Violent Crime Detection System

Conference Paper · September 2018

DOI: 10.1109/STC-CSIT.2018.8526596

CITATIONS
3

READS
368

3 authors, including:

Yaroslaw Dorogyy
National Technical University of Ukraine Kyiv Polytechnic Institute
20 PUBLICATIONS **18** CITATIONS
SEE PROFILE

Vadym Kolisnichenko
National Technical University of Ukraine Kyiv Polytechnic Institute
3 PUBLICATIONS **8** CITATIONS
SEE PROFILE

Violent Crime Detection System

Yaroslav Dorogyy, Vadym Kolisnichenko, Kseniia Levchenko
Department of Automation and Control in Technical Systems
The National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
Kyiv, Ukraine

***Abstract*— The violence is one of the problems of our society. Violence prevention requires violence detection first, which needs a lot of human resources, and therefore it cannot be done instantly and ubiquitously. With the development of technologies it became possible to detect violence with minimal human interaction. In this paper, we describe principles and methods to build an automated system for violence detection. Proposed system is able to make violence detection simple, instant and ubiquitous. It was designed to recognize violence not only during the time it occurs but to predict it for some time before, basing on interpretation of verbal and nonverbal signals of a person. In addition, the system may use different external data sources to get information about the persons and environment to include it into reasoning process, therefore increase prediction accuracy.**

Keywords*— *expert system, artificial intelligence, behavior analysis, violence detection.

I. Introduction

Violence has always been part of mankind and was expressed in different forms. Violence reduction is an important problem, different experts suggest that violence is a hidden cause of poverty [1]. Therefore, it is insisted on the importance of reducing everyday violence.

Violence recognition, which is done only by a human is not effective and takes a lot of resources. Automated system for doing this task should be built. However, the system should not be fully autonomous, it should be customizable, configurable and extendable, and therefore new detection mechanisms can be included when they are developed, and the expert can tune the system to make decisions about possible crimes more accurate.

Automated system for violent crime detection is a necessary part of modern world. The system would help to reduce crime level by making detection simple, instant and ubiquitous. It will allow to predict certain kind of violence beforehand basing on different analysis techniques. The system must be able to integrate with other systems of smart city monitoring and control to make data sharing and management more efficient.

Existing solutions are very limited to the specific source of information [2], and are built to be fully autonomous. They usually rely on video data and machine learning techniques to judge about violence, without using other sources of information (e.g., crime statistics, information from social networks, police databases, hospital databases) and without implementing knowledge about social behavior. Some solutions are broader that crime detection [3], they are about public safety, but they may have same limitations.

II. Domain Analysis

Violent crime is a crime in which an offender uses or threatens force upon a victim. Different countries categorize crimes differently, therefore depending on the jurisdiction, violent crimes may include: homicide, murder, assault, manslaughter, sexual assault, rape, robbery, negligence, endangerment, kidnapping (abduction), extortion, and harassment.

Violence detection can be done by a regular person without any efforts just by seeing or hearing it. A person can see when someone is fighting, shooting or behave abnormally. Also, they can hear screams or cries and hazard a guess about possible crime.

The hardest thing is to recognize violence some time before it happens. It can be done by introducing additional sources of information or by processing (observing) currently available information (visual and auditory information, obtained by senses) more meticulously.

The last one is done by trained people to understand nonverbal and verbal signals of other people [4]. Processing such information gives cues about intents and emotional state of the person in order to anticipate his subsequent actions, which may include a violent character [5].

Verbal communication is the use of sounds and words to send a message. Verbal aggressiveness or abuse can precede the violence. Different languages and slangs could be used, therefore verbal analysis requires understanding them.

Nonverbal communication is the process of sending and receiving wordless messages. It can give information about person's emotional state and intents. It can be separated into different types [6]: haptics (touch), kinesics (body movement, which includes facial expressions, gestures, postures), vocalics (paralanguage), chronemics (the use of time), proxemics (the use of space), oculesics (eye contact). Appearance is also a part of nonverbal communication, it includes hairstyles, clothes, jewelry, body piercings, tattoos, and other. Even small part of such information could tell us a lot about a person and his tendency to do violence, for example, seeing a person in the expensive suite will not make us nervous, unlike a person with bandanas, which may symbolize a gang affiliation. Of course, some cues may vary dramatically between cultures, so

information about the place where an event occurs should be also considered.

Current location, crime statistics in the current region, information about nearby persons, information about their criminal past and medical diagnoses – only a part of additional information that can be used during violence detection to make it more accurate.

Using all this information, an expert can predict the violence beforehand with a high level of confidence. Additional devices may be also used to recognize violence during night time or in areas with limited visibility.

III. System High-Level Overview

The system automates methods described in the previous section that are done by an expert to recognize a violent crime. The information the system would operate on could be divided into primary and supplementary information.

Primary information is internal information that is taken directly from the place where potential crime may occur (e.g., auditory and visual information, location, time).

Supplementary information - external information that is not gained from the place of the event, but stored outside. It is information from different databases (e.g., police, hospital databases), social networks, statistical information, etc. Such information could tell a lot about a person in a scene: criminal past, health issues, mood, etc. Handling most of the supplementary information introduces big latencies due to network communication and big data processing, therefore advance mechanism should be developed to manage it.

Auditory information is audio signals from the place potential violence may occur. It may contain speech, screams or shoots, which is critical for violence recognition. Aggressive speech or speech that contains threats should be interpreted as the elevated probability of violence occurrence. Speech processing must also involve translation, so other languages and slangs can be interpreted correctly.

Visual information is formed by image taking. Other techniques may be also used, like thermal imaging and image intensification. Thermal information may also help to detect the emotional state of a person or give some additional features about him [7, 8] that could be used during reasoning. Night-vision technologies can make violence detection possible at night time. There is no guarantee that the use of specific techniques will improve the detection accuracy by so much that overheads become negligible.

Different environment limitations may lead to limitation of audio sensors, in that situation other techniques may be used to gather audio information. These techniques are based on getting audio information from visual. The first technique is lip reading [9]. It can be also used to compare its results with results obtained by audio sensors to assign audio speech data to relevant persons detected on a scene. The second method is known as visual microphone [10], it allows to recognize the sound from vibrations that can be visually observed, however it may be useless in surveillance conditions.

The location data is used together with crime statistics among regions, got from supplementary sources, to include the possibility of violence occurrence into detection process. Location may be represented by the GPS coordinates or by categorical values like “business borough” or “South East”.

Simplified violence detection process flow is shown in Fig. 1. Primary information is handled inside information processing block to get intermediate data representation (detect faces, actions, emotions, stress, calming nervousness, territorial display behavior, angry speech, screams, etc.). Primary information or intermediate data may be sent to supplementary block to receive additional information. Supplementary information block operates on information that is stored outside the system (usually databases). After passing all information processing layers the data arrives at the classifier, which decides whether violence is present on the scene.

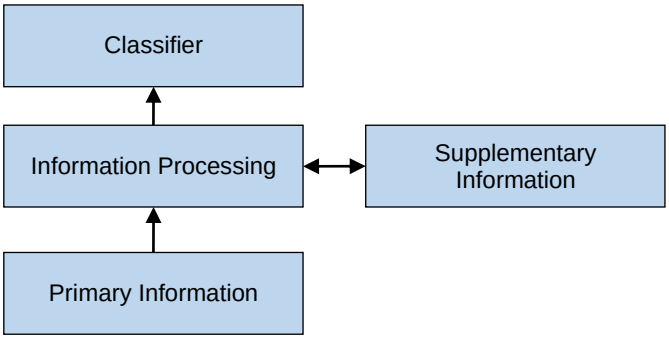


Fig. 1. Simplified violent crime detection process flow.

The system could be built in two ways. The first one is by using information processing block as a single component (general algorithm, neural network) for processing all primary information together. It will extract all features and build its own understanding what is important for reasoning.

The second way is to divide it into components, therefore each component does its own specified job. Components have similar input-output interface, they process data and send results to the next components. Their behavior can be defined by different computer vision algorithms, Bayesian methods, artificial neural networks, spiking neural networks [11], fuzzy logic approaches and other methods. Some supplementary information components, for example, may just communicate with external databases to search information. Such architecture would allow to tune, replace and improve components independently. Also, it makes the visualization of results (outputs) of each component simpler.

Recurrent mechanisms should be involved to process information as sequential, it will lead to better understating of human behavior by the system, which may increase the accuracy.

IV. Processing Example

An example of component-based architecture is presented in Fig. 2. For simplicity, we took only the small subset of components here and used image as a primary information source only. In this section, we will go through the process of

extracting and using human face data to detect emotions and to get additional information that can be used during classification.

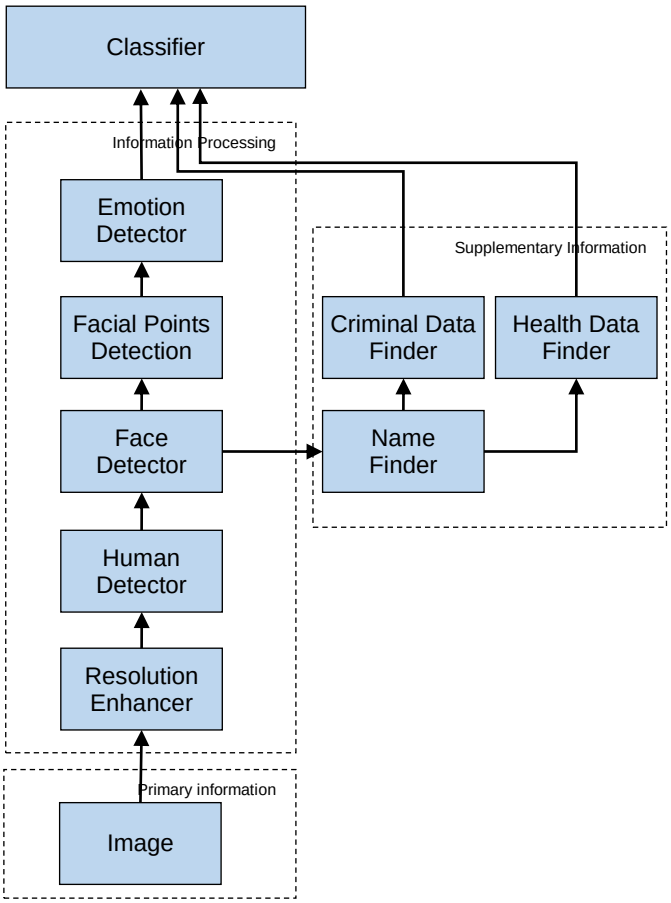


Fig. 2. Simplified process of extracting and using face data during detection.

Different existing or new algorithms may be used inside each component. Several existing solutions were used for components in this example, except human detector, emotion detector, criminal data finder and health data finder components. An image from Multi-Camera Action Dataset (MCAD) was used to show results of each processing step.

The original image is displayed in Fig. 3. This image will be processed by different components to extract useful information that will be used during classification.

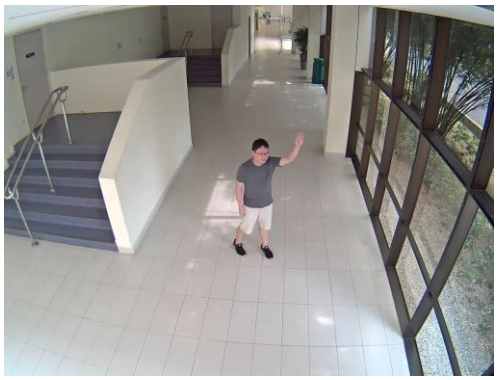
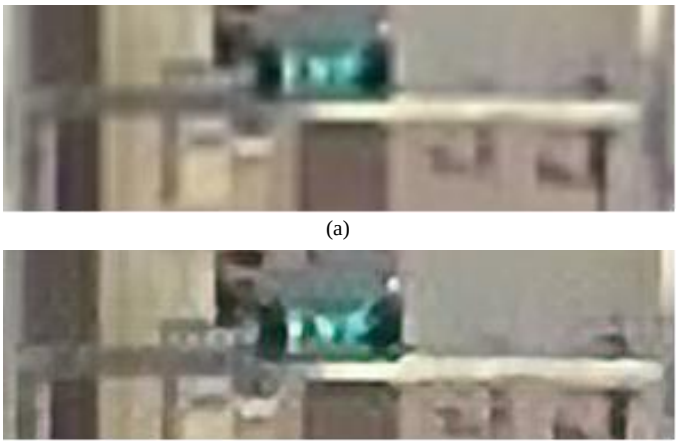


Fig. 3. Image frame from Multi-Camera Action Dataset (MCAD).

The image is then processed by resolution enhancer component. Existing solution [12] was used for this task. Parts of the image before and after enhancing are presented in Fig. 4 and Fig. 5.



(a)

(b)

Fig. 4. Exit sign before (a) and after (b) enhancing.

(a) (b)

Fig. 5. Person's face before (a) and after (b) enhancing.

The next stage is to detect a person on the scene (Fig. 6). This region is then used to detect a face (Fig. 7a). Here, face detection is made by using Haar Cascades [13].

Fig. 6. Human detection.

Facial points are detected (Fig. 7b) using existing technique [14]. They are extracted to be processed in the next component to detect emotions, however, it is also possible to replace these two blocks (facial points detection and emotion detection) with a single one, therefore emotions will be extracted not from facial points coordinates, but from pixels. Other components are also involved into emotion recognition process, which are based on methods described in section 2 and 3, but they are not used in this example.

(a) (b)

Fig. 7. Face detection (a) and facial landmark detection (b).

Emotion detection component recognizes anger, disgust, fear, happiness, sadness, surprise, and contempt. For given facial points it produces "happiness" result.

Neural network based solution was used to recognize a person by his face. Criminal data finder and health data finder components were built to work with databases which consist of fake records for this example, so their outputs were used during classification.

In the next step the classifier makes a decision based on data received from the previous components.

V. Conclusion

Proposed system automates violence detection and makes it simple, instant and ubiquitous. It was designed to detect violence not only during the time it occurs but to predict it for some time before, basing on the interpretation of verbal and nonverbal signals of human. In addition, the system uses different external data sources to get information about the person and environment to include it into reasoning process.

Automated system for violence detection could be extended to recognize other types of crime by adding new sources of information and processing algorithms. At first, it may be crimes, detection of which requires similar information sources and similar techniques (e.g., using visual data for motor vehicle theft detection). Then, it may be improved to detect other crimes (e.g., white-collar crimes), using completely different data types and methods.

Predictive policing methods [15] make possible to identify potential criminal activity before it starts based on mathematical and analytical techniques. Combining these methods with current system functionality may increase the effectiveness of both crime detection and predictive policing activities.

References

[1] G. A. Haugen, V. Boutros, "The Locust Effect: Why the End of Poverty Requires the End of Violence," Oxford University Press, 2014.

[2] G. Mu, H. Cao and Q. Jin, "Violent Scene Detection Using Convolutional Neural Networks and Deep Audio Features," Pattern Recognition: 7th Chinese Conference, CCPR 2016, 2016, pp. 451-463.

[3] Intelligent Video Analytics Platform. [Electronic resource]- Access mode: http://www.nvidia.com/object/intelligent-video-analytics-platform.htm.

[4] J. Navarro, M. Karlins, "What Every BODY is Saying: An Ex-FBI Agent's Guide to Speed-Reading People," William Morrow Paperbacks, 2008.

[5] National Research Council, "Threatening Communications and Behavior," National Academies Press, 2011.

[6] R. G. Jones, "Communication in the Real World: An Introduction to Communication Studies", University of Minnesota Libraries Publishing, 2016.

[7] S. He, S. Wang, W. Lan, H. Fu and Q. Ji, "Facial Expression Recognition Using Deep Boltzmann Machine from Thermal Infrared Images," Humaine Association Conference on Affective Computing and Intelligent Interaction, September 2013, pp. 239-244.

[8] D. Cardone, P. Pinti and A. Merla, "Thermal Infrared Imaging-Based Computational Psychophysiology for Psychometrics," Computational and Mathematical Methods in Medicine, 2015.

[9] J. S. Chung, A. Zisserman, "Lip Reading in the Wild," Asian Conference on Computer Vision, 2016.

[10] A. Davis, et al., "The Visual Microphone: Passive Recovery of Sound from Video," ACM Transactions on Graphics, No. 4, Article 79, 2014, pp. 1-10.

[11] Y. Dorogyy, V. Kolisnichenko, "Designing spiking neural networks," Modern Problems of Radio Engineering, Telecommunications and Computer Science, vol. 6, 2016, pp. 124-127.

[12] Super Resolution for images using deep learning. [Electronic resource] - Access mode: https://github.com/alexjc/neural-enhance.

[13] Face Detection using Haar Cascades. [Electronic resource] - Access mode: http://docs.opencv.org/trunk/d7/d8b/tutorial_py_face_detection.html.

[14] Real-time facial landmark detection with OpenCV, Python, and dlib. [Electronic resource] - Access mode: http://www.pyimagesearch.com/2017/04/17/real-time-facial-landmark-detection-opencv-python-dlib/.

[15] S. Brayne, A. Rosenblat and D. Boyd, "Predictive Policing," Data & Civil Rights: A New Era Of Policing And Justice, 2015.