Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 46 (2025), No. 2, pp. 553–562
DOI : 10.47974/JIOS-1959

Optimizing the deep features using the spectral concept for an efficient video classification

Chowdam Leelavathi *
Department of Computer Science and Engineering
School of Engineering
Dayananda Sagar University
Bangalore
Karnataka
India

and

Department of Computer Science and Engineering - Business Systems
Rajeev Gandhi Memorial College of Engineering & Technology
Nandyal
Andhra Pradesh
India

Gousia Thahniyath †
Department of Computer Science and Engineering
School of Engineering
Dayananda Sagar University
Bangalore
Karnataka
India

K. Rajendra Prasad §
Department of Computer Science and Engineering (Data Science)
Institute of Aeronautical Engineering
Hyderabad
Telangana
India

---

* *E-mail:* **chowleela.04@gmail.com (Corresponding Author)**
† *E-mail:* **gousia-cse@dsu.edu.in**
§ *E-mail:* **krprgm@gmail.com**

Abstract

The identification of suspicious activity is greatly influenced by the efficient learning and classification of video data. Learning large videos requires more processing time than conventional deep learning methods, because deep features have relatively large dimensions. In the present investigation, a spectral-based deep learning method for producing reduced or lower dimensional representations of surveillance videos is developed. There are two essential steps in the proposed method. In order to identify suspicious activity, the first method involves transforming the deep characteristics of video data into lower-manifold spectral space using classification techniques. The lower-rank representation of the video data is found by computing a Laplacian matrix. To show the effectiveness of the suggested strategy in comparison to current methods, experiments are conducted on benchmarked video surveillance datasets.

Subject Classification: *68P30, 68Q30.*

Keywords: *Video classification, Deep learning, Spectral concept, Data sparsity, Curse of dimensionality.*

1. Introduction

Detecting suspicious activity for real-time video surveillance is becoming more popular in recent years for applications related to societal security [1]. Identifying objects and learning their behaviors are the most difficult problems in the classification of suspicious films. One of the most effective approaches for video classification jobs is deep learning [2]. Video frames can have their features extracted batch-wise using current deep learning techniques like VGG16 [3], ResNet50 [4], DenseNet121 [5], MobileNet [6], GoogleNet [7], and EfficientNetB0 [8]. The video data can then be learned in accordance with these features. Many movies that can be used to increase the classification rate or detection efficiency of suspicious behaviors in video surveillance require practical training. After the deep characteristics are extracted, classifier models are trained to classify suspicious activities. The deep features extracted from each video frame are enormous; each movie is arranged with vast frames, and each frame's feature size is enormous as well. Processing the deep characteristics to identify suspicious activity in these situations could lead to real-world scalability problems with regard to the important calculation time metrics. According to others, the main issue with current deep-based classifier models is the curse of dimensionality [8]. Within the current deep-based classifier models, fantastic classifier models have been employed, such as random forest (RF) [9], support vector classifier (SVC) [10], and k-nearest neighbor classifier (kNN) [11].

Initially, the affinity matrix of frame features was created, and then the Laplacian matrix was calculated. The suggested spectral-based deep classifier method (SDCT) consists of the Laplacian matrix (LM) calculations. The number of components ('k') of the subspace of deep features is determined by the features of the lower-rank frame, while the LM investigates the frame's deep features in terms of a large number of Eigenvectors. The first k-largest Eigenvectors are taken to derive the deep features subspace. The subspace of obtained k-Eigen vectors is called spectral space or low-dimensional manifold. The reduced representation of video frames' deep features is indicated by the spectral space. This spectral space preserves the video frame information while mapping the huge size of deep features of frames onto the lower manifold subspace. Instead of using high-dimensional frame features, the top k-Eigen vectors are sufficient to define deep features of the frame data. Lastly, videos are classified using the RF, SVC, and kNN classifier models based on the spectral properties of frames. Three hybrid SDC-VM variations are created using these suggested methods, utilizing the RF, SVC, and KNN models. To overcome the issue of the curse of dimensionality, suggested hybrid variations employ spectral features rather than the deep features of the high-dimensional frame.

The structure of the paper is as follows. Section 2 delineates the literature review of the topic; Section 3 outlines the suggested SDC-VM and visual mining methodologies. Section 4 presents the experimental examination and discussion of the work. Section 5 presents the conclusion and future scope of the effort.

2. Literature Study

Video surveillance has emerged as a crucial application for public safety. The foremost scientific challenge in surveillance is the identification and categorization of anomalous behaviors. In this procedure, the primary concerns in contemporary video analytics research are classification accuracy and computing scalability. Authors in [12], [13], [14], [15], and [16] endeavored to identify or categorize loitering, fighting, and abandoned things by analyzing suspicious behaviors. The majority of these studies utilize offline video surveillance data; in real-time scenarios, post-detection of irregularities or anomalies in videos may prove ineffective. Consequently, the author in [17] devised a real-time blob-matching technique by identifying the temporal characteristics of blobs and activities. It is capable of detecting and categorizing altercations, theft, loitering, fainting, and

similar incidents. Lu and Grauman [18] developed an algorithm for video summary synthesis that extracts specific sub-videos from lengthy recordings based on picture quality parameters to depict the events occurring within the video.

Another approach relies on the key frame sequence, utilizing the event detection mechanism established in [19]. It utilizes frames instead of capturing the video directly. It demonstrates a substantial decrease in video content and enhanced effectiveness in video retrieval. Identifying precise instances within lengthy videos is mostly addressed by the aforementioned methods. Shaohua Wan et al. [20] devised the superframe segmentation algorithm, which eliminates unnecessary frames in a video sequence to mitigate computational burdens during video anomaly classification. The findings and methodology are outlined below, along with the limitations of the study.

Several researchers [21], [22] employed unsupervised methodologies for anomaly identification. It necessitates meticulous classification and management strategies [25], [26] for anomalies. In contrast to unsupervised methods, deep learning-based categorization demonstrates remarkable capabilities in identifying anomalies. Sultani et al. [24] created the 3D convolutional network (C3D), which captures spatiotemporal features and evaluates anomaly scores using a three-layer fully connected network. This results in a significant issue: the scalability of video segmentation prior to feature extraction, owing to the substantial size of the video frames. Yumna Zahid [25] addresses the video segmentation challenge with the ensemble method of IBaggedFCNet, wherein bagging enforces rigorous segmentation and employs the Inception-v3 deep classifier for video classification.

3. Proposed SDCT

The proposed approach employs spectral techniques to mitigate the impact of data sparsity during video categorization. The spectral approach comprises two fundamental steps: i) Determine the affinity matrix, which calculates the weighted matrix for frame features while accounting for affinities. The Laplacian matrix is calculated to provide the Eigen decomposition that reveals appropriate low-dimensional manifolds (or subspaces). Equations (1) through (7) demonstrate the derivation of Eigenvectors, commonly referred to as spectral space, in the proposed study.

Frames and it can be computed with affinity values of F_i, F_j (i.e. $Sigma_i$ and $Sigma_j$)

$$Deep_F = \{F_1, F_2, \ldots, F_n\} \tag{1}$$

$$Sigma_i = d(F_i, F_k) \tag{2}$$

$$d_{ij} = d(F_i, F_j) \text{ and } d_{ji} = d(F_j, F_i) \tag{3}$$

$$W \in F_{nxn}, W = e^{((-d_{ij} * d_{ji})/(\text{Sigma}_i * \text{Sigma}_j))} \tag{4}$$

$$d_{ii} = \Sigma_{i=1}^{n} w_{ij} \tag{5}$$

$$L = D^{-1/2} W D^{-1/2} \tag{6}$$

The values acquired in L are normalized and optimal for subsequent classification or clustering processes. The magnitude of L grows substantial, rendering it unnecessary to depict the high-dimensional space of deep features. In L, the columns denote the Eigenvectors. The dimensions of L become n x n. The initial eigen column vectors in L are typically designated as the largest vectors. We are now selecting the k-largest eigenvalue column vectors for the n frames of deep features. These can be depicted using the stacks of k-largest Eigenvectors, referred to as spectral space 'SV.' The singular value (SV) represented the ideal low-dimensional manifold subspace for the high-dimensional deep features of n frames, characterized by a size of n x k, where k signifies the decreased dimensionality of the spectrum space or the low-dimensional manifold subspace.

The proposed spectral approach for video frames achieves optimal feature representation by extracting the k-largest Eigenvectors, where k denotes the number of major components in spectral space. SV delineates the ideal characteristics of the frame. Upon acquiring the extensive batches of frames, randomly select n batches and input them into the deep layer, adhering to a batch size of 24 (maximum sequence length) and frame dimensions of 224 × 224. The algorithm delineates these phases from step 1 to step 4. Step 5 employs the following deep models independently: VGG16, ResNet50, DenseNet121, MobileNet, GoogleNet, and EfficientNetB0 to extract deep features, with each frame yielding deep features of 2048 dimensions. A significant issue of data sparsity has arisen in the deep features. Consequently, the spectral approach was employed to transform the high-dimensional deep features into a spectral space with reduced data sparsity.

4. Results and Discussion

The experiments are conducted on the 14 classes of the benchmarked video dataset, UCF-CRIME, which is publically available in [23]. Upon commencing the pre-processing, the video for each class is segmented into distinct frames and systematically categorized according to the training and testing labels. Deep features are extracted utilizing the Densenet, Googlenet, Mobilenet, EfficientNetB0, Resnet, and VGG16 models. Additionally, these characteristics are projected onto lower-dimensional manifolds by extracting the principal components (or eigenvectors) using the eigen decomposition of the Laplacian matrix. The reduced dimensional attributes are derived from our proposed SDCT. Table 1 illustrates the outcomes of current deep-based classifier models (i.e., Deep Random Forest, Deep-SVC, and Deep-kNN) alongside the proposed spectral-based deep classifier video models (SDCT-Random Forest, SDCT-SVC, SDCTkNN). The assessment metrics of accuracy, precision, recall, and f-measure are illustrated in [24]

Table 1

Performance Values for the Existing and Proposed Methods

Performance Measure	Models	Deep Random Forest	SDCT-RF	Deep-SVC	SDCT-SVC	Deep-KNN	SDCT-KNN
Accuracy	Densenet	0.8464	0.9607	0.9429	0.9893	0.525	0.6464
	Googlenet	0.8107	0.9786	0.8964	0.9964	0.5893	0.6571
	Mobilenet	0.8786	0.9607	0.9536	0.9929	0.6107	0.6929
	EfficientNetB0	0.8107	0.8464	0.2786	0.7036	0.4071	0.5182
	Resnet	0.8607	0.9821	0.9536	0.9929	0.6107	0.6964
	VGG16	0.8321	0.9393	0.5536	0.9429	0.5464	0.6536
Precision	Densenet	0.8614	0.9642	0.9474	0.9907	0.6185	0.7014
	Googlenet	0.8331	0.9799	0.9091	0.9966	0.6904	0.7361
	Mobilenet	0.8847	0.9664	0.9588	0.9935	0.6522	0.725
	EfficientNetB0	0.8597	0.8317	0.2051	0.7222	0.356	0.3984
	Resnet	0.8775	0.9828	0.9588	0.9935	0.6522	0.7303
	VGG16	0.8578	0.9518	0.6119	0.9504	0.5395	0.5543

Contd...

Recall	Densenet	0.8464	0.9607	0.9429	0.9893	0.525	0.6464
	Googlenet	0.8107	0.9786	0.8964	0.9964	0.5893	0.6571
	Mobilenet	0.8786	0.9607	0.9536	0.9929	0.6107	0.6929
	EfficientNetB0	0.8464	0.8107	0.2786	0.7036	0.4071	0.4871
	Resnet	0.8607	0.9821	0.9536	0.9929	0.6107	0.6964
	VGG16	0.8321	0.9393	0.5536	0.9429	0.5464	0.5536
F-Score	Densenet	0.8432	0.9609	0.9436	0.9895	0.5245	0.6413
	Googlenet	0.8082	0.9785	0.8982	0.9964	0.5837	0.6558
	Mobilenet	0.8775	0.9619	0.9545	0.9929	0.6059	0.6967
	EfficientNetB0	0.8115	0.8447	0.2135	0.6946	0.3574	0.3968
	Resnet	0.8628	0.982	0.9545	0.9929	0.6059	0.7013
	VGG16	0.8299	0.941	0.543	0.9437	0.5071	0.5267

The experimental performance scores indicated that proposed SDCT attained a commendable classification rate relative to current deep learning classifier models, owing to the mitigation of the data sparsity issue in the suggested spectral-based deep classifier models. The accuracy value increased by 10 to 12% in the proposed models relative to the existing models. Correspondingly, the values for precision, recall, and F-score enhanced by 6% to 10%, 5% to 11%, and 4% to 12%, respectively.

5. Discussion and Conclusion

Classifications of video surveillance are becoming essential for societal applications, particularly in the realm of public safety. Threatening, spurious, and other anomalous behavior classifications are advancing in computer vision research. Deep models are the most effective methodologies for video categorization. Nonetheless, there are concerns pertaining to the data sparsity of deep features in video frames. The deep features possess high dimensionality and exhibit sparsity challenges. The proposed SDCT strategy utilizes efficient spectral methods to map deep features to spectral features, thereby mitigating the impact of the sparsity issue. The reduced sparsity of deep features may affect the attainment of a high classification rate for video abnormalities. The results unequivocally demonstrate that spectral-based deep classifier models enhanced classification performance metrics by 7 to 11% relative to deep-based classifier video models. The suggested spectral-based deep models utilize a singular perspective of subspace learning, and there exists potential to

enhance the SDCT through multi-view subspace learning for advancements in future video categorization.

References

[1] E. Şengönül, R. Samet, Q. Abu Al-Haija, A. Alqahtani, B. Alturki, and A. Alsulami, "An Analysis of Artificial Intelligence Techniques in Surveillance Video Anomaly Detection: A Comprehensive Survey," *Appl. Sci.*, vol. 13, no. 8, p. 4956 (2023). [Online] Available: https://doi.org/10.3390/app13084956.

[2] R. Savran Kızıltepe, J. Q. Gan, and J. J. Escobar, "A Novel Keyframe Extraction Method for Video Classification Using Deep Neural Networks," *Neural Comput & Applic*, vol. 35, pp. 24513–24524 (2023). [Online] Available: https://doi.org/10.1007/s00521-021-06322-x.

[3] N. A. Shelke and S. S. Kasana, "Multiple Forgery Detection in Digital Video with VGG-16-Based Deep Neural Network and KPCA," *Multimed. Tools Appl.*, vol. 83, pp. 5415–5435 (2024). [Online] Available: https://doi.org/10.1007/s11042-023-15561-0.

[4] N. Nallappan and R. Velswamy, "Exploring Deep Learning-Based Content-Based Video Retrieval with Hierarchical Navigable Small World Index and ResNet-50 Features for Anomaly Detection," *Expert Syst. Appl.*, vol. 247 (2024), [Online] Available: https://doi.org/10.1016/j.eswa.2024.123197.

[5] B. Li, Y. Karaca, D. Baleanu, Y. D. Zhang, O. Gervasi, and M. Moonis, "Facial Expression Recognition by DenseNet-121," in *Multi-Chaos, Fractal and Multi-Fractional Artificial Intelligence of Different Complex Systems*, Academic Press (2022). [Online] Available: https://doi.org/10.1016/B978-0-323-90032-4.00019-5.

[6] L. Liu, X. Wang, Q. Bao, and X. Li, "Behavior Detection and Evaluation Based on Multi-frame MobileNet," *Multimed. Tools Appl.*, vol. 83, pp. 15733–15750 (2024). [Online] Available: https://doi.org/10.1007/s11042-023-16150-x.

[7] V. Singh, A. Baral, R. Kumar, S. Tummala, M. Noori, S.V. Yadav, S. Kang, and W. Zhao, "A Hybrid Deep Learning Model for Enhanced Structural Damage Detection: Integrating ResNet50, GoogLeNet, and Attention Mechanisms," *Sensors* (2024). [Online] Available: https://doi.org/10.3390/s24227249.

[8] Z. Mutalova, A. Shaushenova, A. Nurpeisova, M. Ongarbayeva, A. Ispussinov, S. Bekenova, and Z. Altynbekova, "Development of a Mathematical Model for Detecting Moving Objects in Video Streams in Real-Time," *IEEE Access*, vol. 12, pp. 169235–169246 (2024). [Online] Available: https://doi.org/10.1109/ACCESS.2024.3487783.

[9] S. Rajagopal, M. Uma Devi, G. Maria Jones, and M. Gomathy Nayagam, "Ensemble Random Forest-Based Gradient Optimization Based Energy Efficient Video Processing System for Smart Traffic Surveillance System," *IETE J. Res.*, vol. 70, no. 9, pp. 7175–7191 (2024). [Online] Available: https://doi.org/10.1080/03772063.2024.2350927.

[10] X. Wang, "Support Vector Machine-Based Video Anomaly Detection Approaches," in *Anomaly Detection in Video Surveillance. Cognitive Intelligence and Robotics*, Springer, Singapore (2024). [Online] Available: https://doi.org/10.1007/978-981-97-3023-0_7.

[11] X. Wang, "K-Nearest Neighbor-Based Video Anomaly Detection Approaches," in *Anomaly Detection in Video Surveillance. Cognitive Intelligence and Robotics*, Springer, Singapore (2024). [Online] Available: https://doi.org/10.1007/978-981-97-3023-0_4.

[12] N. D. Bird, O. Masoud, N. P. Papanikolopoulos, and A. Isaacs, "Detection of Loitering Individuals in Public Transportation Areas," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 167–177, Jun. (2005).

[13] N. Bird, S. Atev, N. Caramelli, R. Martin, O. Masoud, and N. P. Papanikolopoulos, "Real-Time, Online Detection of Abandoned Objects in Public Areas," in *Proc. IEEE ICRA*, pp. 3775–3780 (2006).

[14] L. Sijun, Z. Jian, and D. Feng, "A Knowledge-Based Approach for Detecting Unattended Packages in Surveillance Video," in *Proc. IEEE AVSS*, p. 110 (2006).

[15] S. Blunsden and R. B. Fisher, "The BEHAVE Video Dataset: Ground Truthed Video for Multi-Person Behavior Classification," *Annu. BMVA*, vol. 2010, no. 4, pp. 1–11 (2010).

[16] S. Blunsden, E. Andrade, and R. Fisher, "Non-Parametric Classification of Human Interaction," in *Proc. 3rd Iberian Conf. Pattern Recog. Image Anal.*, Part II, Girona, Spain, pp. 347–354 (2007).

[17] M. Elhamod and M. D. Levine, "Automated Real-Time Detection of Potentially Suspicious Behavior in Public Transport Areas," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 688–699 (2013).

[18] Z. Lu and K. Grauman, "Story-Driven Summarization for Egocentric Video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 2714–2721 , Jun. (2013).

[19] W. Wolf, "Key Frame Selection by Motion Analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. Conf.*, pp. 1228–1231, May (1996).

[20] S. Wan, X. Xu, T. Wang, and Z. Gu, "An Intelligent Video Analysis Method for Abnormal Event Detection in Intelligent Transportation Systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4487–4495, Jul. (2021). [Online] Available: https://doi.org/10.1109/TITS.2020.3017505.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Proc. Adv. Neural Inf. Process. Syst.*, pp. 1097–1105 (2012).

[22] Y. S. Chong and Y. H. Tay, "Abnormal Event Detection in Videos Using Spatiotemporal Autoencoder," in *Proc. Int. Symp. Neural Netw.*, Cham, Switzerland: Springer, pp. 189–196, Dec. (2017).

[23] W. Sultani, C. Chen, and M. Shah, "Real-World Anomaly Detection in Surveillance Videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 6479–6488 (2018).

[24] P. K. Mishra, A. Mihailidis, and S. S. Khan, "Skeletal Video Anomaly Detection Using Deep Learning: Survey, Challenges, and Future Directions," *IEEE Trans. Emerging Topics Comput. Intell.*, vol. 8, no. 2, pp. 1073–1085, Apr. (2024). [Online] Available: https://doi.org/10.1109/TETCI.2024.3358103.

[25] N. Janu, A. Kumar, L. Raja, V. Bhatnagar, A. Kumar, S. K. Ramesh, and C. Poonia, "Development of an Efficient Real-Time H.264/AVC Advanced Video Compression Encryption Scheme," *J. Discrete Math. Sci. Cryptogr.*, vol. 24, no. 8, pp. 2245–2255 (2022).

[26] G. Saini, S. Bhatnagar, L. Raja, S. Sharma, and R. C. Poonia, "Structural Equation-Based Model to Investigate the Moderating Effect of Fear of COVID Using Partial Least Square Method," *J. Interdiscip. Math.*, Taylor & Francis, 22 Feb (2022).

Received November, 2024