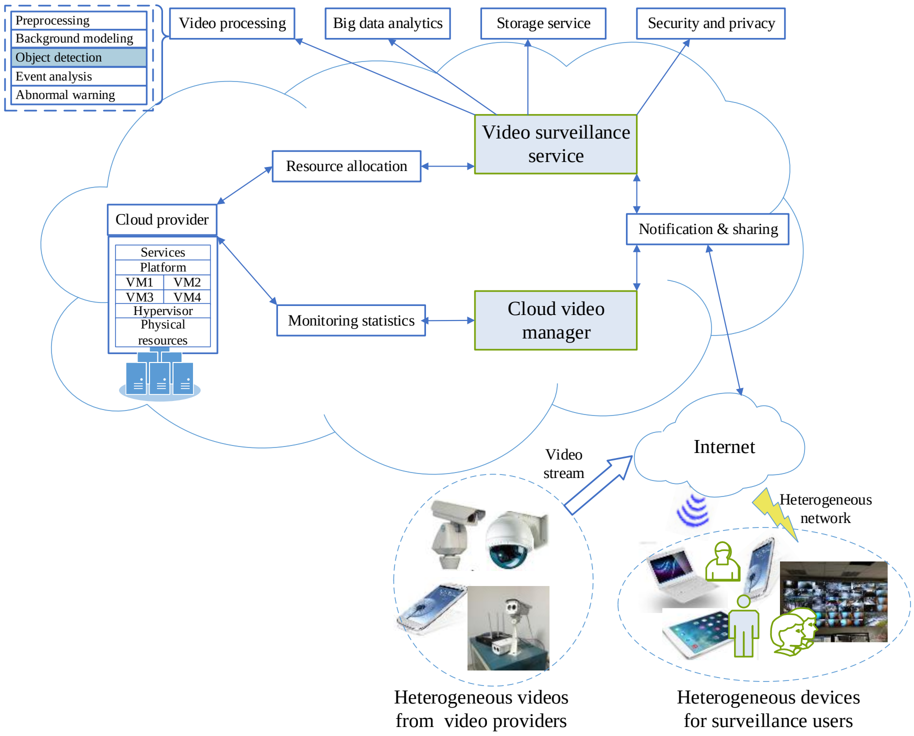


Fig. 2. The designed architecture for a cloud-based video surveillance system.

Video Surveillance Service. The system is a service oriented system in essence, so all function components are designed as services over the internet, including video processing, big data analytics, storage, security and privacy.

1. *Video Processing.* The cloud-based processing seems promising due to the enormous video processing capability that can be leveraged. The tasks of video processing are preprocessing, background modeling, object detection, event analysis, and abnormal warning, in Fig.2. Due to different video types, networks and devices, object detection is the first and foremost task to identify abnormal events and generate warnings.
2. *Big Data Analytics.* Big data analytics service is used to determine that whether it improves the fidelity of information or timelines of response.
3. *Storage Service.* Storage service provides a database for intelligent video processing. Important factors like vendor lock-in, disaster recovery capability, elasticity, and payment structure are concerned.
4. *Security and Privacy.* Security and privacy policy must be in place for deployment to secure videos, identify authentication and enforce privacy for access control.

3.4 System Prototype Development

To implement the function of framework, a prototype system is developed on a private cloud platform. Two instances are launched, one is to store the captured warning and querying information, and the other is for various web services. Dahua cameras are used to capture image and to connect the cloud. In the prototype system, the video processing service has been developed. For example, background modeling and object detecting are illustrated in Fig.3(a). Users is allowed to freely choose between continuous video and event record. In Fig.3(b), there is an example of the user browsing the warning list and event record from his or her view. Besides, web querying service is implemented on SQL server 2014.

(a) (b)

Fig. 3. (a) Real-time object detection in the system, (b) user browse the warning list and event record in the system.

4 Proposed Method for Object Detection in Video Processing

As mentioned above, object detection is one of the key task for video processing service in the system. During video processing, GMM has been widely applied in background model and object detection. However, it have some weakness of convergence, sensitive to ambient noise and sudden light change, and prone to detect false target. Therefore, to overcome the disadvantages, we improve traditional GMM in OpenCV library and propose a novel EFD-GMM approach for object detection.

The flowchart of method is described in Fig.4. Firstly, frame difference is introduced into GMM background model, which quickly distinguishes the background and moving region to extract the foreground. Then, the background model is mixed with the edge frame difference, and different updating rates are adopted during modeling to accelerate the speed of convergence for noise suppression and illumination independent. Finally, image AND operation is performed among the foreground information detected by background model, the blob information calculated by frame difference, and contour information gotten from edge frame difference. The approach has improvements on noise suppression, shadow removal and false target elimination, because of the processes of adding frame difference to GMM and integrating edge frame difference.

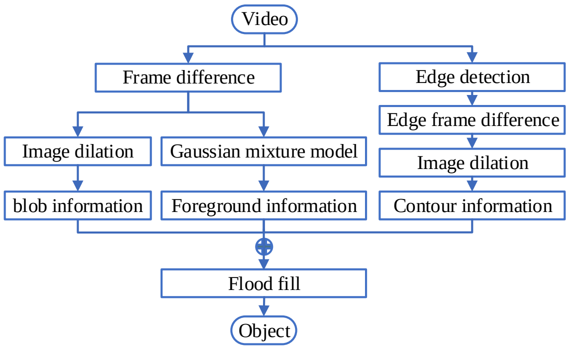


Fig. 4. Flowchart of proposed method.

4.1 Adding Frame Difference to GMM

GMM uses several frames to model background at the beginning. In the complicated scene, the background could hardly be modeled, for the background is shaded by the moving objects in most of the time. Thus, we consider the frame difference method [22] as the compensation for foreground extraction, and then rectify it using the blob information dilated from the frame difference image.

We first calculate the difference image among successive frames, using a threshold T to get a binary image to distinguish background and foreground coarsely. The process is described as follows. A count value C for each pixel ($0 \le C \le m$) is initialized as $m/2$ (m is the upper limit of C), two frames f_t and f_{t-1} are used to detect the foreground in

$$C = \begin{cases} C - 1/\beta & |f_t(i,j) - f_{t-1}(i,j)| \le T \\ C + 1 & |f_t(i,j) - f_{t-1}(i,j)| > T \end{cases}, \qquad (1)$$

$$\alpha = \frac{2C}{m} \times \alpha \qquad (0 \le C \le m)$$

where β is a coefficient determined by camera parameters. The pixel is identified as the background and the count C increases if the difference between two frames are not more than T, and vice versa. The updating rate α is dynamic according with the C. If the frame variation is great, α will increase. The threshold T consists of T_c and T_r. T_r is a fixed empirical value 30, and T_c is an optimum factor with the change of frame.

$$T = T_c + T_r, \quad T_c = \frac{1}{N} \sum_{i,j} |f_t(i,j) - f_{t-1}(i,j)|. \qquad (2)$$

Moreover, if the pixels changes rapidly in the video frame, GMM will be updated sooner with the updating rate α for background modeling. To construct GMM, each pixel is modeled by a mixture of K Gaussian distributions. The sample value of a certain pixel point $P(x,y)$ is $\{X_1, X_2, \ldots, X_t\}$, the probability of the present observed pixel value X_t in t^{th} time is

$$P(X_t) = \sum_{i=1}^{k} \omega_{i,t} \eta\left(X_t, \mu_{i,t}, \Sigma_{i,t}\right), \tag{3}$$

$$\eta(X_t, \mu_{i,t}, \Sigma_{i,t}) = \frac{1}{(2\pi)^{n/2} \left|\Sigma_{i,t}\right|^{1/2}} e^{-\frac{1}{2}(X_t - \mu_{i,t})^T \sum_{i,t}^{-1}(X_t - \mu_{i,t})}. \tag{4}$$

Where, K is the quantity of model components, and $\omega_{i,t}$, $\mu_{i,t}$ and $\Sigma_{i,t}$ are the weighted value, mean value and covariance matrix of the i^{th} Gaussian distribution of the model in t^{th} time, respectively. If the difference value between the present pixel and the background model is within a certain range, it can be considered as the background. That is, if it meet the condition $\left|X_t - \mu_{i,t-1}\right| > 2.5\sigma_{i,t-1}$, the model will be updated by

$$\begin{cases} \omega_{k,t} = (1-\alpha)\omega_{k,t-1} + \alpha\left(M_{k,t}\right) \\ \mu_t = (1-\rho)\mu_{t-1} + \rho X_t \\ \sigma_t^2 = (1-\rho)\sigma_{t-1}^2 + \rho\left(X_t - \mu_t\right)^T\left(X_t - \mu_t\right) \\ \rho = \alpha\eta(X_t \left| \mu_k, \sigma_k\right.) \end{cases}. \tag{5}$$

The updating speed of GMM mainly depends on learn rate α. $M_{k,t}$ is 1 for the matched model or 0 for the remaining distributions. If none of them match the current pixel, the least probable distribution will be replaced by a new with the current mean, the initialized high variance, and a low prior weight. α is determined by the frame difference process, the learning rate ρ, μ and σ are updated. Finally, for background estimation [24], we can sort order of K Gaussian distributions according to ω/θ, and suppose the first B distributions as the background models.

$$B = \arg\min\left(\sum_{k=1}^{b} \omega_{k,t} > T\right). \tag{6}$$

4.2 Integrating Edge Frame Difference

Although the combination of frame difference and GMM make the convergence of model better, the updating and estimation of model is still vulnerable to the sudden light change. We adopt the method of Canny edge detection [22] on frame difference image, then integrate the contour information from edge frame difference (EFD) to the foreground mask for sake of removing the false target. The basic idea is that the edge frame difference in false area is robust to the moving object in foreground, displayed in Fig.5. Surely, there is a drawback of edge frame difference that the foreground contour has some inner cavity. Therefore, we perform the image AND operation among the blob, foreground and contour information, then flood fill the foreground to get a complete object. In fact, the experimental result in the next section validate that the EFD and GMM are really mutual reinforcing to suppress noise, remove shadow and eliminate false target. Moreover, the performance of EFD-GMM can pave the basis for the next steps of event analysis and abnormal warning. In our system, a LIB linear SVM [23] classifier is learned to determine whether the event of object is abnormal or not.

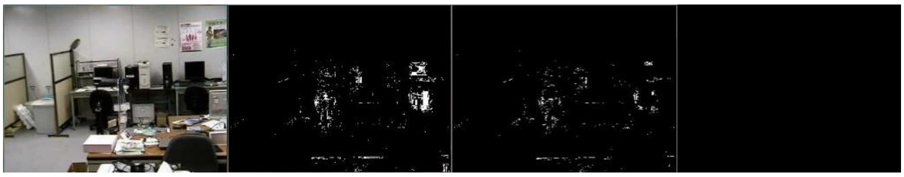
Fig. 5. The original image and edge frame difference image.

5 Experimental Result and Analysis

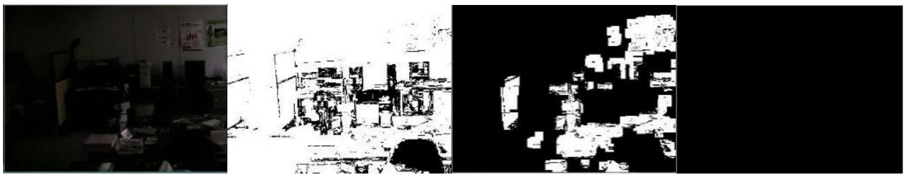
To verify the proposed algorithm, the comparison experiments are made in video sequences from different scenes. They are carried out on the client of the developed system prototype. We run the experiments on Pentium(R) E700@3.2GHz CPU unit, dual CPU core, 2GB memory, drive with 64-bit windows file system in Microsoft server. Two datasets belong to Kyushu University and Institute of Automation of Chinese Academy of Sciences are used to test various situations, such as sudden light changes, a pedestrian comes into the scenery and stay for a while, and ambient noise. Besides, we set the parameters of our algorithm K, α and T to 5, 0.03 and 0.85 in trials.

In Figure 6, we compare the original GMM [19], EFD [27] and the proposed EFD-GMM methods to validate the performance on an indoor video with quick lighting changes. The examples from the 25th frame, 541th frame, 857th frame in the original video are located in the first column (corresponding to Fig.6(a1), Fig.6(b1), Fig.6(c1), respectively). The results of object detection by original GMM are in second column in Fig.6, the results of EFD are in the third column, and the EFD-GMM results are in fourth column. It is obvious that the method of EFD-GMM gets the best performance in the extracted foreground and object detection, for we exclude lots of unnecessary moving points in background updating. Moreover, by comparison between the second and third column, EFD can indeed perform better than GMM when light is changing.

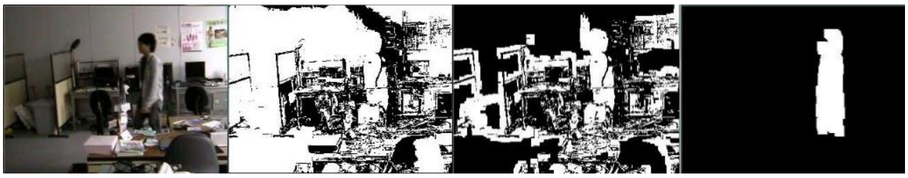
In Figure 7, we also compare the GMM, EFD and EFD-GMM to validate the performance on an outdoor video while a pedestrian starts walking instead of squatting. The example from the 100th frame in the original video are Fig.7(a), and the results of object detection by GMM, EFD and EFD-GMM are in Fig.7(b), Fig.7(c) and Fig.7(d), respectively. It shows that there is a false target around the object detected by GMM, because pixels of the long-time stationary target are gradually recognized as the background, and part of them is left in the foreground once the target moves suddenly. In contrast, our proposed EFD-GMM can update model timely to remove the false target.



(a1) (a2) (a3) (a4)

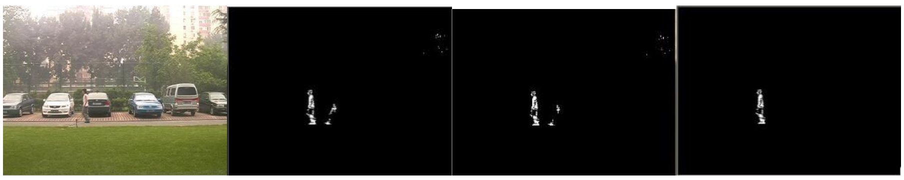


(b1) (b2) (b3) (b4)



(c1) (c2) (c3) (c4)

Fig. 6. The examples from the 25th frame, 541th frame, 857th frame with sudden light changes.



(a) (b) (c) (d)

Fig. 7. The example from the 100th frame while a pedestrian starts walking instead of squatting.

In Figure 8, the results in a complicated outdoor situation with waving leaves are compared. The example from the 331th frame in the video are Fig.8(a), and the results of object detection by GMM, EFD and EFD-GMM are in Fig.8(b), Fig.8(c) and Fig.8(d), respectively. In Fig.8(b) and Fig.8(c), the leaves are detected as the object, but they are ambient noise in fact. EFD-GMM is robust on this situation and has the ability of noise suppression in Fig.8(d) compared with the other methods. In total, 3338 frame videos are involved in the experiments. The object detection rate (DR) and false alert rate (FAR) of EFD-GMM is 98% and 2%, and DR has an increase of 14% and FAR has a decrease of 7%. To sum up, the proposed EFD-GMM method for object detection in video processing service of our developed system prototype.

(a) (b) (c) (d)

Fig. 8. The example from the 331th frame with waving leaves.

6 Conclusion

To integrate lots of heterogeneous cameras and analyze numerous video for detecting the objects of potential security threats, this paper develop a new cloud-based video surveillance system. Based on the system requirements analysis, the architecture framework is designed with two core system modules. In the video surveillance service, to fulfill the key task of video processing, a novel EFD-GMM approach is proposed for object detection. A prototype system is developed on the designed cloud architecture to validate the proposed approach. The experimental results show that the approach is effective and robust than GMM in complicated scenes from various video providers. However, the classifier for abnormal analysis needs improvement, and the cost remains a decisive factor to embrace the cloud-based surveillance. Therefore, future works may be directed to these issues for video surveillance application.

Acknowledgement. The work was supported in part by the Natural Science Foundation of China under Contract 61272052 and Contract 61473086, in part by PAPD, in part by CICAEET, and in part by the National Basic Research Program of China under Grant 2015CB352501. The work of B. Zhang was supported by the Program for New Century Excellent Talents University within the Ministry of Education, China, and Beijing Municipal Science & Technology Commission Z161100001616005. Baochang Zhang is the corresponding author.

References

1. Raty, T.D., "Survey on Contemporary Remote Surveillance Systems for Public Safety", IEEE Trans. on Systems, Man and Cybernetics C: Applications and Reviews, vol. 40, no. 5, pp. 493–515, 2010.
2. Ren, Y.J., Shen, J., Wang, J., Han, J., Lee, S., "Mutual Verifiable Provable Data Auditing in Public Cloud Storage", Journal of Internet Technology, vol. 16, no. 2, pp. 317-323, 2015.
3. Karimaa, A., "Video Surveillance in the Cloud: Dependability Analysis", Proc. of 4th Int. Conf. on Dependability, pp. 92–95, 2011.
4. Hossain, M.S., Hassan, M.M., Qurishi, M.A., Alghamdi, A., "Resource Allocation for Service Composition in Cloud-based Video Surveillance Platform", Proc. of IEEE Int. Conf. on Multimedia and Expo Workshops, pp. 408–412, 2012.

5. Lin, C.F., Yuan, S.M., Leu, M.C., Tsai, C.T., “A Framework for Scalable Cloud Video Recorder System in Surveillance Environment”, Proc. of 9th Int. Conf. on Ubiquitous Intelligence & Computing and Autonomic & Trusted Computing, pp. 655–660, 2012.
6. Hossain, M.A., “Analyzing the Suitability of Cloud Based Multimedia Surveillance Systems”, Proc. of 15th IEEE Int. Conf. on High Performance Computing and Communications, 2013.
7. Neal, D., Rahman, S., “Video Surveillance in the Cloud”, Journal of Cryptography and Information Security, vol. 2, no. 3, 2015.
8. Sabahi, F., “Cloud Computing Security Threats and Responses”, Proc. of 3rd Int. Conf. on Communication Software and Networks, pp. 245– 249, 2011.
9. Venters, W., Whitley, E.A., “A Critical Review of Cloud Computing: Researching Desires and Realities”, Journal of Information Technology, vol. 27, no. 3, pp. 179–197, 2012.
10. Hossain, M.A, “Framework for a Cloud-based Multimedia Surveillance System”, Journal of Distributed Sensor Networks, 2014.
11. Cucchiara, R., Prati, A., Vezzani, R., “Designing Video Surveillance Systems as Services”, Proc. of 2nd Workshop on Video Surveillance Projects, Italy, 2011.
12. Li, J., Li, X.L., Yang, B., Sun, X.M., “Segmentation-based Image Copy-move Forgery Detection Scheme”, IEEE Trans. On Information Forensics and Security, vol. 10, no.3, pp.507-518, 2015.
13. Shi, J., Malik, J., “Normalized Cuts and Image Segmentation”, IEEE Trans. on PAMI, vol. 22, no. 8, pp. 888–905, 2000.
14. Li, S.Z., “Markov Random Field Modeling in Image Analysis”, Springer, 2009.
15. Zhang, B.C., Perina, A., Li, Z.G., Murino, V., Liu, J.Z., Ji, R.R., “Bounding Multiple Gaussians Uncertainty with Application to Object Tracking”, IJCV, 2016.
16. Zhang, B.C., Li, Z.G., Perina, A., Bue, A.D., Murino V., “Adaptive Local Movement Modelling (ALMM) for Object Tracking”, IEEE Trans. CSVT, 2016.
17. Zhang B.C., Perina, A., Murino, V., Bue, A.D., “Sparse Representation Classification with Manifold Constraints Transfer”, Proc. of CVPR, pp.4557-4565, 2015.
18. Mahmood, A.M., Maras, H.H., Elbasi, E., “Measurement of Edge Detection Algorithms in Clean and Noisy Environment”, Proc. of 8th Int. Conf. on Application of Information and Communication Technologies, pp. 1-6, 2014.
19. Shen, R., “Building a Cloud-enabled File Storage Infrastructure”, Tech Republic White Paper, F5 Network, 2013.
20. Sotomayor, B., Montero, R.S., Lorente, I.M., Foster, I., “Virtual Infrastructure Management in Private and Hybrid Clouds”, IEEE Internet Computing, vol. 13, no. 5, pp. 14–22, 2009.
21. Beloglazov, A., Buyya, R., “Energy Efficient Resource Management in Virtualized Cloud Data Centers”, Proc. of 10th IEEE ACM Int. Symposium on Cluster, Cloud, and Grid Computing, pp. 826–831, 2010.
22. Liu, X.D., Yu, Y., Liu, B., Li, Z., “Bowstring-based Dual-Threshold Computation Method for Adaptive Canny Edge Detector”, Proc. of Int. Conf. of Image and Vision Computing, pp. 13-18. New Zealand, 2013.
23. Gu, B., Sheng, V.S., Tay, K.Y., Romano, W., Li, S., “Incremental Support Vector Learning for Ordinal Regression”, Trans. Neural Netw. Learn. Syst., vol. 26, no. 7, pp.1403-1416, 2015.
24. Pan, Z.Q., Kwong, S., Sun, M.T., Lei, J.J., “Early MERGE Mode Decision Based on Motion Estimation and Hierarchical Depth Correlation for HEVC”, IEEE Trans. On Broadcasting, vol. 60, no. 2, pp. 405-412, 2014.