A deep Spatio-temporal network for vision-based sexual harassment detection

Md Shamimul Islam
Dept. of CSE
Manarat International University
Dhaka, Bangladesh
shamimulislam@manarat.ac.bd

Md Mahedi Hasan
IICT
BUET
Dhaka, Bangladesh
mahedi0803@gmail.com

Sohaib Abdullah
Dept. of CSE
Manarat International University
Dhaka, Bangladesh
sohaib@manarat.ac.bd

Jalal Uddin Md Akbar
Dept. of CSE
International Islamic University Chittagong
Chittagong, Bangladesh
jalaluddinmdakbar00@gmail.com

N H M Arafat
Dept. of Computer Science and Technology
Henan Polytechnic University
454003, Jiaozuo, Henan, P.R. China
arafat.nhm@gmail.com

Saydul Akbar Murad
Faculty of Computing
Universiti Malaysia Pahang
26600, Pahang, Malaysia
saydulakbarmurad@gmail.com

2021 Emerging Technology in Computing, Communication and Electronics (ETCCE) | 978-1-6654-8364-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/ETCCE54784.2021.9689891

***Abstract*—Smart surveillance systems can play a significant role in detecting sexual harassment in real-time for law enforcement which can reduce the sexual harassment activities. Real-time detecting of sexual harassment from video is a complex computer vision because of various factors such as clothing or carrying variation, illumination variation, partial occlusion, low resolution, view angle variation etc. Due to the advancement of convolutional neural networks (CNNs) and Long Short-Term Memory (LSTM), human action recognition tasks have achieved great success in recent years. But sexual harassment detection is addressed due to presences of large-scale harassment dataset. In this work, to address this problem, we build a video dataset of sexual harassment, namely Sexual harassment video (SHV) dataset which consists of harassment and non-harassment videos collected from YouTube. Besides, we build a CNN-LSTM network to detect the sexual harassment in which CNN and RNN are employed for extracting spatial features and temporal features, respectively. State-of-the-art pretrained models are also employed as a spatial feature extractor with an LSTM and three dense layer to classify harassment activities. Moreover, to find the robustness of our proposed model, we have conducted several experiments with our proposed method on two other benchmark datasets, such as Hockey Fight dataset and Movie Violence dataset and achieved state-of-the-art accuracy.**

***Index Terms*—sexual harassment, surveillance systems, deep learning**

I. Introduction

Sexual harassment, one of the most unwanted and intolerable criminal activities in the world, which according to the UN, is unsolicited sexual advances, asking for sexual favors, and other physical or verbal activities which are sexual in nature [1]. As it is accomplished against one's will, it results in hurting the victim's dignity. In a variety of circumstances, it may occur in different places such as in workplaces, in the military, in academic institutes, during transportation [2], in public places, etc. The harasser can be a direct supervisor, an indirect supervisor, a coworker, an instructor, a peer, or a colleague and may have any gender and any type of connection with the victim. It can happen in physical or verbal form and both online and offline. In this paper, our focus is on the onsite physical form of harassment as it is one of the most serious offenses and if not protected in due time may lead to greater crimes like rape.

Nearly one in five women will experience sexual harassment in any given year, according to the result of a survey named "National Intimate Partner and Sexual Violence Survey" [3]. Victims of sexual harassment can experience severe psychological consequences, including anxiety, headaches, depression, sleep disorders, losing or gaining weight, nausea, reduced self-esteem, and sexual dysfunction. Using advanced intelligent surveillance-based systems, we can reduce this unwanted event. Although sexual harassment cannot be completely prevented, the sexual harassment detection system can reduce the frequency, if it can accurately recognize a harassment incident and generate an alert.

In this work, we have explored several deep learning algorithms to classify harassment in videos and experimented with those on the Sexual Harassment Video (SHV) dataset which has been collected from YouTube. The dataset contains harassment videos that have been collected from different dramas, harassment training videos, harassment consciousness videos, etc. We have applied CNN-LSTM in our collected dataset where CNN is employed as a spatial feature extractor and LSTM is used as a temporal feature extractor. We have evaluated our proposed model extensively in two of the most complicated and benchmark datasets, the Hockey fight dataset [6], and Movie Violence Dataset [6]. The proposed method achieved state-of-the-art accuracy in these datasets. The following are the paper's major contributions:

- We have developed a novel Sexual Harassment Video (SHV) dataset which contains 300 sexual harassment videos belonging to two classes: harassment and non-harassment situations. We have collected events that occurred in different locations in such a manner that it can represent both in-house and open site events.

978-1-6654-8364-3/21/31.002021*IEEE*