Metadata of the chapter that will be visualized in SpringerLink

Book Title	Proceedings of the 2nd International Conference on Data Engineering and Communication Technology
Series Title	
Chapter Title	An Effective Video Surveillance Framework for Ragging/Violence Recognition
Copyright Year	2019
Copyright HolderName	Springer Nature Singapore Pte Ltd.

Corresponding Author	Family Name	**Sharma**
	Particle	
	Given Name	**Abhishek**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	Graphic Era University
	Address	Dehradun, Uttarakhand, India
	Email	abhisheksharma13598@gmail.com
Author	Family Name	**Tripathi**
	Particle	
	Given Name	**Vikas**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	Graphic Era University
	Address	Dehradun, Uttarakhand, India
	Email	
Author	Family Name	**Gangodkar**
	Particle	
	Given Name	**Durgaprasad**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	Graphic Era University
	Address	Dehradun, Uttarakhand, India
	Email	

Abstract

Ragging is one among the biggest problems in the education system in the current time. This paper proposes an efficient computer vision-based framework to detect ragging/violence during the actual time of occurrence of the event. The framework of the system is based on motion analysis. To determine the motion in the video, the use of optical flow has been made. The code has been trained and tested by a collection of several videos. Three different reference frame delays are considered, and the result analytics are reviewed for effective reference frame selection. Passage of the different frames of the video through the descriptor takes place in order to wrench out feature from the frames. The use of two descriptors HOG and LBP has been made individually with optical flow to generate dataset. Further analysis has been done through Random Forest to validate our results. Our proposed framework is able for effective detection of ragging and normal activities with an accuracy of 88.48%.

Keywords (separated by '-')	Computer vision - Optical flow - HOG - LBP - Ragging

Author Proof

An Effective Video Surveillance Framework for Ragging/Violence Recognition

Abhishek Sharma, Vikas Tripathi and Durgaprasad Gangodkar

Abstract Ragging is one among the biggest problems in the education system in the current time. This paper proposes an efficient computer vision-based framework to detect ragging/violence during the actual time of occurrence of the event. The framework of the system is based on motion analysis. To determine the motion in the video, the use of optical flow has been made. The code has been trained and tested by a collection of several videos. Three different reference frame delays are considered, and the result analytics are reviewed for effective reference frame selection. Passage of the different frames of the video through the descriptor takes place in order to wrench out feature from the frames. The use of two descriptors HOG and LBP has been made individually with optical flow to generate dataset. Further analysis has been done through Random Forest to validate our results. Our proposed framework is able for effective detection of ragging and normal activities with an accuracy of 88.48%.

Keywords Computer vision · Optical flow · HOG · LBP · Ragging

1 Introduction

Ragging is one of the most common problems that has a deep root in the education society. Since many years, several students have been victim to this malpractice. Although there are many laws against ragging in India, still this practice prevails in most of the institutions of higher studies. The problem is more severe in medical and engineering colleges. Physical violence, mental torture, and rape are some of the darkest sides of ragging. Forcing the juniors to indulge in unhealthy activities or using them to get through their work is another widespread problem in ragging. Ragging can cost life of a student. According to the RTI reply, UGC received 640

A. Sharma (✉) · V. Tripathi · D. Gangodkar
Graphic Era University, Dehradun, Uttarakhand, India
e-mail: abhisheksharma13598@gmail.com

© Springer Nature Singapore Pte Ltd. 2019

A. J. Kulkarni et al. (eds.), *Proceedings of the 2nd International Conference on Data Engineering and Communication Technology*, Advances in Intelligent Systems and Computing 828, https://doi.org/10.1007/978-981-13-1610-4_25

UNCORRECTED PROOF
455037_1_En_25_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:14/7/2018 Pages: 11 Layout: T1-Standard

Author Proof

and 543 complaints in year 2013 and 2014, respectively. Even an FIR was filed in 9 and 57 cases, respectively, in year 2013 and 2014 [1].

Almost all institutes are well equipped with the CCTV cameras, and almost every part of the campus is covered with these cameras. Most of the time these cameras are used for recording, but they are used after some accident has occurred. So, we came up with the thought, '*Why not use CCTV cameras to their full potential so that they can do something more than just recording?*' If the cameras can detect ragging, then there will be no need for the victim to register complaint and the culprits would be caught and the case will be handled accordingly. In this paper, we proposed a framework that will be self-sufficient to analyze the video and capture ragging. We have used optical flow, HOG, LBP, and other techniques to develop the desired system. The paper is organized as follows: In Sect. 2, we provide the previous work performed on the violence detection in videos. Methodology of the proposed work has been explained in Sect. 3. Section 4 contains the analysis of the result and the last section concludes the paper.

2 Literature Review

The first among those with the ideas to detect violence in video was Nam et al. [2], it gives the idea of recognizing unusual activities in videos using blood detection and fire in the video and analyzing the movement, also by analyzing the sounds in the case of abnormal activity. Cheng et al. [3] took into consideration the gunshots, car-breaking, and explosions in audio. Giannakopoulos et al. [4] also put forward the idea of detecting violent activities using sound. Clarinet al. [5] talks about the framework that uses a Kohonen framework to recognize blood pixels and human skin in the video and detecting motion involved in violent activity with blood in the action involved. Zajdel et al. [6] proposed a framework, which is based on feature extraction from the motion in video and distress audio to detect unusual activity like violence in the videos.

Another work involves that of Gong et al. [7]. He put forward the idea of violence detection using video, sound, and sound effects to detect abusive and unusual activities in the movies. Chen et al. [8] used the feature extraction from motion and collection of dataset in order to detect aggressive behaviors. Lin and Wang [9] used motion in the movie along with the sound training followed by the sound of exploding and blood detection classifier. Giannakopoulos et al. [10] used framework for detecting violent scene in movies on the basis of audio statistics, combined with the statistic of visuals and motion statistics to determine whether a video contains violent scene or not. Some other recent work involves the papers breaking down violence scene by Acar et al. [11] in the year 2015 and violence detection using oriented violent flow by Gao et al. [12] in the year 2016. Finally, to conclude the basic theme of all the papers mentioned, we can say that most ideas are based on the analysis of videos along with the audio. But these systems may fail in the case where there are no audios. Not all CCTV cameras are accompanied by voice recording system.

UNCORRECTED PROOF

455037_1_En_25_Chapter ✓ TYPESET DISK LE ✓ CP Disp.:14/7/2018 Pages: 11 Layout: T1-Standard

Author Proof

Generally, the CCTV cameras are used only to record the video. Another important criteria used in many papers is the recognition of blood to detect violence. Since the quantity of blood may not be significant to detect for a camera like that of a CCTV which has low resolution. AQ3

3 Methodology

In this section, we have given a detailed explanation of the steps that were followed in getting an efficient result. Firstly, we will discuss the datasets that are used in this experiment, and later, the procedure of the experiment is laid out.

3.1 Datasets

The video dataset is divided into two parts; one is the ragging, and the other is normal activity. The main source of the video is YouTube, and few scenes are from movie 'Hostel of 2011 by Manish Gupta [13].' Although there are a few videos that we shot on our own. AQ4

The ragging dataset contains fight scene (pushing, kicking, punching, bullying activities, etc.). The length of each video varies from 30 s to a minute long. In the test set, the total video length is around 7–8 min. Figure 1 shows some scenes of the videos from the ragging activity dataset.

Fig. 1 Images from ragging activities dataset

UNCORRECTED PROOF

455037_1_En_25_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:14/7/2018 Pages: 11 Layout: T1-Standard

Author Proof

Fig. 2 Images from normal activities dataset

The normal dataset contains the usual activities in the day-to-day life on the road, institutes, and other public places. The activities involved are walking, talking, sitting, etc. We tried to cover all the normal activities that are observable in the public place. In the test set, the total video length is around 7–8 min. Scenes of some normal activities from normal activity dataset are shown in Fig. 2.

3.2 Proposed Frame Work

The main ideology behind the framework is combining the optical flow with the descriptors such as HOG [14] and LBP. Algorithm 1 shows the steps involved in the implementation of framework where window size refers to reference frame delay of 5, 10, and 15 frames while calculating the optical flow. These steps involve acquiring video, calculating the optical flow, passing the images to a descriptor, classifying the image using random forest, etc.

UNCORRECTED PROOF

Author Proof

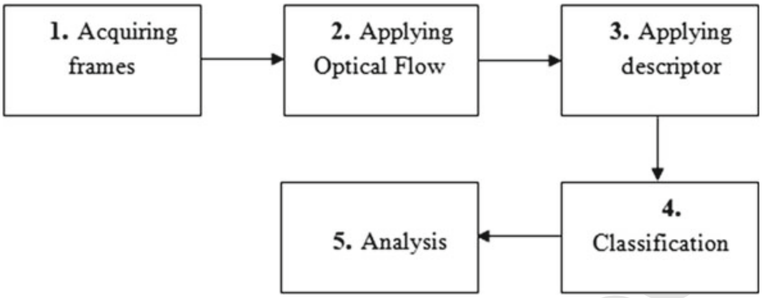

Fig. 3 Flow diagram of proposed framework

Algorithm 1 for Generation of Descriptor

Input : Video.
Output : Compute descriptor.

1. Compute frames.
2. Initialize p with 0.
3. Initialize n with window size
4. While p < frames do
   - I. If (p % n = 0)
     - a. Calculate optical flow.
     - b. Extract features.
     - c. Store features.
     - d. Increment p by 1.
5. Dataset generated.

Figure 3 shows the flow diagram of the implementation of the frame work. There are mainly five blocks to implement this framework. Our first step was to read the video files and extract the frames from the video. The video frames are acquired, and the color images are converted into gray scale. The videos are '.mp4' format.

The next step involves applying optical flow on the frames. In order to achieve more accuracy, we have taken variations in applying optical flow. We have taken reference frame delay of 5, 10, and 15. This means that for a frame delay of 5, the current frame will be evaluated with the fifth frame in the video in order to calculate optical flow. Optical flow is nothing but the movement of the pixel in 'x' and 'y' direction. To calculate this 'x' and 'y,' there are two popular methods; one is Horn and Schunck [15] and Lucas–Kanade method. Image in Fig. 4 depicts optical flow in the image. These images were in RGB and were converted into gray scale. The lines in the image depict optical flow.

We have used two well-known descriptors HOG and LBP. Then, an analysis is made among these descriptors. We obtain 1×288 arrays in the case of HOG and 1×348 in the case of LBP. There are two classes; one is named ragging, and the other is normal. So for all the frames, we store values in a file.

Next step involves classifying data. There are almost 75,000 instances in training set with almost equal number of instances for ragging and normal. Similarly, with

UNCORRECTED PROOF

Author Proof

Fig. 4 Images with optical flow (lines represent optical flow)

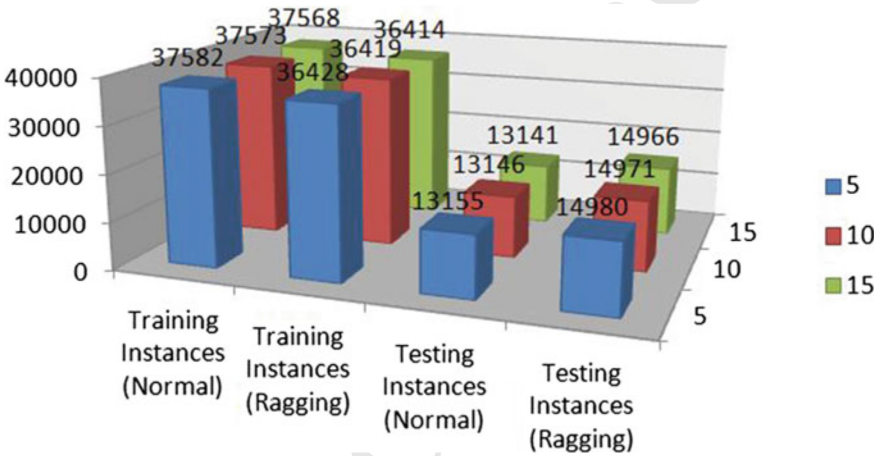

Fig. 5 Graph representing instances of test and train in terms of frames

almost equal number of instances for ragging and normal dataset, there are about 28,000 total instances in testing dataset. The videos are in the format 480 × 854. We have used random forest for classification purpose. Figure 5 depicts the information regarding instances of both train and test for both ragging activity and normal activity classes.

4 Result and Analysis

The framework is coded in MATLAB; training and testing are conducted in weka. The system on which this framework was developed has 4 GB RAM, Intel core i3, 2.13 GHz processor. The frame resolution chosen is 480 × 854. The videos are in '.mp4' format, and there length varies from 10 s to 2 min. The frame rate is 29 frames/s. Table 1 gives the accuracy in term of percentage.

UNCORRECTED PROOF

Author Proof

Table 1 Accuracy with different method

Method	5	10	15
LBP	87.3645	86.7816	88.4805
HOG	75.4256	76.9966	77.3983

Table 2 Detailed result statistics for LBP with 15 frame delay optical flow

Class	True-positive ratio	False-positive ratio	Precision value	Recall value	F-measure	ROC
Normal	0.873	0.104	0.88	0.873	0.876	0.956
Ragging	0.896	0.127	0.889	0.896	0.892	0.956
W. Average	0.885	0.117	0.885	0.885	0.885	0.956

Table 3 Confusion matrix for LBP with 15 frame optical flow

	Ragging	Normal	Total
Ragging	13,416	1564	14,980
Normal	1677	11,478	13,155
Total	15,093	13,042	28,135

In Table 1, values 5, 10, and 15 in first row are referring to the reference frame delay of frames while calculating optical flow. For example, 5 from first row and method LBP in above table imply that the descriptor is LBP with optical flow having reference frame delay of five frames. From table above, we can observe that the best accuracy is obtained by LBP with reference frame delay of 15 frames. AQ6

From the Table 2, the value of true-positive ratio, false-positive ratio, and precision gives very promising results. True-positive ratio is the proportion of correctly identified instances. Similarly, we have false-positive ratio that generates a false alarm. Precision value is the refinement in the calculations, recall means the fraction of the relevant instances that are retrieved, and F-measure takes the combination of precision and recall to provide a single value of measurement.

Table 3 represents the confusion matrix. Here, we can see that about 11,478 instances of normal and 13,416 instances of ragging are correctly classified. These results are quite promising as our framework identifies 24,894 instances correctly out of 28,135 instances which is 88.48%.

Figures 6 and 7 are showing threshold model for ragging activities as well as threshold model for normal activities, respectively. The ROC curve is a plot which is plotted in between false-positive rate and true-positive rate for different cutoff points. The points on the ROC curve represent the ratio of true-positive rate and false-positive rate with respect to a particular decision threshold. The ROC curve passes to the upper left corner shows the high sensitivity and specificity which means area under the ROC curve close to 1 represents the high accuracy. The area under both of our model is 0.9958 which shows an excellent test.

UNCORRECTED PROOF

Author Proof

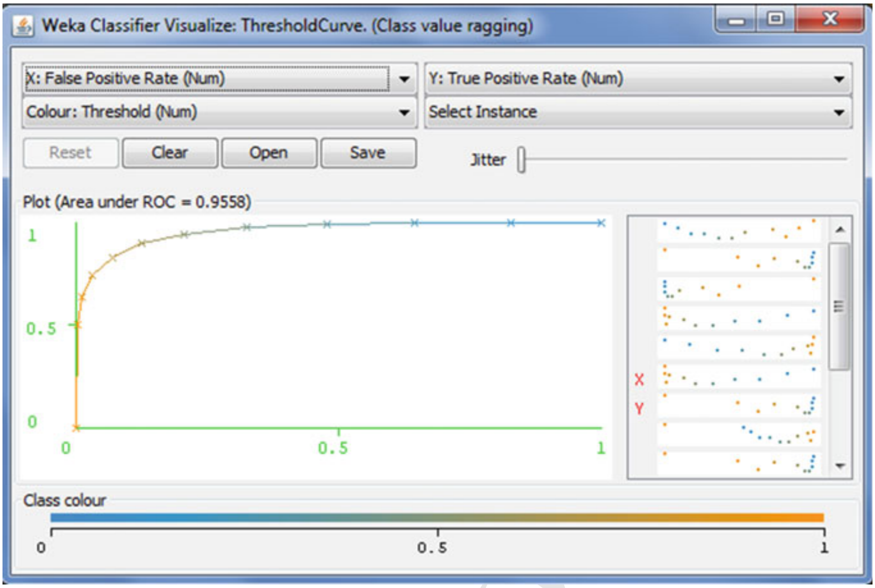

Fig. 6 Threshold model (ragging activity)

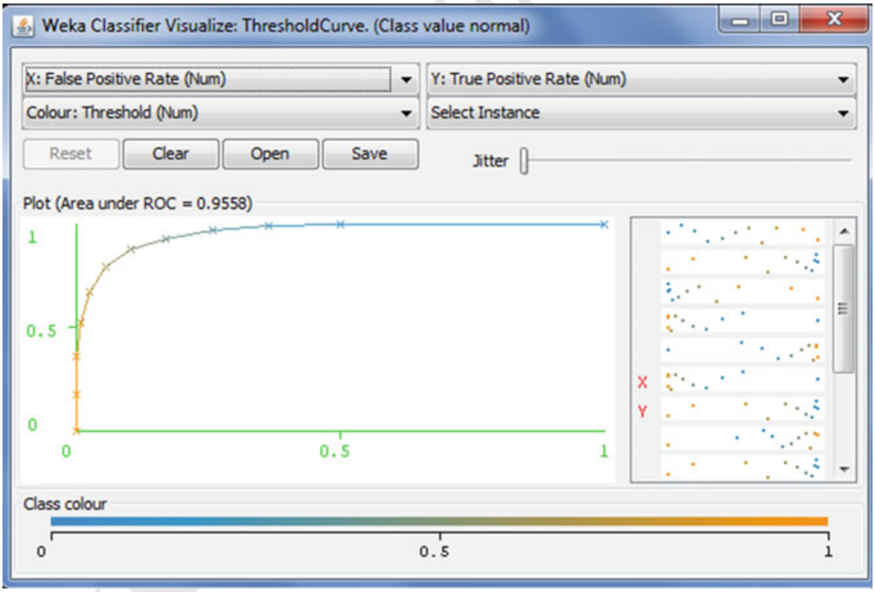

Fig. 7 Threshold model (normal activity)

Author Proof

5 Conclusion

In this paper, a framework for better prevention of ragging using video surveillance in college campus or similar premises has been presented. In particular, this paper discusses the recognition of ragging and normal activity in the concerned area. We have taken into consideration three different cases for optical flow (reference frame delay) and two different types of descriptor (HOG and LBP). Thus, a total of six different cases to observe the accuracy are presented. From the different cases, maximum accuracy in the case of HOG is 77.39% for reference frame delay of 15 frames and in the case for LBP it is 88.48% for reference frame delay of 15 frames. However, this framework is open to improvizations and advancements such as fusion of framework with audio analysis, application of new descriptors or addition of various combinations of optical flow and descriptors.

References

1. Times of India Homepage, http://timesofindia.indiatimes.com/home/education/news/1183-ragging-cases-66-FIRs-in-two-years/articleshow/47477316.cms. Last accessed 12 Aug 2017
2. Nam J, Alghoniemy M, Tewfik AH (1998) Audio-visual content-based violent scene characterization. In: Proceedings of international conference on image processing, pp 353–357
3. Cheng WH, Chu WT, Wu JL (2003) Semantic context detection based on hierarchical audio models. In: Proceedings of the ACM SIGMM workshop on multimedia information retrieval, pp 109–115
4. Giannakopoulos T, Makris A, Kosmopoulos D, Perantonis S, Theodoridis S (2006) Violence content classification using audio features. Advances in Artificial Intelligence, Lecture Notes in Computer Science, vol 3955, pp 502–507
5. Clarin CT, Ann J, Dionisio Michael M, Echavez T, Naval Jr PC (2005) Detection of movie violence using motion intensity analysis on skin and blood. In: University of the Philippines, pp 150–156
6. Zajdel W, Andringa T, Gavrila DM, Krijnders JD (2007) Audio-video sensor fusion for aggression detection. In: IEEE conference on advanced video and signal based surveillance, pp 200–205
7. Gong Y, Wang W, Jiang S, Huang Q, Gao W (2008) Detecting violent scenes in movies by auditory and visual cues. In: Proceedings of the 9th Pacific rim conference on multimedia, pp 317–326
8. Chen D, Wactlar HD, Chen MY, Hauptmann A (2008) Recognition of aggressive human behaviour using binary local motion descriptors. In: Engineering in medicine and biology society, pp 5238–5241
9. Lin J, Wang W (2009) Weakly-supervised violence detection in movies with audio and video based co-training. In: Proceedings of the 10th pacific rim conference on multimedia, pp 930–935
10. Giannakopoulos T, Makris A, Kosmopoulos D, Perantonis S, Theodoridis S (2010) Audio-visual fusion for detecting violent scenes in videos. In: 6th Hellenic conference on AI SETN Athens Greece, pp 91–100
11. Acar E, Hopfgartner F, Albayrak S (2016) Breaking down violence scene. Neurocomputing 208:225–237
12. Gao Y, Liu H, Sun X, Wang C, Liu Y (2016) Violence detection using oriented violent flow Image and Vision Computing. Image Vis Comput 48:37–41

UNCORRECTED PROOF

455037_1_En_25_Chapter ✓ TYPESET DISK LE ✓ CP Disp.:14/7/2018 Pages: 11 Layout: T1-Standard

Author Proof

13. Youtube Homepage. https://www.youtube.com/watch?v=1ktaw31FyiY. Last accessed 21 June 2017
14. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: International conference on computer vision & pattern recognition (CVPR). IEEE Computer Society, vol 1. San Diego, USA, pp 886–893
15. Horn BKP, Schunck BG (1981) Determining optical flow. In: Artificial intelligence, pp 81–87

UNCORRECTED PROOF

Author Proof

Author Queries

Chapter 25

Query Refs.	Details Required	Author's response
AQ1	Please check and confirm if the author names and initials are correct.	Everything is correct
AQ2	Please check the clarity of the phrase 'is able for effective detection' in the sentence 'Our proposed framework is able…'.	Fine. Nothing to change
AQ3	Please check the clarity of the sentence 'Since the quantity of blood…'.	Fine. Nothing to change
AQ4	Please check the clarity of the sentence 'Although there are a few videos that…'.	Fine. Nothing to change
AQ5	Please check the term 'weka' with regard to capitalization in the sentence 'The framework is coded in…'.	Initial letter of the word in 'weka' should be in caps
AQ6	Please check whether the edit made in the sentence 'For example, 5…' conveys the intended meaning.	No change is required

UNCORRECTED PROOF

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	*Textual mark*	*Marginal mark*
Leave unchanged	• • • under matter to remain	
Insert in text the matter indicated in the margin		New matter followed by or
Delete	/ through single character, rule or underline or through all characters to be deleted	or
Substitute character or substitute part of one or more word(s)	/ through letter or through characters	new character / or new characters /
Change to italics	under matter to be changed	
Change to capitals	under matter to be changed	
Change to small capitals	under matter to be changed	
Change to bold type	under matter to be changed	
Change to bold italic	under matter to be changed	
Change to lower case	Encircle matter to be changed	
Change italic to upright type	(As above)	
Change bold to non-bold type	(As above)	
Insert 'superior' character	/ through character or where required	or under character e.g. or
Insert 'inferior' character	(As above)	over character e.g.
Insert full stop	(As above)	⊙
Insert comma	(As above)	,
Insert single quotation marks	(As above)	or and/or or
Insert double quotation marks	(As above)	or and/or or
Insert hyphen	(As above)	
Start new paragraph		
No new paragraph		
Transpose		
Close up	linking characters	
Insert or substitute space between characters or words	/ through character or where required	
Reduce space between characters or words	between characters or words affected	