Optical Flow Statistics for Violent Crowd Behavior Detection

Pallavi D. Chakole[a] and Vishal R. Satpute[b]

Image Processing and Computer Vision Laboratory, Department of Electronics and Communication Engineering, Visvesvaraya National Institute of Technology, Nagpur, Maharashtra, 440010, India

Keywords: Human Activity Detection, Crowd Behavior Analysis, Anomaly Detection, Optical Flow, Similarity Index, Correlation Coefficients.

Abstract: We proposed an approach for identifying human violent behavior by evaluating the optical flow of a series of sequences obtained from a video. The term Violent or Violence refers to an event that arises, causing of unexpected displacement of a crowd. "Crowd Behaviour Analysis" is an important research topic that falls under the area of image processing and computer vision, machine learning, and deep learning, which have been investigated by researchers. Proceeding with this attitude, a simple and novel method based on the amount of movement present in the current frame with respect to its previous frame has to be presented here. The methodology employed is as follows: the optical flow of two consecutive frames will be calculated. Further, correlation coefficients will be calculated by considering the magnitude of the optical flow of successive frames. From those correlation values, we can know how much the successive frames are similar or correlated. High correlation coefficients pointed that, there will be less movement in the crowd, a lower rate of change of velocity, and thus normal behavior or non-violent event. On contradictory if the correlation coefficients seem to be low, there will be more movement in the crowd, a high rate of change of velocity, and thus abnormal behavior or violent event detected. Decision criteria have to be set for a particular threshold value that has been selected adaptively, below which we can get violent events. Implementation has to be done on MATLAB R2021b, using the UMN video dataset consisting of 11 videos of three different scenarios. Evaluation results concluded that the proposed methodology can able to detect violent anomalies somehow accurately.

1 INTRODUCTION

We are constantly insecure as the population grows, human behavioural elements diversify, and highly crowded settings emerge all around us. People require a security guard to counteract this insecurity. However, keeping a constant eye out for suspicious activity, especially in crowded areas, is a difficult assignment for a guard. As a result of this, surveillance cameras are being developed. However, there are a few flaws. When it comes to CCTV surveillance, we save real-time recordings in our database, and anytime something suspicious occurred, we searched these databases for reasons and actions. Finding anomalies in a busy environment, however, remains difficult.As a result, it is critical to design a real-time suspicious activity detection system or anomaly detection system for constantly monitoring crowded places, crowd management, and preventing anomalies in advance.

In response to these disadvantages, researchers and practitioners in the fields of image processing and computer vision, machine learning, and deep learning devote themselves to detecting such anomalies in human behaviour. Detecting unusual humanity in public locations is a significant challenge that everyone must investigate. Crowds are common in public places such as public places, financial institutions, and roadways in the modern era. Again, large crowds may attend public events such as gatherings, concerts, sports, protests, and demonstrations. There is always a systemic risk, insecurity, and management required. However, gathering management and security appear to be ineffective.We are always looking for isolated incidents in real-time, but all we have are recorded videos. When something unexpected occurs in public places, we go into these stored database systems to figure out what happened. How did it happen? Who should be held accountable? However, the damage had already been done at this point. We always glance for an automotive crowd behavioral assessment system that takes all of these factors into account.

[a] https://orcid.org/0000-0001-7714-2005
[b] https://orcid.org/0000-0001-9944-9489

Chakole, P. D. and Satpute, V. R.
Optical Flow Statistics for Violent Crowd Behavior Detection.
DOI: 10.5220/0013628300004664

In *Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025) - Volume 3*, pages 557-564
ISBN: 978-989-758-763-4

Paper is organized in a sequence as, literature review, proposed methodology, result analysis, conclusion and at last references are listed.

2 LITERATURE REVIEW

Researchers must be able to recognize human behavior from video sequences. Behavior recognition is concerned not only with human behaviors but also with their mental capacities or psychological state, as well as their facial expressions and gestures. Due to several real-time issues such as occlusion, crowded surroundings, illumination effect, scaled variations, changes in views, and so on, recognizing actions from movies or steady frames is once again a difficult task (Parate et al., 2018). It's a multimodal investigation of human activity recognition. (Vrigkas et al., 2015) proposes unimodal and multimodal modalities-based work. The novel expresses how the influence of human behavior is dependent not just on individuals' behavior but also on group behavior, their interactions, associations, body language, and interaction with objects. By differentiating the terms 'Activity' as a sequence of actions related to body part movements, and 'Behavior' as a movement of body parts along with facial expressions, moods, and gestures. The hierarchical structure of human activities has been structured with respect to different layers in (Zhang et al., 2017). Layers are categorized from bottom to top as 'action primitives layer', 'action/activity layer', and 'complex interactions layer'. Adopting a different approach to detect activities of human Kinect sensors has been used (Nale et al., 2021) which can able to extract features as per the user's need. Skelton-based unusual activities can be detected by (Franco et al., 2020). Another approach to extract features to detect suspicious human activities has been implemented in (Abrishami Moghaddam and Zare, 2019) by defining the term 'spatiotemporal wavelet correlogram' work based on correlation of wavelet coefficients with respect to space and time. This method is overcome for the background subtraction method to detect foreground objects. The term, 'Groups' and 'Crowd' are defined based on the number of individuals involved, interactions within an individual, their velocity, and direction of movements. According to (Murino et al., 2017) 'Groups' are those where two or more people are involved with some interactions, share the same spatial and temporal adomain, as well as same velocity and direction of movements. On the other hand, 'Crowds' are those where more than two individuals are involved, with no interaction, with different spatial and temporal domains, with different velocities and directions of movement. Based on the definition of groups and crowds the evaluation of activity detection has been done by means of adopting two different approaches as 'Microscopic' and 'Macroscopic' approaches (Bour et al., 2019). The microscopic method is concerned with groups in which each individual behavior is treated separately.In juxtaposition, the macroscopic approach is concerned with crowds, where the entire crowd is regarded as a single entity. A summary of an anomalous group and crowd behavior analysis has been summarised in (Afiq et al., 2019) (Sawarbandhe et al., 2019). In his review authors differentiates various detection techniques that have been implemented in the last 5 year. The methods has been categorised and subcategorized as 'Gaussian of Mixture Model (GMM)' (Chavan et al., 2018)(Naveen et al., 2014), 'Hidden Markov Model (HMM)' (Satpute et al., 2014)(Gangal et al., 2014a)(Gangal et al., 2014b), 'Optical Flow (OF)' (Nayan et al., 2019)(Chen and Lai, 2019) and 'Spatio-temporal techniques (STT)'. Optical flow with the multiresolution concept has been proposed in (Meinhardt-Llopis and Sánchez, 2012). Additionally, the work-based of people aggregation based on groups and crowd has been explained in (Mohammadi et al., 2016b). In this work, the authors categorized behavior analysis strategy as 'model-based strategy' and 'motion-based strategy'. Under the model-based approach, some algorithm has to de learned by the model to do a specific task as per users' requirement, such as 'Social Force Model (SFM)' (Mehran et al., 2009), 'Behavior Heuristic Model (BHM)' (Mohammadi et al., 2016a). Another approach motion-based behavior analysis has been done by means of optical flow vectors (Horn and Schunck, 1981) (Lucas et al., 1981), which can able to gives information based on the amount of motion has to de done frame by frame in a video. for example, violence flow method (Hassner et al., 2012), the substantial derivative method based on fluid mechanics (Mohammadi et al., 2015), optical flow analysis based on divcurl(Chen and Lai, 2019), entropy (Cheggoju et al., 2021), correlation coefficients (Nayan et al., 2019), histogram of magnitude and momentum (Bansod and Nandedkar, 2020). Detection and classification has been done by different techniques by different authors like, SVM (Patil and Biswas, 2017)(Thombare et al., 2021), Bag of Visual Words (BoW)(Sharma and Dhama, 2020), tracking techniques (Jirafe et al., 2021), (Ghutke et al., 2016). An 'Intelligent video surveillance' (Sreenu and Durai, 2019) deep learning method-based review for crowd behavior analysis, object detection(Gajbhiye et al., 2017), violent detection has been summarized from different journals and years, along with challenges (Gupta et al., 2016), motivations (Pawade et al., 2021), applications, drawbacks, and results.

3 PROPOSED METHODOLOGY

The methodology has been proposed shown in the figure.1, starting with taking a video as an input. From video, frames are extracted for further optical flow analysis. Optical flow is the method that can able to give information about movements present within successive frames in the form of motion vectors (motion in x and y direction), magnitude of motion vectors, and orientation of motion. The further process has been done by considering the magnitude of optical flow only. Correlation can be estimated between two successive frames. Correlation can give how much the two frames are similar to each other. For smaller motion, we get large correlation coefficients and for larger motion, we get small correlation coefficients. Further differences of correlation coefficients are calculated to know about the amount of motion. Depending on the value of differences of the magnitude of correlation coefficients of the successive frame we get the frame number from which motion gets start increases. Accordingly, from that frame number, we can select the correlation coefficient value as a threshold value to set decision criteria. By comparing with this selected threshold we can able to detect violent and non-violent events successfully. Step by step process is explained below in detail.

3.1 Background Subtraction

As shown in figure 1, initially CCTV video is taken as an input. For the CCTV system before getting started initial frame is captured as the background reference frame. From that reference frame, each extracted current frame gets subtracted to get foreground objects as shown in figure2.

$$O(x,y) = C(x,y) - B(x,y) \tag{1}$$

Where $O(x,y)$ is the foreground object frame, that we got by differences of $C(x,y)$ current frame, and $B(x,y)$ background frame for corresponding pixels coordinates x and y.

3.2 Optical Flow

As an estimation of motion-based approach method based on optical flow, Horn & Schunk (Horn and Schunck, 1981), Lucas & Kanade (Lucas et al., 1981) methods are well known. Optical flow is able to detect movements within consecutive frames. Optical flow is a global estimation, which means calculations are done pixel by pixel. It is based on some assumptions like ‘Brightness constancy, Spatial coherence, and Temporal persistence’. By means of brightness constancy, the brightness of the small area remains unchanged even if, the area gets displaced by a small amount. Due to the property of special coherence, the velocity of nearby points in the scene is usually the same since they belong to the same surface. Over time, the picture motion of a surface patch varies due to temporal persistency.

For a video sequence I(x,y,t), by assuming brightness constancy, pixel intensities remains same over time ie.

$$I(x,y,t) = I(x+dx, y+dy, t+dt) \tag{2}$$

By applying Taylor series expansion, some sort of substitutions, and Crammer’s rule on eq (2) can be written in the form of,

$$I_x \frac{dx}{dt} + I_y \frac{dy}{dt} + It = 0 \tag{3}$$

Further substitutions for $\frac{dx}{dt} = u$ and $\frac{dy}{dt} = v$ as motion vectors for x and y directions respectively in eq (3),

$$I_x u + I_y v + I_t = 0 \tag{4}$$

This is nothing but an optical flow constraint equation for motion vectors u and v respectively for x and y directions.

Further u and v are calculated by iterative methods applied on minimized cost function of brightness constancy term along with the smoothness factor.

$$u^{n+1} = u^{-n} - I_x\left(\frac{I_x u^{-n} + I_y v^{-n} + I_t}{\alpha^2 + I_x{}^2 + I_y{}^2}\right) \tag{5}$$

$$v^{n+1} = v^{-n} - I_y\left(\frac{I_x u^{-n} + I_y v^{-n} + I_t}{\alpha^2 + I_x{}^2 + I_y{}^2}\right) \tag{6}$$

where n is no of iterations, u^{-n} is local average velocity of u for an n^{th} iteration, which can be used for subsequent iteration i.e. $n+1$ iteration. Simillarly v^{-n} is local average velocity of v for an n^{th} iteration. α is the scale factor to be used for smoothness constraint.

The number of iterations gets stopped if convergence happened or stopping criteria gets fulfilled, whichever gets earlier (Meinhardt-Llopis and Sánchez, 2012). Further magnitude and orientation can be calculated by using u and v getting from eq.(5) and eq.(6) as

$$Mag = \sqrt{u^2 + v^2} \tag{7}$$

$$\theta = tan^{-1}\left(\frac{v}{u}\right) \tag{8}$$

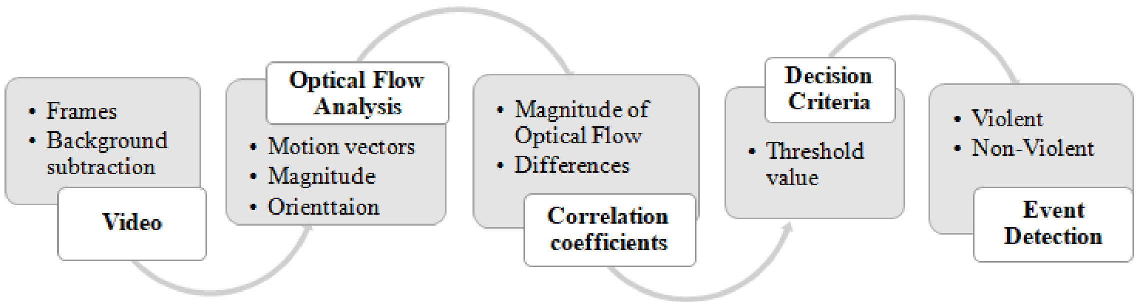

Figure 1: Flow for proposed methodology

(a) (b) (c)

Figure 2: Background Subtraction, a)Background Frame, b)Current Frame, c)Foreground object frame.

As of now by applying optical flow on image sequences we can know about motion present in a present image with respect to the previous image. u, v, Mag, θ can able to give information about motion in the x-direction, motion in the y-direction, total motion, and direction where motion happened respectively. This information can be used further for predictions about events.

3.3 Correlation Coefficients

By considering the optical flow magnitude of a series of sequences we can find out the correlation coefficients between the present image and the next consecutive image. Correlation coefficients can tell about, how much similar kind of motion that the images have? Higher the value of the coefficient, similar or slow in motion. Lower the value of coefficients,fast in motion. Correlation coefficients of an image I_n with respect to I_{n+1} can be calculated by eq.(9).

$$R_{\mathrm{n,n+1}} = \frac{\sum\sum(I_{\mathrm{n}}(x,y)-\mu_n)^2(I_{\mathrm{n+1}}(x,y)-\mu_{n+1})^2}{\sqrt{(\sum\sum(I_{\mathrm{n}}(x,y)-\mu_{\mathrm{n}})^2)(\sum\sum(I_{\mathrm{n+1}}(x,y)-\mu_{\mathrm{n+1}})^2)}} \tag{9}$$

Here in eq.(9) R used to represent correlation coefficients between image I_n and I_{n+1}. μ_n is the mean of image I_n and μ_{n+1} is the mean of image I_{n+1}.

As we know correlation coefficients give an idea about how much motion is present inside each image. Here in this paper events may be considered as sudden and quick movements within a crowd. That can be observed by calculating differences of consecutive correlation coefficients. Smaller the differences, slower movements within a crowd, larger the differences, large movements within a crowd. By evaluating these differences we can know the region where anomaly exists.

3.4 Decision Criteria

Differences in correlation coefficients values give an idea about the region where the event occurrences can happen. A threshold value can be selected for making a decision about an event, that the value of correlation coefficients of that frame from where differences of correlation coefficients start increasing. For a selected threshold value we can compare correlation coefficients of all images. If the value of the correlation of an image sequence is less than the selected threshold then a non-violent event can be detected. But if the value of correlation coefficients of a particular sequence is larger than a selected threshold then we can say that a violent event can be detected.

Figure 3: UMN dataset frames, row one is of normal event, row two is of Abnormal event frame.

$R \leq Threshold$; Violent event detected
$Otherwise$; Non-Violent event detected

4 RESULT ANALYSIS

The implementation has been applied on UMN dataset (Bansod and Nandedkar, 2020; Chen and Lai, 2019; Nayan et al., 2019). The simulation was done on a CPU with 32GB of RAM, a 3.40GHz Intel(R) Core(TM) processor.

4.1 Dataset

The dataset was generated by the University of Minnesota, especially for unusual activity detection for both indoors and outdoor activities. The video dataset consists single video combination of eleven videos of three different weather scenarios as perfect daylight, perfect illumination effect, and poor illumination effects. The resolution of video is 320 by 240 pixels, with frame rate of 30fps. Some sample frames are shown in the figure 3. For evaluation purposes, a single video gets separated into a total of 11 different videos of a 30fps frame rate. As shown in the table. 1 no of frames each video consists and frame number at which event gets detected are tabulated.

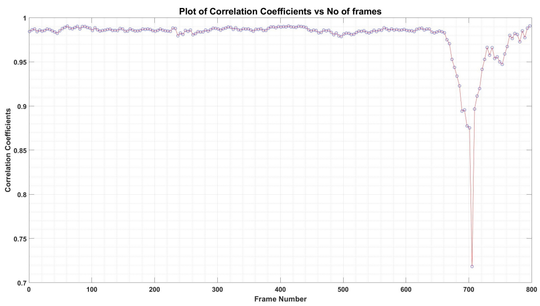

Figure 4: Plot of correlation coefficients of optical flow magnitude versus frame number

4.2 Analysis on Dataset

Test video no.-2: A very second video of the UMN dataset is of outdoor activity video with slightly low illumination. Video is having a total of 828 frames of resolution 320 by 240 pixels. The proposed methodology was implemented on the video to get results for violent detection. Simulation graphs are shown in figure.4 and 5 of correlation vs frame number and event detected graph along with threshold line respectively. Valley is the region where a sudden fall of correlation coefficients happened due to quick movements within a crowd. A threshold value can be selected adaptively by means of using a difference of correlation coefficients. Depending upon the selected threshold value event can be detected as a violent or non-violent event successfully in figure 6.

Figure 5: Threshold plot for an video with threshold selected at th=0.9480

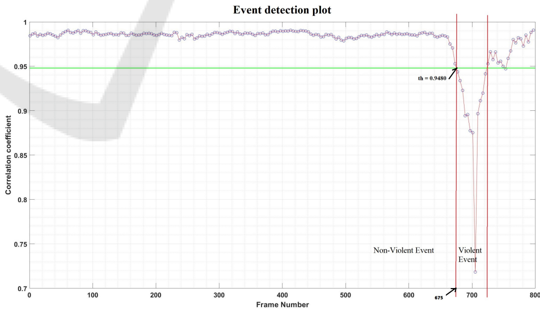

Figure 6: Event detection plot with violent and non-violent event regions

Test video no.-4: The fourth video of the UMN dataset of indoor activities with poor illumination of duration 22 sec, with a frame rate of 30fps, each frame is of the size of 320 x 240. The simulation plots for a video are shown in figure.7,8, and 9 are of correlation plot, threshold plot, and event detection plot respectively. Further, the evaluation of all videos of the UMN dataset has been tried over algorithm. The table 1 gives the statistics about number of frames, se-

Table 1: Evaluation table for UMN dataset

Video No.	No. of frames in Each Video	Simulation Time (sec)	Threshold Value (th)	Event Occurring at frame No.	Ground truth
Video 1	625	24.69	0.9082	495	485
Video 2	828	35.86	0.9480	675	679
Video 3	549	19.43	0.9948	323	320
Video 4	685	27.47	0.9966	573	674
Video 5	768	31.97	0.9940	502	496
Video 6	579	20.42	0.9976	453	460
Video 7	895	41.02	0.9948	743	741
Video 8	667	25.19	0.9968	463	469
Video 9	658	25.00	0.8369	569	552
Video 10	677	25.93	0.8090	599	578
Video 11	808	33.39	0.8954	739	725

lected threshold, and frame no at which event can be detected. It can be observed that event detected frame is very close to ground truth frame number.

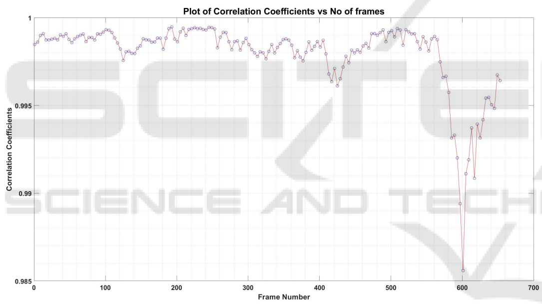

Figure 7: Plot of correlation coefficients of optical flow magnitude versus frame number

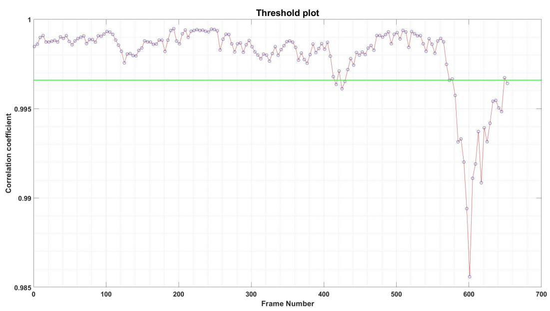

Figure 8: Threshold plot for an video with threshold selected at th=0.9966

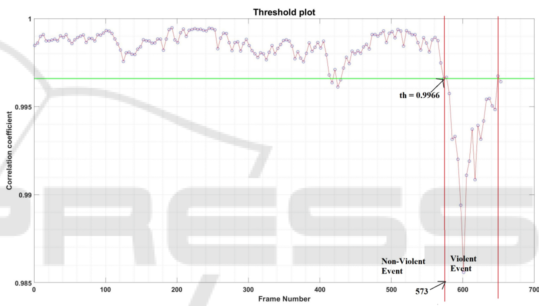

Figure 9: Event detection plot with violent and non-violent event regions

5 CONCLUSION

This paper is all about the optical flow method used to detect violent events within a crowded scene video dataset. Violent is referred to as an instant at which people start dispersing suddenly or quick dispersal of people within the crowd. Correlation coefficients are used to analyze the relationship between successive frames so that to know about the number of movements. If successive frames have high correlation means frames have low motion, whereas if correlation coefficients are low then motion within the frame is large. Here we need to target the frames where we get large motions as large motions pointed towards the quick and sudden action hence violent event.

REFERENCES

Abrishami Moghaddam, H. and Zare, A. (2019). Spatiotemporal wavelet correlogram for human action recognition. *International Journal of Multimedia Information Retrieval*, 8(3):167–180.

Afiq, A., Zakariya, M., Saad, M., Nurfarzana, A., Khir, M. H. M., Fadzil, A., Jale, A., Gunawan, W., Izuddin, Z., and Faizari, M. (2019). A review on classifying abnormal behavior in crowd scene. *Journal of Visual Communication and Image Representation*, 58:285–303.

Bansod, S. D. and Nandedkar, A. V. (2020). Crowd anomaly detection and localization using histogram of magnitude and momentum. *The Visual Computer*, 36(3):609–620.

Bour, P., Cribelier, E., and Argyriou, V. (2019). Crowd behavior analysis from fixed and moving cameras. In *Multimodal Behavior Analysis in the Wild*, pages 289–322. Elsevier.

Chavan, R., Gengaje, S., and Gaikwad, S. (2018). Multi-object detection using modified gmm-based background subtraction technique. In *International Conference on ISMAC in Computational Vision and Bio-Engineering*, pages 945–954. Springer.

Cheggoju, N., Nawandar, N. K., and Satpute, V. R. (2021). Entropy: A new parameter for image deciphering. In *Proceedings of International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications*, pages 681–688. Springer.

Chen, X.-H. and Lai, J.-H. (2019). Detecting abnormal crowd behaviors based on the div-curl characteristics of flow fields. *Pattern Recognition*, 88:342–355.

Franco, A., Magnani, A., and Maio, D. (2020). A multimodal approach for human activity recognition based on skeleton and rgb data. *Pattern Recognition Letters*, 131:293–299.

Gajbhiye, P., Naveen, C., and Satpute, V. R. (2017). Virtue: Video surveillance for rail-road traffic safety at unmanned level crossings;(incorporating indian scenario). In *2017 IEEE Region 10 Symposium (TENSYMP)*, pages 1–4. IEEE.

Gangal, P., Satpute, V., Kulat, K., and Keskar, A. (2014a). A novel approach based on 2d-dwt and variance method for human detection and tracking in video surveillance applications. In *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, pages 463–468. IEEE.

Gangal, P., Satpute, V., Kulat, K., and Keskar, A. (2014b). Pnew approch for object detection and tracking using 2d-dwt. In *2014 International Conference on Electronics and Communication Systems (ICECS)*, pages 1–5. IEEE.

Ghutke, R. C., Naveen, C., and Satpute, V. R. (2016). A novel approach for video frame interpolation using cubic motion compensation technique. *International Journal of Applied Engineering Research*, 11(10):7139–7146.

Gupta, A., Stapute, V., Kulat, K., and Bokde, N. (2016). Real-time abandoned object detection using video surveillance. In *Proceedings of the International Conference on Recent Cognizance in Wireless Communication & Image Processing*, pages 837–843. Springer.

Hassner, T., Itcher, Y., and Kliper-Gross, O. (2012). Violent flows: Real-time detection of violent crowd behavior. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–6. IEEE.

Horn, B. K. and Schunck, B. G. (1981). Determining optical flow. *Artificial intelligence*, 17(1-3):185–203.

Jirafe, A., Jibhe, M., and Satpute, V. (2021). Camera handoff for multi-camera surveillance. In *Applications of Advanced Computing in Systems*, pages 267–274. Springer.

Lucas, B. D., Kanade, T., et al. (1981). An iterative image registration technique with an application to stereo vision. Vancouver.

Mehran, R., Oyama, A., and Shah, M. (2009). Abnormal crowd behavior detection using social force model. In *2009 IEEE conference on computer vision and pattern recognition*, pages 935–942. IEEE.

Meinhardt-Llopis, E. and Sánchez, J. (2012). Horn-schunck optical flow with a multi-scale strategy. *Image Processing on line*.

Mohammadi, S., Kiani, H., Perina, A., and Murino, V. (2015). Violence detection in crowded scenes using substantial derivative. In *2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6. IEEE.

Mohammadi, S., Perina, A., Kiani, H., and Murino, V. (2016a). Angry crowds: Detecting violent events in videos. In *European Conference on Computer Vision*, pages 3–18. Springer.

Mohammadi, S., Setti, F., Perina, A., Cristani, M., and Murino, V. (2016b). Groups and crowds: Behaviour analysis of people aggregations. In *International Joint Conference on Computer Vision, Imaging and Computer Graphics*, pages 3–32. Springer.

Murino, V., Cristani, M., Shah, S., and Savarese, S. (2017). *Group and crowd behavior for computer vision*. Academic Press.

Nale, R., Sawarbandhe, M., Chegogoju, N., and Satpute, V. (2021). Suspicious human activity detection using pose estimation and lstm. In *2021 International Symposium of Asian Control Association on Intelligent Robotics and Industrial Automation (IRIA)*, pages 197–202. IEEE.

Naveen, C., Satpute, V. R., Kulat, K. D., and Keskar, A. G. (2014). Video encoding techniques based on 3d-dwt. In *2014 IEEE Students' Conference on Electrical, Electronics and Computer Science*, pages 1–6. IEEE.

Nayan, N., Sahu, S. S., and Kumar, S. (2019). Detecting anomalous crowd behavior using correlation analysis of optical flow. *Signal, Image and Video Processing*, 13(6):1233–1241.

Parate, M. R., Satpute, V. R., and Bhurchandi, K. M. (2018). Global-patch-hybrid template-based arbitrary object tracking with integral channel features. *Applied Intelligence*, 48(2):300–314.

Patil, N. and Biswas, P. K. (2017). Detection of global abnormal events in crowded scenes. In *2017 Twenty-third National Conference on Communications (NCC)*, pages 1–6. IEEE.

Pawade, A., Anjaria, R., and Satpute, V. (2021). Suspicious activity detection for security cameras. In *Applications of Advanced Computing in Systems*, pages 211–217. Springer.

Satpute, V., Kulat, K., and Keskar, A. (2014). A novel approach based on 2d—dwt and variance method for human detection and tracking in video surveillance applications (an alternative approach for object detection). In *2014 9th International Conference on Industrial and Information Systems (ICIIS)*, pages 1–6. IEEE.

Sawarbandhe, M. D., Satpute, V. R., Cheggoju, N., and Sinha, S. (2019). Combining learning and handcrafted based features for face recognition in still images. In *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, pages 2502–2506. IEEE.

Sharma, S. and Dhama, V. (2020). Abnormal human behavior detection in video using suspicious object detection. In *ICDSMLA 2019*, pages 379–388. Springer.

Sreenu, G. and Durai, M. S. (2019). Intelligent video surveillance: a review through deep learning techniques for crowd analysis. *Journal of Big Data*, 6(1):1–27.

Thombare, P., Gond, V., and Satpute, V. (2021). Artificial intelligence for low level suspicious activity detection. In *Applications of Advanced Computing in Systems*, pages 219–226. Springer.

Vrigkas, M., Nikou, C., and Kakadiaris, I. A. (2015). A review of human activity recognition methods. *Frontiers in Robotics and AI*, 2:28.

Zhang, S., Wei, Z., Nie, J., Huang, L., Wang, S., and Li, Z. (2017). A review on human activity recognition using vision-based method. *Journal of healthcare engineering*, 2017.