J. Electrical Systems 20-4s (2024): 353-362

[1]PingXu

[2]Aziguli Abudouwaili

Identification of Unusual Visitor Activity Using a Hybrid Gaussian Model in Picturesque Locations

Abstract: - The management of tourists in picturesque locations is becoming increasingly crucial due to the rapid growth of the tourism industry. The subsequent issue, though, is the rise in strange visitor behaviours, like carelessly crossing hazardous locations and doing illicit graffiti, among other things. These strange behaviours harm tourist destinations' reputations while also raising the possibility of safety hazards. Because of this, this paper uses a regional feature analysis approach to detect abnormal pedestrian behaviour. Specifically, we begin by modelling the background, obtain the moving target region information through moving target detection, and use the minimum outer rectangular box as the regional feature. Lastly, we compute the aspect ratio of this rectangle, perform curve fitting and prediction, and finally conclude the abnormal behaviour detection process. The method's simplicity and speed are demonstrated by the experimental findings.

Keywords: video surveillance; area characteristics; abnormal behaviour

I. INTRODUCTION

The tourism industry has been growing in recent years, and the system for regulating scenic areas has been improved. However, in these tourist destinations, there are frequently instances of abnormal behaviour, such as people randomly climbing rocks or scrawling, crossing over a railing into a scenic area that is not open, falling off a lakeside over a railing into the water, and so forth. At the same time, many visitors are aware of these incidents from the news reports that highlight the abnormal behaviour[1-3]. Dangerous and uncivilised behaviours are examples of abnormal behaviour in tourist sites. Uncivilised actions include drawing and scribbling everywhere. Among the risky habits are falling into water, going over railings, and climbing everywhere. These are typical aberrant habits found in tourist destinations. The identification of abnormal behaviours by tourists and the prompt sending of early warnings to staff have become increasingly common as a result of the frequent occurrence of these behaviours. In contrast, manual identification of abnormal behaviours through real-time surveillance videos is labor-intensive and inefficient, and behavioural identification technology has not kept up with the diverse range of application scenarios[4-6]. People are starting to focus on machine learning as a means of maximising the role that specific behaviour recognition technology can play. By applying machine learning techniques to the human body, specific behaviour recognition technology can be applied to achieve rapid recognition as well as increased recognition accuracy [7]. The use of machine learning techniques and behaviour recognition technologies for the identification of anomalous behaviour in tourism scenarios can enhance the manpower-saving, real-time, accurate, and supervisory system of beautiful locations.

In the tourism sector, keeping an eye on and controlling unusual visitor behaviour in picturesque locations

[1] School of TAIYUAN TOURISM COLLEGE Taiyuan City, Shanxi Province 030031

[2]Haian (Beijing) Data Technology Co., Ltd，Beijing 100023

has always been a challenging issue. Monitoring blind zones and a lack of personnel are two issues with traditional monitoring methods, which primarily rely on manual patrols [8]. The advancement of computer vision and pattern recognition technologies has made it increasingly practical to use video surveillance for real-time monitoring of tourists in picturesque locations. Enhancing detection accuracy is largely dependent on the backdrop model, which is one of the fundamental components of video anomaly detection.

The Gaussian model is a traditional background modelling technique that has been extensively researched and used in the field of video anomaly detection. Nevertheless, the conventional Gaussian model is prone to erroneous detection and performs poorly against dynamic backdrops and lighting changes [9]. Therefore, enhancing Gaussian models to adapt to complex and dynamic scenic settings is one of the key areas of study in current research. The attractive area is a public open space with a complex and varied environment, and the behaviour of its visitors is very erratic. More intelligent and adaptive background models are required to handle these issues because traditional background models are prone to failure when dealing with high numbers of tourists and complex lighting circumstances. The primary focus is on a few particular behaviours found in tourist destinations, such as climbing, painting, and overtopping[10-12]. However, in a particular scene, particularly in the tourism scene, the abnormal behaviour of tourists recognition technology is still relatively rare. Many researchers have used the machine learning method of the human body in the scene of the behaviour recognition technology in a variety of scenes, while in real-time and accuracy to carry out continuous improvement, the technology has become the focus of attention of countless researchers in recent years.

In light of this, the goal of this work is to optimise the hybrid Gaussian model in order to increase the precision and stability of identifying anomalous tourist behaviour in picturesque locations. By utilising multi-moment information fusion and incorporating deep learning techniques, a more intelligent backdrop model is built. This research aims to develop precise backdrop models in intricate scenic area settings so that anomalous tourist activity may be tracked in real time. Better detection accuracy, a lower false alarm rate, and the capacity to adjust to various scenarios and challenging lighting conditions are the benefits.

II. IDENTIFICATION OF ABNORMAL PEDESTRIAN BEHAVIOUR USING AREA FEATURES

When it comes to scenic video surveillance, the camera is typically fixed, meaning that the background is largely constant. A single Gaussian model and a mixed Gaussian model have been developed to characterise the statistical distribution for each background pixel, which follows a certain statistical distribution on the time axis. The mixed Gaussian model is more comprehensive.

2.1 Hybrid Gaussian model

A hybrid Gaussian model is used to model each sequence of pixel points over time $\{X_1, X_2, ..., X_n\}$, with the probability of a pixel value at the current observation defined as.

$$P\left(X_t\right)=\sum_{n=1}^{K}\omega_{n,t}\eta\left(X_t,\mu_{n,t},\Sigma_{n,t}\right) \quad (1)$$

Where: $\omega_{n,t}$ is the weight of the nth Gaussian model at time t, satisfies; K is the number of Gaussian models, which typically takes a value of 3 to 5.

The mean and variance of the nth Gaussian model at time t are, respectively, $\sum_{n=1}^{K}\omega_{n,t}=1$,t=1, $\mu_{n,t}$, and $\Sigma_{n,t}$.

The Gaussian probability density function is represented by η.

$$\eta=\left(\mathrm{X_t},\mu,\Sigma\right)=\frac{1}{2\pi^{\frac{n}{z}}\,|\,\Sigma\,|^{\frac{1}{2}}}\mathrm{e}^{-\frac{1}{2}(\mathrm{X_t}-\mu_\mathrm{t})^{\mathrm{T}}\Sigma^{-1}(\mathrm{X_t}-\mu_\mathrm{t})} \quad (2)$$

For reasons of computational complexity, this is usually taken to be $\Sigma_{n,t}=\sigma_n^2\mathrm{I}$.

Since the background of a long-running video programme will inevitably change, each hybrid model's distribution model needs to be updated [13]. A new model is generated and put at the end of the model queue, marking the pixel as foreground, if an observed pixel value does not match any of the existing models. When an observed pixel value matches one of the existing models, the weight of that model is increased, the variance of that model is decreased, and the mean is recalculated.The hybrid model was revised as follows by Stauffer and Grimso n：

$$\begin{aligned}
\omega_{\mathrm{n,t}}&=(1-\alpha)\omega_{\mathrm{n,t-1}}+\alpha\mathrm{M_{n,t}}\\
\mu_{\mathrm{n,t}}&=\left(1-\rho_{\mathrm{n,t}}\right)\mu_{\mathrm{n,t-1}}+\rho_{\mathrm{n,t}}\mathrm{X_t}\\
\sigma_{\mathrm{n,t}}^2&=\left(1-\rho_{\mathrm{n,t}}\right)\sigma_{\mathrm{n,t-1}}^2+\rho_{\mathrm{n,t}}\left(\mathrm{X_t}-\mu_{\mathrm{n,t}}\right)^{\mathrm{T}}\left(\mathrm{X_t}-\mu_{\mathrm{n,t}}\right)
\end{aligned} \quad (3)$$

where the learning rate, or $\rho_{\mathrm{n,t}}=\alpha\eta\left(\mathrm{X_t}|\ \mu_{\mathrm{n,t}},\sigma_{\mathrm{n,t}}\right)$, α , indicates how quickly updates in the background. The model matching operator, $\mathrm{M_{n,t}}$, returns 1 in cases where a new pixel matches and 0 in cases where it does not.

2. 2 The lowest aspect ratio of an exterior rectangle

Since the aspect ratio of the moving target's minimal outer rectangle is highly sensitive to the shadows cast by the target and is directly related to the motion target detection algorithm, we employ hybrid Gaussian background modelling in HSV space [14].The following is the formula for converting between RGB and HSV modes:

$$\begin{aligned}
&\mathrm{V}=\frac{\mathrm{R+G+B}}{3}\\
&\mathrm{H}=\frac{1}{360}\left[90-\mathrm{Arc\,tan}\left(\frac{\mathrm{F}}{\sqrt{3}}\right)+\{0,\mathrm{G}>\mathrm{B};180,\mathrm{G}<\mathrm{B}\}\right] \quad (4)\\
&\mathrm{S}=1-\frac{\min(\mathrm{R,G,B})}{\mathrm{V}}
\end{aligned}$$

where， $F = \frac{2R - G - B}{G - B}$.

After the foreground, that is, the motion target is segmented by the mixed Gaussian model, and after morphological filtering and connected component analysis, a binary image can be obtained containing only the motion target with a minimum outer rectangular box outside the target, and then mapped to the original image as shown in Figs. 1 (a) and (b). In HSV mode, when a pixel is covered by shadows, its luminance varies greatly, while the chromaticity signal does not change much. Thus, the adaptive mixed Gaussian model method in the literature [15] is used.

a b

Figure 1 Experimental Results

The aspect ratios for pedestrians who are standing, jogging, and sitting were extracted in this work; the outcomes are displayed in Fig. 2.

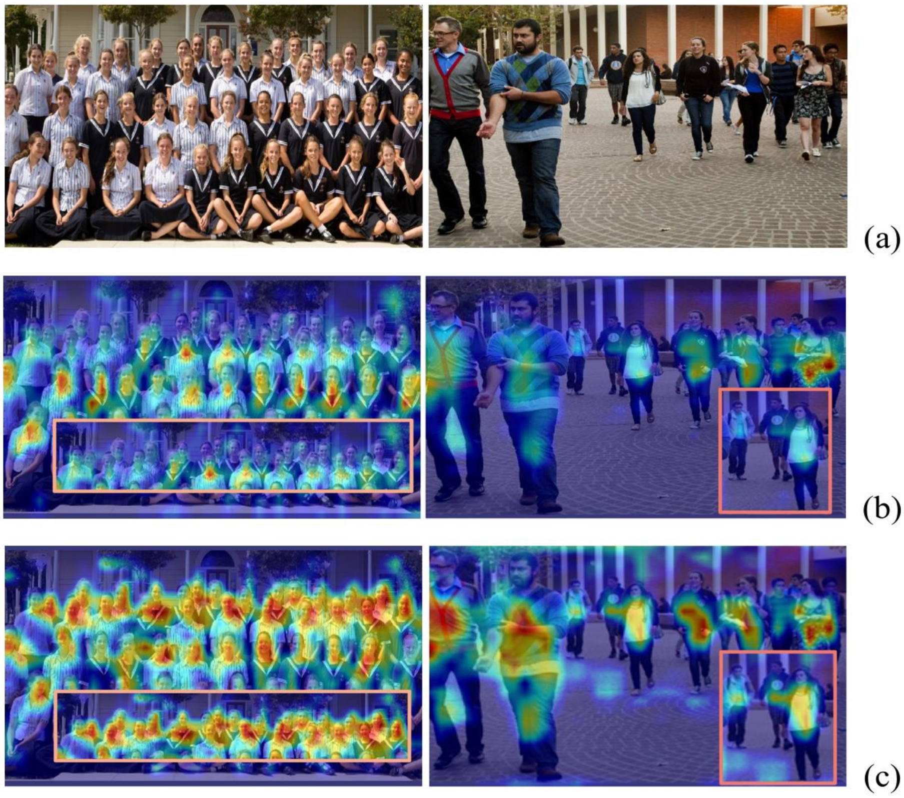

Figure 2 Test Results

Based on experiments, this paper adopts this feature for the detection of pedestrian behaviour. Specifically, it shows that in video sequences, the change of the minimum outer rectangular box of normal pedestrians is analytical. For example, when walking normally, the aspect ratio of this rectangular box is periodic; when running is similar to walking, but the magnitude and frequency of the change are different.

2.3 Curve fitting

The fitted function for periodic functions often has the following form:

$$f(x) = a + \sum_{k=1}^{K} \left[b_k \sin(kx) + c_k \cos(kx) \right] \quad (5)$$

This research employs the fitting form below in light of the calculation's intricacy and real-time nature：

$$\mathrm{f}(\mathrm{x}) = \mathrm{a} + \mathrm{b}_1 \sin\left(\mathrm{c}_1\mathrm{x} + \mathrm{d}_1\right) + \mathrm{b}_2 \cos\left(\mathrm{c}_2\mathrm{x} + \mathrm{d}_2\right) \quad (6)$$

We applied the least squares criterion and restricted the fitting to the $1.5 < \mathrm{a} < 10, 0 < \mathrm{b}_1, 0 < b_2$ restriction in light of the real circumstances.

III. RESULTS

In order to test the system's recognition capability, the ROIs of segmented sequential infrared images were fed into the Le Net-7 system, which was built with the intention of creating CPU-fast-running convolutional neural networks. The system ran on an Intel i7 9750H processor running at 4. 5 GHz with 32 GB of RAM. The testing in this work is conducted using the Ohio State University-provided OTCBVS Benchmark Dataset database [16] and Terravic Motion IR Database database [17]. All ROIs are normalised to 32×32 to ensure that the test set and training set are of the same size.

Infrared person detection investigations are carried out using three distinct sets of infrared images. Test set 1 is a multi-person test set that contains pedestrians adhering to each other (shading), and it is derived from the OSU Thermal Pedestrian Database. It comprises of 23 photos with 101 human targets. The second test set, which comes from the Terravic Motion IR Database Database, has 230 photos totaling 460 human targets—including pedestrians who are obstructing one another—each of which has two armed targets. Test set 3 is a single human test set with 200 photos, each having a single human target, taken from the OSU Color-Thermal Database database. Figure 4 displays a subset of the detection results across the three test sets. The detected pedestrians' location region is indicated by the green rectangular box. Figure 4 shows that even in cases when there is a slight obstruction between the human bodies, the system is still able to accurately identify the human targets in the consecutive infrared images, demonstrating the method's viability and effectiveness.

The following performance metrics were selected to gauge the system's competence and display the detection effect [18].

$$\text{Detection rate} = \frac{TP}{TP+FP} \times 100\%$$
$$\text{False alarm rate} = \frac{FN}{TP+FN} \times 100\% \quad (7)$$

The variables TP, FP, and FN represent the number of correctly detected human targets, wrongly detected human targets, and non-human targets that were mistakenly categorised as pedestrians. The approach is contrasted with the conventional Fast R-CNN model and SVM classification coupled with the Histogram of Orientation Gradient (HOG). The FastR-CNN open source structure typically consists of three types: VGG-CNN-M1024, ZF, and VGG16 [19–20]. Each type has a deeper network level, with the deepest VGG16 network being used for comparison with the well-established Le Net-7 network. Table 1 presents the findings.

Table 1 Experimental Results

Test set	Different methods	Number of pedestrians	TP	FP	FN	Detection rate/%	False alarm rate	Detection time/ms
1	HOG+SVM	100	97	3	2	98	1.8%	402.52
	Fast R-CNN		99	2	1	97	1.0%	31.06
	Proposed method		100	2	1	98	1%	14.89
2	HOG+SVM	450	450	8	0	99	0.2%	319.05
	Fast R-CNN		452	6	0	98.8	0	28.56
	Proposed method		455	4	0	99.9	0	14.5
3	HOG+SVM	200	200	0	0	100	0	204.6
	Fast R-CNN		200	0	0	100	0	28.23
	Proposed method		200	0	0	100	0	10.26

This paper's method has a lower false alarm rate and a higher detection rate than the two classic methods, as Table 1 illustrates. HOG+SVM and Fast R-CNN perform identically to this approach in the easier test set 3 (one-line pedestrian target). The detection rate of the method presented in this study is higher and no false alarm occurs in test set 2, where there are more examples of adhesion and occlusion between pedestrians. While the Fast R-CNN model is hampered by the environment, the detection rate is lower and there is one false judgement of pedestrian targets in test set 1 (multiple pedestrian targets), which is a complex environment, where the detection rate of this method is the highest and there is no false judgement of pedestrians[21].

Because fixed-size, single-shape feature vectors are used as the input for linear classification and because there are too many traversal windows when the image is in a complex context, the classic HOG+SVM has the longest detection time. The detection time is significantly decreased when convolutional neural networks are used in place of traditional feature vector matching. When this paper's method is compared to Fast R-CNN, the detection time is considerably lower, which can cut the detection time of Fast R-CNN by around half. The detection time of this paper's method is approximately one-third of Fast R-CNN's in the single-row human test set 3, suggesting that this paper's method significantly improves detection time (see Fig. 3).

Figure 3 Partial experimental results

Fig. 4(a) and (d) displays the segmentation results of the detection of human behaviour in the picturesque environment.

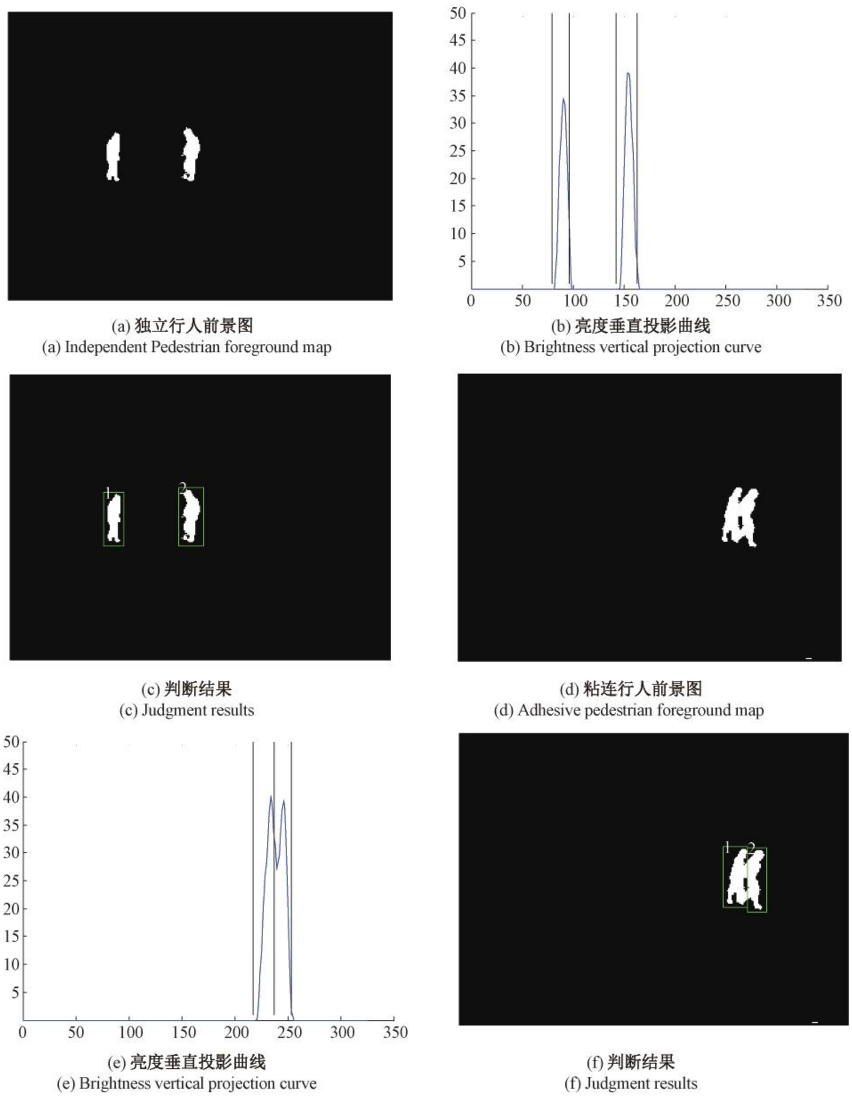

Figure 4 Differences in pedestrian characteristics

As seen in Fig. 4(b),(e), the image segmented by Gaussian model thresholding is vertically projected to the X-axis, yielding the image's grey scale vertical projection curve. Figure 4(b) The human body's greyscale region is depicted as a high peak devoid of any sticking phenomena. In order to obtain a series of luminance bands perpendicular to the X-axis, the midpoints of the ascending and descending curves are calculated quantitatively on both sides of the raised peaks in the projection curves. These two midpoints are then used as the start and end points of a luminance band, respectively, and the areas where the human body may be present are included in the luminance bands. Figure 4(e) The vertical projection curves require additional processing since they are sticking[22,23].

When the brightness bands produced by vertical and horizontal projection are simultaneously placed in the appropriate locations inside the original image, the image can be divided into numerous high-brightness rectangular sections, as seen in Fig. 4(c).

The ROIs of independent, unobstructed objects may be found using the aforementioned technique, and the trained convolutional neural network can then use the ROIs to detect pedestrians. Nevertheless, this approach is not precise enough to handle the situation where pedestrians obstruct one another (Fig. 4(d)), where the system's detection rate is low. In order to identify the ROIs of occluded pedestrians, this research uses the directional projection method based on Gaussian background modelling.

IV. CONCLUSION

Regarding the scenic spot surveillance system, identifying abnormal tourist behaviour demands very high real-time and accuracy requirements; in contrast, the old method has great accuracy but is difficult to achieve in real-time. This study is expected to provide more intelligent and effective monitoring means for the management of the scenic spot, effectively improve the level of safety management of the scenic spot, and give tourists a safer and more enjoyable tour experience through the optimisation of the Gaussian model of the background model in the scenic spot tourists abnormal behaviour detection through the application of in-depth research.

Funding

Project type: Shanxi Province art science planning project

Project name: Empirical research on the integration of culture and tourism to enhance the value of smart tourism

Project number: 23AA008

REFERENCES

[1] Korn, J. M., Kuruvilla, F. G., McCarroll, S. A., Wysoker, A., Nemesh, J., Cawley, S., ... & Altshuler, D. (2008). Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nature genetics*, *40*(10), 1253-1260.

[2] Zheng, X., Han, J., & Sun, A. (2018). A survey of location prediction on twitter. *IEEE Transactions on Knowledge and Data Engineering*, *30*(9), 1652-1671.

[3] Visuwasam, L. M. M., & Raj, D. P. (2023). Spatio Temporal Tourism Tracking System Based on Adaptive Convolutional

Neural Network. *Comput. Syst. Sci. Eng.*, *45*(3), 2435-2446.

[4] Bao, J., Zheng, Y., Wilkie, D., & Mokbel, M. (2015). Recommendations in location-based social networks: a survey. *GeoInformatica*, *19*, 525-565.

[5] Zhao, E., Du, P., Azaglo, E. Y., Wang, S., & Sun, S. (2023). Forecasting daily tourism volume: A hybrid approach with CEMMDAN and multi-kernel adaptive ensemble. *Current Issues in Tourism*, *26*(7), 1112-1131.

[6] Kaiser, M. S., Lwin, K. T., Mahmud, M., Hajializadeh, D., Chaipimonplin, T., Sarhan, A., & Hossain, M. A. (2017). Advances in crowd analysis for urban applications through urban event detection. *IEEE Transactions on Intelligent Transportation Systems*, *19*(10), 3092-3112.

[7] Chen, W., Wang, W., Yin, H., Zhao, L., & Zhou, X. (2023). HFUL: a hybrid framework for user account linkage across location-aware social networks. *The VLDB Journal*, *32*(1), 1-22.

[8] Chen, W., Wang, W., Yin, H., Zhao, L., & Zhou, X. (2023). HFUL: a hybrid framework for user account linkage across location-aware social networks. *The VLDB Journal*, *32*(1), 1-22.

[9] Wang, X., Zhao, Y. L., Nie, L., Gao, Y., Nie, W., Zha, Z. J., & Chua, T. S. (2014). Semantic-based location recommendation with multimodal venue semantics. *IEEE Transactions on Multimedia*, *17*(3), 409-419.

[10] Bharathi, G., & Anandharaj, G. (2023). Classification of review information based on tourist's places of interest in India using support vector machine. *International Journal of Business Information Systems*, *44*(2), 202-218.

[11] Beheshti, R., & Sukthankar, G. (2015). A hybrid modeling approach for parking and traffic prediction in urban simulations. *AI & society*, *30*, 333-344.

[12] Wang, S., Yu, Y., Chen, J., & Liu, J. (2023). Impact of climate change on cherry blossom viewing tourism: analysis and simulation based on Weibo proxy data. *Current Issues in Tourism*, *26*(5), 718-734.

[13] Afyouni, I., Al Aghbari, Z., & Razack, R. A. (2022). Multi-feature, multi-modal, and multi-source social event detection: A comprehensive survey. *Information Fusion*, *79*, 279-308.

[14] Rawat, Y. S., & Kankanhalli, M. S. (2016). ClickSmart: A context-aware viewpoint recommendation system for mobile photography. *IEEE Transactions on Circuits and Systems for Video Technology*, *27*(1), 149-158.

[15] Xiang, T., & Gong, S. (2008). Incremental and adaptive abnormal behaviour detection. *Computer Vision and Image Understanding*, *111*(1), 59-73.

[16] Antonakaki, P., Kosmopoulos, D., & Perantonis, S. J. (2009). Detecting abnormal human behaviour using multiple cameras. *Signal Processing*, *89*(9), 1723-1738.

[17] Xiang, M., Rao, H., Tan, T., Wang, Z., & Ma, Y. (2019). Abnormal behaviour analysis algorithm for electricity consumption based on density clustering. *The Journal of Engineering*, *2019*(10), 7250-7255.

[18] Yu, W. (2021). Research on feature extraction of vehicle abnormal driving behaviour based on 5G internet of vehicles. *International journal of vehicle design*, *86*(1-4), 124-142.

[19] Lloyd, K., Rosin, P. L., Marshall, D., & Moore, S. C. (2017). Detecting violent and abnormal crowd activity using temporal analysis of grey level co-occurrence matrix (GLCM)-based texture measures. *Machine Vision and Applications*, *28*, 361-371.

[20] Jung, J. H., Ha, K. S., Chae, M., Kang, W. S., Lee, G. D., & Lee, S. H. (2010). Abnormal behaviour in colour tracking characteristics of the fringe-field switching liquid crystal display. *Liquid Crystals*, *37*(3), 303-309.

[21] Ali, Jehad, Rutvij H. Jhaveri, Mohannad Alswailim, and Byeong-hee Roh. "ESCALB: An effective slave controller allocation-based load balancing scheme for multi-domain SDN-enabled-IoT networks." Journal of King Saud University-Computer and Information Sciences 35, no. 6 (2023): 101566.

[22] Ren, X., Ahmed, I., & Liu, R. Study of Topological Behavior of Some Computer Related Graphs. Journal of Combinatorial Mathematics and Combinatorial Computing, 117, 3-14.

[23] C. Zhang, M. Li and D. Wu, "Federated Multidomain Learning With Graph Ensemble Autoencoder GMM for Emotion Recognition," in IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 7, pp. 7631-7641, July 2023, doi: 10.1109/TITS.2022.3203800.