INSTITUTE FOR HOMELAND SECURITY

Sam Houston State University

# REMOTE DETECTION OF ANGER:

## CCTV AND LARGE-SCALE SENSOR NETWORKS


**Institute for Homeland Security**

**Sam Houston State University**

Zeno Franco, PhD

# Remote Detection of Anger - CCTV and Large-Scale Sensor Networks

Author: Zeno Franco, PhD
For: Sam Houston State University

## Executive Summary

In light of recent and historical active shooter and other violent episodes in places of work, strategies that can rapidly identify high risk incidents using remote sensing systems are of increasing importance. The use of CCTV and other massive sensor networks to detect physical gestures that are associated with anger, frustration, and other high valence emotions. Strategies involving a SmartCities framework that optimize early detection of violence through CCTVs and related technologies as a distributed sensor network has implications for multiple sectors including healthcare, education, transportation, energy, chemical, government, and entertainment. Smaller private businesses and rural municipalities may also benefit from such approaches as these systems may ultimately reduce the number of personnel dedicated to monitoring. Compared to other emotions, fairly little is known about anger. Similarly with sensor-based activity detection considerable gaps remain in understanding how to detect the prodromal phase of anger in video feeds. Currently, large amounts of gestural information is available from existing CCTV and related systems that goes unanalyzed.

## Background

Over the past few decades, closed circuit television (CCTV) and related technologies became increasingly popular in a wide range of passive monitoring applications. This has only increased with the advent of internet-based video security systems in recent years and been further complicated by live social media video feeds. However, overwhelming amount of video data, estimates that more than 99% of surveillance video is unviewed [1, 2]. Security professionals are unable to view, analyze or respond to video, which is often providing information with multiple views and numerous locations.

Automated analysis of CCTV, live social media, and other video streams is recognized as one of the only viable avenues for leveraging information about unfolding high-risk events, however a number limitations persist. Several high-level strategies for video based behavioral risk analysis exist including detection abnormal individual behaviors, abnormal small group behaviors (e.g. dyads or triads), and monitoring crowd level dynamics [3, 4]. These approaches broadly fit into a SmartCities framework [5, 6], but also have important implications for the private sector [7, 8].

Increasing concerns around active shooter incidents or other internal or external threats to employee safety suggest that optimizing the use of passive video security systems into early violence detection systems, which can trigger alerts and generate rapid crisis response are important next steps [9]. Ultimately, well-designed automated systems that use strategies garnered from machine learning can interrogate massive amounts of video data without human intervention. When an event that has a high probability of turning violent is detected, an alert to a human monitor for verification can be generated. This sociotechnical systems approach, with human cross-check, is important in order to minimize issues around false positives [10]. Further, because the human monitor would likely be receiving only complex visual information that requires attention and thought, the role of video monitoring becomes an engaging rather than a boring tasks [11, 12]. The human monitor on determining a real threat exists can request security or law enforcement intervention with specific location and situational dynamics information. However, several issues remain and substantive improvements in each of these areas are needed to ensure automated CCTV based crisis detection systems are sufficiently accurate.

# Problems

A number of research problems exist in this field. First among them is the lack of long, untrimmed video with realistic fights occurring. Most prior research focused on snippets of video, where prodromal events are removed, and where the fight sequence information is already pre-identified [1, 13] . Most available data sets are based on general activity that does not include fights, involve snippets extracted from Hollywood movies, or are created though faked violent actions [1]. One study collected 1,000 CCTV videos containing real world fights obtained from YouTube. Footage includes actions like punching, kicking, pushing, wrestling, with two or more people. Notably, videos that were not CCTV footage were dropped. This approach suggests that paths toward creating high quality marked data for machine learning is possible, but special attention should be paid to ensuring that the time before the altercations is also included so that all precursor events or weak signals indicating that a fight is about to occur can also be studied.

Another issue is that frame level data label is time-consuming, which limits the number of large-scale violent video data sets [13]. Hand-crafted features to extract audio information [13]– look at low level features that may not be robust. Some specific problems emerge with the variability in available videos for analysis, including lighting differences, camera vantage points, image quality, stability of camera, and color vs. black and white images [14]. Further, many initial studies examining video data for fights or expressions of anger focus only on visually available information, but not audio. Some authors have emphasized that early indicators of aggression may be best detected through vocalization, thus underscoring the importance and complexity of approaches that involve large numbers of distinct sensor types. With this, issues around multi-channel data fusion and pre-processing prior to machine learning emerges.

Notably, the considerations around detection of overt anger and fighting behaviors in CCTV feeds may serve as a distraction from the more subtle issue, which include the identification and detection of the prodromal signs that rage is about to occur, which often do not produce signal in the typical sense. Pauses in action or pauses in verbal exchange, stand-offs where two or more people are arguing but have not yet engaged in an altercation, and other areas where lack of overt aggression are key to understanding anger must also be factored in. Early detection of the potential for violence, rather than simply detecting the occurrence of violence, should remain the ultimate goal [1].

## Current State of the Art

Some of the more technical considerations around current state of the art in video-based activity detection systems include findings that motion information (optical flows) improves performance over RGB-only models [1]. A number of machine learning approaches have been examined, including weakly supervised violence detection [13], scale-invariant feature transformation, spatial-temporal interest point (STIP), histogram of orientated gradient (HOG), histograms of oriented optical flow (HOF), motion intensity, deep convolutional neural networks (CNN), multi Stream CNN [14], as well as 2D and 3D skeletal representations for action detection [14]. Some algorithms for violent scene detection in video exist include MediaEval and Fighting detector [13]. Notably, surveillance cameras have more stable backgrounds than many of the dataset exemplars, making the detection problem easier to resolved. Existing labeled violent video data sets include: HockeyFight, Peliculas, Technicolor, Violent Flows, UCF-Crimes, Surveillance Camera Fight Dataset [14].

## Future Solutions

Advanced considerations in this space include intelligent decision support for CCTV monitoring stations, providing humans-in-the-loop with guidance about the types of strengths and weakness of artificial intelligence driven video violence detection, offering suggested courses of action dynamically given the situational parameters, and feeding real-time information about perpetrators to law enforcement once a human monitor has determined that a violent act is actually taking place [15]. The human-in-the-loop step is an important consideration for legal and constitutional considerations as well as practical ones [16]. Information for law enforcement might include the number of perpetrators, location, direction, escalation patterns (e.g. if a fist fight has escalated to include weapons), etc. To achieve systems with reasonable true positive rates that also adequately minimize both false positive and false negative concerns, fusing data from multiple sensor platforms (multiple video vantage points, audio, context data, etc.) will be an important goal [17-19]. Dynamic integration of other data sources, including livestreamed social media from an incident scene are other areas for advanced work [20-22].

## Conclusion

Artificial intelligence activity detection for violent action using large scale networks is an area of increasing interest and focus in both academic and business communities. In the wake of active shooter incidents in multiple sectors of society, the ability to rapidly detect a violent actor using massive amounts of video data produced by CCTV and other passive sensors is becoming a priority. A number of video analysis strategies have been utilized with marked success, but problems in terms of early detection of warning signs remain. Ultimately data fusion, integrating multiple types of data and feature selection that optimizes available information in these data streams is likely to be a key strategy to optimizing accuracy. Strategies that balance public safety, privacy, legal, and ethical considerations will need to be developed. This approach fits broadly into the Smart & Safe Cities initiatives, but also has implications for businesses of all sizes and rural municipalities.

# References


[1] M. Perez, A.C. Kot, A. Rocha, Detection of real-world fights in surveillance videos, ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 2662-2666.
[2] M. Bendali-Braham, J. Weber, G. Forestier, L. Idoumghar, P.-A. Muller, Recent trends in crowd analysis: A review, Machine Learning with Applications 4 (2021) 100023.
[3] J. Bruggeman, D. Weenink, Collective action in violent conflicts and other highly uncertain situations, arXiv preprint arXiv:2112.05088 (2021).
[4] R. Philpot, Beyond the dyad: The role of groups and third-parties in the trajectory of violence, (2017).
[5] P.W. Khan, Y.-C. Byun, N. Park, A data verification system for CCTV surveillance cameras using blockchain technology in smart cities, Electronics 9(3) (2020) 484.
[6] G. Oatley, T. Crick, D. Bolt, CCTV as a smart sensor network, 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, IEEE, 2015, pp. 1878-1885.
[7] D. Harkin, Regulating private sector security provision for victims of domestic violence, Theoretical criminology 23(3) (2019) 415-432.
[8] E. Pauwels, Artificial Intelligence and Data Capture Technologies in Violence and Conflict Prevention, Global Center on Cooperative Security. September (2020).
[9] O. Bazgir, D. Nolte, S.R. Dhruba, Y. Li, C. Li, S. Ghosh, R. Pal, Active shooter detection in multiple-person scenario using RF-based machine vision, IEEE Sensors Journal 21(3) (2020) 3609-3622.
[10] P.F. Wu, Y. Qu, J.J. Preece, Why an emergency alert system isn’t adopted: The impact of socio-technical context, People and Computers XXII Culture, Creativity, Interaction 22 (2008) 101-104.
[11] G.J. Smith, Behind the screens: Examining constructions of deviance and informal practices among CCTV control room operators in the UK, Surveillance & Society 2(2/3) (2004).
[12] L. Hesse, The transition from video motion detection to intelligent scene discrimination and target tracking in automated video surveillance systems, Security Journal 15(2) (2002) 69-78.
[13] P. Wu, J. Liu, Y. Shi, Y. Sun, F. Shao, Z. Wu, Z. Yang, Not only look, but also listen: Learning multimodal violence detection under weak supervision, European conference on computer vision, Springer, 2020, pp. 322-339.
[14] Ş. Aktı, G.A. Tataroğlu, H.K. Ekenel, Vision-based fight detection from surveillance cameras, 2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA), IEEE, 2019, pp. 1-6.

[15] L.M. Burgess, A. Garinger, G. Carrick, C. Systematics, Integrating Computer-Aided Dispatch Data with Traffic Management Centers, United States. Federal Highway Administration. Office of Operations, 2021.
[16] G. Matuszek, CCTV systems–technological and legal aspects. The present and the prospects for future, Zeszyty Naukowe SGSP/Szkoła Główna Służby Pożarniczej (2020).
[17] X. Li, Q. Yu, B. Alzahrani, A. Barnawi, A. Alhindi, D. Alghazzawi, Y. Miao, Data fusion for intelligent crowd monitoring and management systems: A survey, IEEE Access 9 (2021) 47069-47083.
[18] A.F. Smeaton, E. O'Connor, F. Regan, Multimedia information retrieval and environmental monitoring: Shared perspectives on data fusion, Ecological informatics 23 (2014) 118-125.
[19] S. Qiu, H. Zhao, N. Jiang, Z. Wang, L. Liu, Y. An, H. Zhao, X. Miao, R. Liu, G. Fortino, Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges, Information Fusion 80 (2022) 241-265.
[20] S. Subramani, H. Wang, H.Q. Vu, G. Li, Domestic violence crisis identification from facebook posts based on deep learning, IEEE access 6 (2018) 54075-54085.
[21] R.T. Sheffield, Facebook Live as a recordmaking technology, Archivaria 85 (2018) 96-121.
[22] P. Papailias, Witnessing to survive: Selfie videos, live mobile witnessing and black necropolitics, Image Testimonies, Routledge2018, pp. 104-120.

# INSTITUTE FOR HOMELAND SECURITY

Sam Houston State University™

The Institute for Homeland Security at Sam Houston State University is focused on building strategic partnerships between public and private organizations through education and applied research ventures in the critical infrastructure sectors of Transportation, Energy, Chemical, Healthcare, and Public Health.

The Institute is a center for strategic thought with the goal of contributing to the security, resilience, and business continuity of these sectors from a Texas Homeland Security perspective. This is accomplished by facilitating collaboration activities, offering education programs, and conducting research to enhance the skills of practitioners specific to natural and human caused Homeland Security events.

Institute for Homeland Security
Sam Houston State University



Franco, Z. (2022). **Remote Detection of Anger: CCTV and Large-Scale Sensor Networks**. (Report No. IHS/CR-2022-2047). The Sam Houston State University Institute for Homeland Security. https://ihsonline.org/Research/Technical-Papers/CCTV-and-Sensor-Networks-Remote-Detection-of-Anger