Video Surveillance and Privacy: A Solvable Paradox?

Belainesh Mehari

INTRODUCTION

Video surveillance concerns models, techniques, and systems for acquiring and processing videos about the external world, detecting targets along time and space, recognizing interesting or dangerous situations, generating real-time alarms, and recording meaningful data about the controlled scene. While the target of surveillance systems can be the whole environment, for example, natural events or moving vehicles, the most complex and addressed target is surely the human being: where people are, what movements or actions they are performing, what their behavior is, and whether that behavior is affecting security or safety.

Through experiments on action recognition and natural language description, we show that the paradox of surveillance and privacy can be solved by artificial intelligence and that respect for human rights is not a chimera

RESEARCH FEATURE

when surveillance systems were governed by human inspectors only and stored video frames that could later be retrieved and used against the privacy of the depicted persons. At the beginning of the 21st century, instead, computer vision advancements made automatic processing effective both in surveillance and privacy-preserving solutions.

Most privacy-preserving approaches were oriented to visual anonymization, achieved by covering face appearances on pictorial data. Also, pseudoanonymization techniques were developed, such as encryption and data scrambling, to store privacy-concerning information in a way impossible to be retrieved by human eyes

PRIVACY REGULATION

Many worldwide regulations concern the principle of data minimization (that is, the need to use a minimal amount of data), opting out (that is, the option to delete data if required by the data owner), the limitation of data storage (for example, in the GDPR), and the inappropriate use of data. New proposals, such as the Artificial Intelligence Act, recently approved by the European Commission, instead, make explicit reference to machine learning technologies and regulate the deployment and usage of highrisk applications (for example, health, security, enrollment, education, and finance) in order to assure the key Points of trustworthy AI, that is, the principles of human oversight, transparency, accountability, compliance with human rights, and privacy. They pose several limitations on the design of new AI-based applications, such as video surveillance

The 2000s: The boom of computer vision video surveillance

In the first years of the 21st century, video surveillance spread for three concurrent reasons: hardware availability, computer vision improvements, and the need for social security created by the terrorist attacks in the United States and Europe

People detection started to be formulated by considering people target models recognizable by a two-class classifier (human presence versus non human presence). The birth of people detection started probably with the seminal work of Dalai and Triggs, in 2008.6 Since that work, hundreds of approaches concerning people were developed. They can be categorized according to three different aspects: 1. Which (handcrafted) features to employ? Several general purpose descriptors were proposed, such as histogram of oriented gradients6 covariance and structured part-based descriptors.7 2. Which classifier? Detectors should be coupled with suitable classifiers, such as neural networks, support vector machines (SVMs), LogitBoost, and AdaBoost, showing a true boom in pattern recognition techniques for people detection. 3. Which search space? Often, searching everywhere is not necessary; thus, many proposals focused on improving both efficiency and precision/recall, for example, with sliding window or hierarchical (pyramidal or multiresolution) windows. As well, the idea of region proposals started to be defined in order to look at regions according to a probability density function p(X | Z), where X is the state of the person and Z is the observation. For instance, in Gualdi et al.,8 a multistage paticle window provided fast and accurate multistage probabilistic sampling for boost and SVM classifiers.

The past decade: Deep learning, surveillance, and privacy

In the latest decade, the paradox between video surveillance adoption and privacy was exacerbated by the rebirth of neural networks and the incredible results of deep learning after the ImageNet challenge10 and the development of convolutional neural networks (CNNs) capable of classifying, recognizing, and detecting targets. While video surveillance systems became widespread and widely used, the debate on privacy and human rights shifted toward social data collection and dual use of data, culminating in scandals like the Cambridge Analytica one in 2018. The technology of people detection, tracking, recognition, and action analysis became very sophisticated thanks to several results.

PRIVACY PRESERVING ACTION RECOGNITION

The task of modifying images to not be visible by human experts, leaving inside useful semantic content for a task. Examples of image processing approaches are pixelization, pixel scrambling, and shape hiding in video

PRIVACY PRESERVING IMAGE TEXTUAL DESCRIPTION

Working on privacy-compliant data and providing useful tasks for people surveillance and action analysis could be cumbersome when systems are starting to deal with foundation models or, in general, large pretrained systems that provide textual descriptions of a scene. Image captioning is a new way of generating image descriptions in a natural language way. Also in this case, networks can be trained on anonymized visual data and reduce the loss in performance by using knowledge distillation

Conclusion

Paradox of surveillance and privacy could be solved.
First, video understanding can now be effectively provided by machine learning approaches, and there is no need for continuous human monitoring. Human oversight is necessary only in case of dangerous situations
Second, to understand what people do, we do not need information about their identity, their face, or their appearance.
Third, we can start discussing the controllability of privacy for pretrained networks, constraining them to give answers in both privacy-by-design and privacy-by-default methods. New attempts show that this could be achievable, and we hope that this will be the future of AI-based systems: to be designed for human well-being and thus for human security and safety, without affecting human rights and, in particular, the freedom of privacy.

Authors

Rita Cucchiara , Lorenzo Baraldi , Marcella Cornia , and Sara Sarto , University of Modena and Reggio Emilia

REFERENCES

1. S. Rodotà, "La privacy tra individuo e collettività,

2. S. Warren and L. Brandeis, "The right to privacy,"

3. C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-time tracking of the human body,

4. K. W. Bowyer, "Face recognition technology: Security versus privacy,

5. A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction,"

6. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection

7. P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral channel features,

8. G. Gualdi, A. Prati, and R. Cucchiara, "Multistage particle windows for fast and accurate object detection,"