Towards Systematic Evaluation of Privacy Risks in IoT Video Analytics

Hadleigh Schwartz
University of Chicago Computer Science Master's Thesis

ABSTRACT

Today's video analytics systems can parse scenes and human activity from videos collected by IoT cameras and are increasingly being deployed for critical tasks such as security and traffic management. Yet their ability to continuously monitor and identify individuals also raises significant privacy concerns. In this work, we propose a method for systematically evaluating the privacy risks of video analytics using real world IoT surveillance video. Our method takes into account policy, edge computation, and camera placement constraints to explore possible configurations and risks. Our evaluation shows that gait recognition is a significant privacy threat that should be considered alongside facial recognition and accounted for in modern privacy policies. Furthermore, we find edge-based verification methods still perform significantly worse than those methods requiring cloud computing. Lastly, we find evidence that color recognition, which we use to identify individuals' shirt colors, can be a helpful modality for increasing identification confidence and efficiency. We believe that the approach we present provides a starting point for reasoning about privacy risks of IoT video analytics under various policy and computation constraints, in both the present and future.

1 INTRODUCTION

In the last two decades, IoT cameras have become ubiquitous. These cameras are part of a growing class of everyday objects that are embedded with sensors and processing and communication abilities, forming an "internet of things" around us. In recent years, IoT cameras have become not only widespread but also more computationally powerful. Fueled by innovations in edge computing hardware, IoT cameras can process much of the video they collect locally (even running machine learning models) rather than having to stream video to servers in the cloud for processing. This proliferation of powerful IoT cameras has coincided with the development of computer vision models for video analytics. Each year, these models - for tasks including object detection, segmentation, face recognition, pose estimation, gait analysis, scene recognition, and social interaction recognition - become faster, more accurate, and smaller in storage size.

The immense data collection power of IoT cameras, coupled with the data analysis power of modern machine learning, has enabled IoT camera deployments to produce previously unseen quantities and types of information. A growing class of applications leverages machine learning models to extract information from video captured by IoT camera endpoints. These *IoT video analytics* systems are increasingly being deployed by governments and businesses to improve interactions with city spaces and enhance safety and security. However, the potential for these systems to do good is in tension with concerning privacy risks.

In the absence of explicit protection measures or regulations, the entities who deploy IoT video analytics systems can continuously and passively monitor all individuals who pass through their cameras' fields of view. They may collect and retain information on people's behaviors, relationships, locations, and identities. As IoT video analytics systems spread, understanding the extent of the privacy risk they pose becomes critical.

In this work, we propose a *blackbox evaluation of risk*. The goal of a blackbox evaluation is to explore the space of possible IoT video analytics system configurations and understand privacy risk under each. We present a method for systematically exploring this space using three practical deployment constraints - policy constraints, edge computation constraints, and camera deployment constraints - to narrow down practical configuration candidates. For each configuration, we evaluate privacy risk using real-world IoT surveillance video. This method provides a starting point for reasoning about privacy risks of IoT video analytics under various policy and computation constraints, in both the present and future.

In our evaluation, we consider face, gait, and shirt color as "identification modalities." We implement face recognition, gait recognition, and color recognition pipelines on state-of-the-art edge and cloud hardware. This forms a re-configurable IoT video analytics testbed in which we validate our proposed evaluation method.

We find that while policies and camera constraints can prevent the use of face recognition, this does not significantly reduce privacy risk due to the flexibility and power of gait recognition. In our experiments, gait recognition using cloud computing outperforms all other identification methods. Furthermore, it is amenable to modern-day surveillance scenarios because it is less sensitive to camera position than face recognition. We also find that, while the color of a person's shirt cannot alone identify him or her, it can be used as a "secondary modality" in conjunction with gait or face recognition to reduce processing in the cloud and increase identification confidence.

2 BACKGROUND AND RELATED WORK

The Rise of IoT Video Analytics Systems The global video analytics market is projected to grow from $6.35 billion to $28.37 billion in value in the next seven years alone [35], as public and private entitites deploy video analytics systems for critical tasks. Researchers in industry and academia have developed IoT video analytics systems for enhancing safety and security. Such applications include traffic accident detection [1], fall detection in nursing homes [6], COVID-19 social distance monitoring [57], and crime and abnormal activity detection [30, 33, 59]. Many governments are deploying IoT video analytics systems as part of smart city

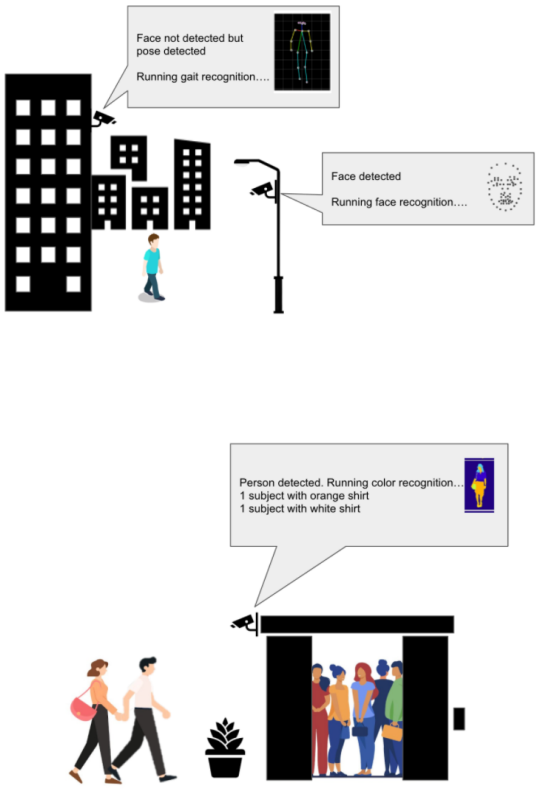


Figure 1: IoT video analytics can continuously and passively monitor human activity, posing privacy risks. Consider how different identification modalities can be leveraged to identify individuals, like in the two scenarios above.

programs, with the goal of improving interactions with city spaces and promoting sustainable urban development [11, 38] .

Video Analytics and Invasions of Visual Privacy Privacy in computer science most often refers to data privacy - a person's ability to choose when, how, and to what extent her personal information is shared with others. This information can range from her name and location to her online or real-life behaviors [20]. Visual privacy [47, 51] is a form of data privacy that encompasses a person's right to control how his or her visual information (i.e, information in the form of videos and images) is collected and used. Recent events, like Clearview AI's collection of over 20 billion face images for facial recognition [5, 32] and the New York City Police Department's use of 15,280 cameras for ML-aided surveillance [36], have drawn attention to the large-scale tracking and recognition capabilities of IoT video analytics systems.

Even systems that explicitly avoid usage of biometric identifiers, like some behavior and anomaly detection systems [59], may make individuals uncomfortable. Surveys have found that even when people favor the use of IoT video analytics for perceived benefits such as enhanced security, they are concerned about lack of notification and consent, potential unauthorized secondary use of visual data for purposes such as tracking and verification, and generally, how their visual data is treated, accessed, and stored [25, 44, 64].

Privacy Policies Visual privacy invasions are just one aspect of growing misuse of personal data. In light of this trend, governments and institutions are increasingly implementing policies to regulate the types of personal information that entities can analyze, transmit, and store. Many of these regulations directly apply to IoT video analytics systems. The European Union's General Data Protection Regulation (GDPR) is possibly the most sweeping data privacy policy to date. The GDPR prohibits the processing of any information "relating to an identified or identifiable natural person," except under special cases in which consent has been obtained or a rigorous assessment that weighs privacy-security tradeoffs has been carried out. GDPR's definition of biometric identifiers as "personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person," [37] is arguably broad enough to encompass any form of data collected or generated through IoT video analytics, from images of faces to gait embeddings.

There is no federal law comparable to the GDPR in the United States, though individual states such as Illinois, Washington, Texas, Virginia, and New York have passed their own legislation on biometric data usage [31]. The Illinois Biometric Information Privacy Act, for instance, defines a biometric identifier as "a retina or iris scan, fingerprint, voiceprint, or scan of hand or face geometry" and prohibits the collection of biometric identifiers without explicit, informed consent of the identifier's owner [4]. Notably, this definition of biometric identifiers does not include gait. Despite these policy examples, many IoT video analytics systems in the United States remain unregulated, especially those operated by governments. NYPD's Domain Awareness System, for example, legally streams and stores video collected by networked CCTV cameras for ML analysis in the cloud.

Some non-government institutions have developed privacy policies for their deployments. For instance, the SAGE (Software-Defined Sensor Network) project, established by Northwestern University in 2019, convened community stakeholders to develop a privacy policy for their network of urban sensor nodes located in the city of Chicago. They found that citizens were weary of video being stored in the cloud. Thus their policy allows only one raw image per 15 minutes to be sent off the edge. All other video is processed on the node itself and deleted immediately after processing [3].

Existing Work on Privacy Risks in IoT Video Analytics Most work relating to privacy risk in IoT video analytics falls into one of two categories: (1) detailing new forms of visual privacy invasion or (2) proposing specific privacy safeguards for IoT video analytics systems.

In the first category, computer vision methods for analyzing individuals are constantly being developed and improved. These methods are often discussed in contexts outside of privacy despite having clear visual privacy ramifications. Facial recognition is perhaps the most well-known and widespread of these technologies.

Other techniques use soft biometrics - personal attributes like gender, age, ethnicity, hair color, height, weight, and clothing - to recognize or re-identify individuals [22, 24, 27, 28, 39, 41, 55, 66]. A recent paper [16] has even claimed that it has achieved a computer vision system for personality identification.

Perhaps the most alarming new mode of visual privacy invasion is gait recognition. Gait, or the way an individual walks, has been shown to be a highly distinctive feature and is the subject of much computer vision ML research [10, 13, 26, 34, 42, 43, 50, 53, 58, 61]. Gait recognition provides a robust alternative to facial recognition, which may not always be possible. Consider the difficulties the COVID-19 pandemic, which necessitates widespread mask-wearing, has posed for facial recognition. Furthermore, while someone may obscure their face, consistently changing the shape of their body and that they walk is not as straightforward. Gait recognition technology is already being deployed at large scales in Beijing and Shanghai [48]. Given that gait recognition will likely gain traction in the coming years, we consider it as a key recognition modality in our evaluation.

Significant work has been done to mitigate invasions of visual privacy through computer vision-based mechanisms. [9, 12, 14, 17, 19, 40, 56, 60, 62, 63, 65], propose to detect privacy-sensitive regions of images and videos using existing person and face detection methods, and then encode these regions in a manner that reduces their sensitivity (e.g., through masking, blurring, encryption, denaturing). [46] obscures people in scenes to varying degrees based on the context. In a similar vein, [52] obscures people to varying degrees based on their expressed privacy preferences. The goal of these techniques is to protect individual privacy while preserving the overall utility of the video. In contrast, [67] prevents a scene from being photographed all together using smart LEDs.

Other solutions focus on data management and storage in IoT video analytics systems. Several works propose ways of enhancing 'digital sovereignty,' or ensuring that a citizen has control over the data collected on him or her. [23] discovers IoT cameras and notifies individuals of their presence. [18] obfuscates individuals in video streams and allows users to determine who may de-obfuscate them.

While these "attack" and "defense" papers are critical to understanding the privacy risk posed by IoT video analytics, they differ from our work along two main axes. Firstly, they often involve arbitrary attack scenarios and specific system configurations (e.g., malicious facial recognition using cloud processing). Our goal is to zoom out from specific assumptions and scenarios and understand the factors that enable and impact risk itself: machine learning models, edge and cloud computing, policies, and IoT cameras. Secondly, the main contribution of these papers is demonstration of novel risks. In this paper, we do not aim to uncover new risks but rather to demystify the process of evaluating IoT video analytics privacy risk in quickly evolving policy and computation landscapes.

3 THE NEED FOR A BLACKBOX EVALUATION OF RISK

Whitebox and blackbox testing are two techniques for evaluating systems. In whitebox testing, specific internal structures are tested to ensure their functionality. In blackbox testing, the inner workings of the system are not known. Instead, testing is conducted by observing system output in response to specific inputs and execution conditions [45].

These terms accurately describe what we see as two approaches for evaluating the privacy risks posed by IoT video analytics systems. Most existing works on video analytics privacy risk take what we deem a whitebox approach. These works assume a specific configuration and then evaluate the risk this configuration poses using test data. A configuration refers to an end-to-end pipeline consisting of IoT cameras, machine learning models, and edge and cloud computing hardware that is used to analyze video. For example, many papers assume a configuration in which cameras are positioned so that faces are visible, allowing cloud-based facial recognition models to be used.

We believe that these whitebox approaches alone are inadequate because they address just a subset of the space of possible configurations and privacy risks. In reality, there are countless possible configurations, each entailing specific types of degrees of privacy risk. IoT cameras themselves can be positioned in various scenes, at varying heights, angles, etc. Humans have various modes of identifiable information embedded in their appearance and behavior, from their gait and facial features to the color of their clothing. An IoT video analytics deployment can consist of various computer vision models distributed across various computers, at the edge and in the cloud. How these models are distributed and deployed will depend on a combination of policy and computation constraints (i.e., where models are allowed to be placed, if at all, and where models can be placed), both of which vary and will likely evolve in the coming years.

Those who have stake in evaluating the privacy risks posed by IoT video analytics systems, such as privacy policy-makers, are not always given specific pipelines to analyze and cannot necessarily base their decisions on individual deployments. Instead, they need a high-level understanding of privacy risk that takes into account the diversity and flexibility of modern day IoT video analytics systems.

This type of understanding can be achieved through a blackbox evaluation of risk. In a blackbox evaluation, we have no knowledge of the video analytics system's configuration. To examine privacy risk, we must vary our input and conditions to explore a range of practical configurations. In this paper, we establish a method for identiying these configurations to conduct a blackbox evaluation of risk.

3.1 Navigating the Blackbox Evaluation Using Policy, Computation, and Camera Constraints

We propose that three types of constraints- policy, edge computing, and camera constraints - can be used to systematically enumerate configurations. We believe these categories reflect the many practical considerations required to operationalize a deployment and that they also provide a logic for evaluating risk. To see how this is the case, suppose a system designer seeks to deploy an IoT

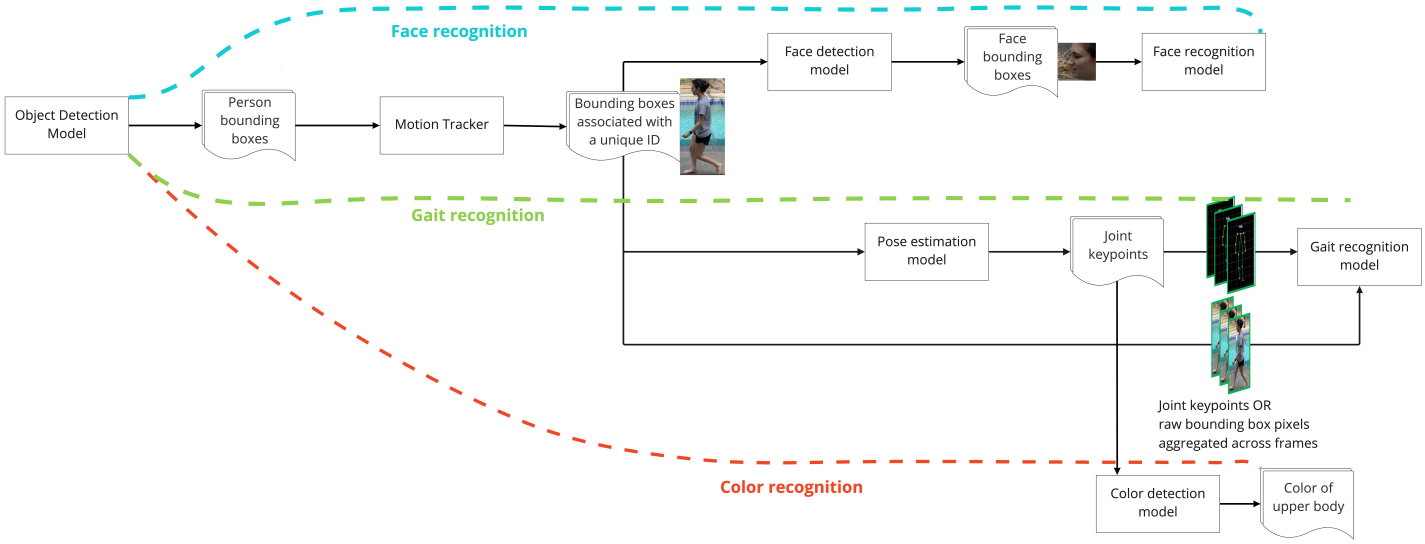


Figure 2: We consider three identification modalities: face, gait, and clothing color. This dependency graph illustrates the flow of computer vision models needed to identify an individual based on each modality.

video analytics application for verifying the presence of specific individuals on the street.

Camera constraints Firstly, the system designer is limited by camera parameters that influence what information can even be observed in individuals. Are faces visible, enabling facial analysis? Is the resolution of the portion of the image containing the face high enough for facial feature extraction to succeed? Are people walking throughout the scene, enabling gait analysis? If people are walking throughout the scene, are they walking parallel to the camera or at unconstrained angles? These questions are determined by camera parameters, like the distance of the camera from subjects, the camera's location (e.g., overlooking a sidewalk or courtyard), or its resolution. They determine the identification modalities (e.g., face, gait, color) that can be used, which influences the types of models that need to or can be deployed (e.g.,object detection, face detection). This constraint is the most specific to an individual deployment, and one system may even consist of many cameras with many different parameters.

Edge computing constraints Secondly, the system designer faces computation constraints that limit which implementations of model types she may deploy and where those models can be run - the edge or the cloud. A typical machine learning video analytics application involves IoT cameras streaming video to the cloud, where powerful GPUs can run machine learning models and aggregate insights. However, to reduce network strain and scale their deployments, many system designers are shifting processing to the edge, or the IoT camera themselves. Under this paradigm, IoT cameras - equipped with compact GPU modules - can run video analysis locally, meaning only extracted information must be sent off the device. Like its cloud counterpart, edge video analytics must be done in near real-time, invoking the classic ML accuracy-inference speed tradeoff.

Furthermore, hardware at the edge is typically significantly less powerful than that at the cloud. Though models specialized for mobile computing environments exist, they are usually less accurate than the models intended for larger GPU's. Some particularly complex models have yet to be adapted for mobile contexts. During our experiments, for instance, the edge GPU ran out of memory when attempting to run an existing gender, color, and height identification model [29]. Thus, even if the system designer chooses the fastest, smallest models for the edge, she may often still need to leverage cloud computing for parts of the processing.

Policy constraints Thirdly, the system designer may need to take into account the policies restricting her deployment possibilities. Given local or institutional regulations, what types of information are allowed to be transmitted and stored, and what processing is allowed to be run where? This constraint is the broadest in that it will apply to many different deployments depending on high-level factors like entity type (e.g., government or private institution) or geographic location.

Policy, edge computing, and camera constraints each narrow down the space of possible deployment configurations. Each also affects the quantity and type of identity factors embedded in human subjects that can be extracted, which in turn impacts the deployment's ability to successfully verify individuals. Thus, each combination of camera, policy, and computational constraint, and each configuration it permits, can be associated with different forms and degrees of privacy risk. If we can enumerate a range of configurations and establish a measure of privacy invasion under each, we can understand privacy risks broadly posed by IoT video analytics. In a sense, we use the very diversity of policy and system considerations that makes understanding the privacy risk of IoT analytics systems difficult, as a roadmap for our blackbox evaluation.

4 OUR BLACKBOX EVALUATION

We have established that policy, edge computing, and camera constraints constitute a method for enumerating configurations. In this

section we detail how we practically apply our method to conduct a blackbox evaluation of privacy risk in IoT video analytics systems.

Identification Modalities and Models Considered In our evaluation, we consider three modalities: face, color, and gait. Face and gait are biometric identifiers explored in many computer vision papers. The color of a person's clothing can also be used to distinguish a person. For instance, if the goal of a deployment is to track an individual across multiple checkpoints along a walking commute, the color of the person's shirt can be paired with face or gait information to boost tracking confidence.

Face, gait, and color recognition all require preliminary analysis by other models. For instance, identifying someone's gait requires detecting and tracking them throughout the video in the first place. To capture these relationships, we present a dependency graph of computer vision model types and inputs and output (Figure 2). This dependency graph formalizes the flow of models needed to analyze a person with a specific modality. Each node represents either a model or a form of extracted information. In total, we consider seven types of machine learning models: object detection, motion tracking, face detection, face recognition, pose estimation, gait recognition, and color detection.

Impact of Camera Constraints Camera constraints inherently determine which modalities (color, face, or gait) can be applied in a configuration. We organize camera constraints along three axes: quality of the extracted image of the face, walking manner, and whether samples were collected on the same day. The latter consideration is relevant because, if samples were collected on the same day, we might assume that people are wearing the same shirt (a necessary assumption for color recognition).

Edge Computing Constraints in our Testbed We deploy publicly available models for the seven computer vision tasks discussed above on two different GPU's: one comparable to those that are available at the edge, and one comparable to those that are used for cloud computing. For our edge hardware, we use a NVIDIA Jetson AGX Xavier Developer Kit, a state-of-the-art GPU module intended for edge computing use cases. To simulate cloud hardware, we use a NVIDIA Titan RTX GPU. When possible, we test multiple models for each task to explore accuracy-speed tradeoffs. These models and their inference speeds on both edge and cloud hardware are listed in Table 1.

In our experiments, we set 5 FPS as the threshold for what is considered real-time. While this is a low framerate, it is sufficient to capture motion and activity. Furthermore, we found that gait recognition performed well at framerates at or above 5 FPS, but began to suffer at lower framerates. Note that we assume that GaitNet can only be run in the cloud, since a gait recognition runtime of over 21 seconds on the edge would significantly interfere with local real-time processing of video. However, all processing leading up to gait recognition with GaitNet (i.e., person detection and tracking), can be run on the edge.

Policies Considered We consider three policies:

- Policy 1: None
- Policy 2: No facial recognition
- Policy 3: No streaming of raw images or video to the cloud

Policy 1 is relevant for adversarial cases or cases in which there are genuinely no policies pertaining to video surveillance. Policy 2 is one of the most common policies today (e.g., Illinois Biometric Information Privacy Act). Policy 3 may be motivated by concerns over storing sensitive video in bulk in the cloud or even network bandwidth usage considerations. For instance, the Chicago Array of Things, whose developers consulted with many community stakeholders to establish privacy policies, only sends one image off the edge every fifteen minutes and deletes all video immediately after processing it on the edge [3]. This also makes the deployment much more scalable, as individual nodes no longer need to consume large amounts of network resources to stream video to the cloud.

4.1 Applying Our Method to Enumerate Configurations

Camera, policy, and edge computing constraints culminate in a set of feasible deployment configurations. A configuration entails a set of of model type nodes in the graph (Figure 2) associated with a set of specific models and model deployment locations. Figure 3 illustrates the process of enumerating configurations using our method. The example in Figure 3 assumes that the camera is positioned so that individuals are always walking parallel to it (this is realistic if the camera is positioned facing a sidewalk, for example) and faces are not visible. Since faces are not visible, the modality of gait is used. Since people are always walking parallel to the camera, either GaitGraph or GaitNet can be used (recall that GaitGraph is angle-variant, meaning it can only function if people walk at the same angle relative to the camera each time).

Without taking into account edge computing or policy constraints, there are at least six possible configurations for gait recognition. (There are more, equally valid hypothetical configurations that could result from using alternative models for object detection and pose estimation; for the sake of brevity, these are not shown). Taking into account the computation constraints captured in Table 1, in which some models run too slowly on the edge for real-time analysis, only some of the configurations in Figure 3 are feasible. Specifically, Configuration 4 is not feasible (GaitNet would take too long on the edge). Now, if we assume that raw video and images are not permitted to be sent off the camera, Configurations 3, 5, and 6 are also not feasible. Thus, if we assume that the edge is constrained and raw video cannot be sent off the camera, only Configurations 1 and 2 are valid. Note that all configurations are feasible under the "no facial recognition" policy, since the camera constraints themselves prevents face recognition from being of use.

The example in Figure 3 demonstrates that, combinatorially, the number of possible configurations is large. Luckily, policy and computing constraints provide a way to narrow down the space of possibilities. With this framework for enumerating configurations, we can explore privacy risk.

4.2 Methodology and Datasets

We enumerate feasible configurations under various combinations of the camera, policy, and edge computing constraints we consider.

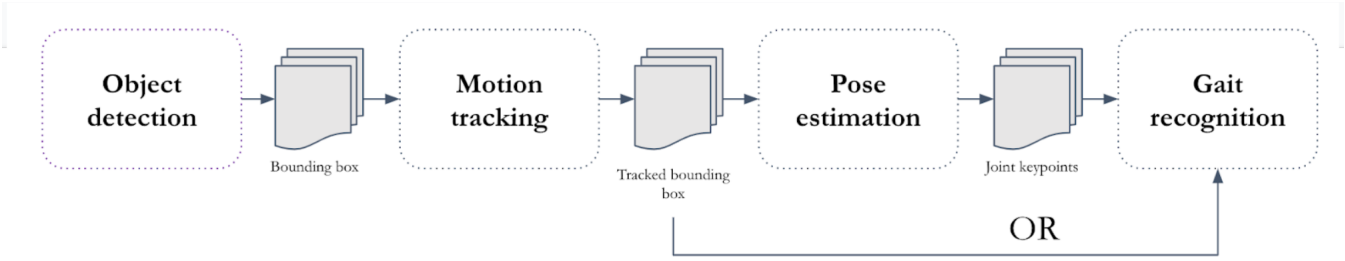
Measuring Privacy Risk For each configuration, we examine the privacy risk it poses by asking, "how well can verification be

Model	Task	Input	Compute Cost in Cloud	Compute Cost on Edge	Test Flow	Capable of Running on Edge?
YOLOX [15]	Object Detection	Raw video	42 FPS	8 FPS	Video->YOLOX	Yes
Faster RCNN [15]	Object Detection	Raw video	8 FPS	0.5 FPS	Video->Faster RCNN	No
Centroid Tracking (not deep) [2]	Motion tracking	Bounding box coordinates	Negligible	Negligible	Video->YOLOX->Centroid Tracking	Yes
BlazeFace [8]	Face detection	Image of person	23 FPS	5 FPS	Video->YOLOX->Centroid Tracking->BlazeFace	Yes
Topdown heatmapwith Resnet50 Backbone [21]	Pose estimation	Image of person	30 FPS	5 FPS	Video->YOLOX->Ccentroid Tracking->Topdown Heatmap with Resnet50	Yes
Topdown heatmap with Alexnet backbone [21]	Pose estimataion	Image of person	33 FPS	6.5 FPS	Video->YOLOX->Centroid tracking->Topdown Heatmap with Alexnet	Yes
Color recognition with Topdown heatmap Resnet50 backbone input	Color recognition	Coordinate of shoulder keypoint	30 FPS	5 FPS	Video->YOLOX->Centroid Tracking->Topddown Heatmap Resnet50 Backbone->Color recognition	Yes
Color recognition with Topdown heatmap Alexnet backboneinput	Color recogntion	Coordinate of shoulder keypoint	33 FPS	6.5 FPS	Video->YOLOX->Centroid Tracking->Topdown Heatmap Alexnet Backbone->Color recognition	Yes
GaitGraph [58]	Angle-variant gait recognition	Aggregated joint keypoints of person walking at same angle each time	Not tested =*	.20 seconds per subjectv	Aggregated joint keypoints->GaitGraph	Yes
GaitNet [53]	Angle-invariant gait recognition	Aggregated frames of person walking; may be walking at different angles each time	21 seconds per subject (CPU Only)	NOt tested*	Aggregated frames of person walking -> GaitNet	No
FaceNet [49]	Face recognition	Image of face	Not tested*	.015 seconds per face	Image of face->FaceNet	Yes

Table 1: Model Zoo. Note that Facenet and GaitGraph were not tested in the cloud since it is assumed their runtimes would be negligible there given their fast run times on the edge. Due to hardware compatibility issues, GaitNet was only able to be run in the cloud and only the CPU-only version was able to be tested. We can safely assume the GaitNet runtime on the edge would be significantly higher than that of the cloud. Also note the distinction between angle-variant and angle-invariant gait recognition. In angle-variant gait recognition, the individual must be walking at the same angle relative to the camera each time. In angle-invariant gait recognition, individuals can walk at any angle.

achieved?" We focus on verification because it provides a concrete and relevant point of departure for understanding the privacy risks of IoT video analytics. Specifically, verification is commonly achieved through visual means (i.e., video or photo). Verification can serve as the basis for tracking. It can also be viewed as a weaker formulation of recognition. That is, verification can become recognition if the biometric being used to verify is added to a recognition gallery (e.g., verifying that two pictures are both of John Smith is equivalent to facial recognition if the first picture is John Smith's passport picture and is present in a face recognition gallery).

Datasets To evaluate verification accuracy, we use real-life surveillance video as input to a configuration in our testbed. Using the



Configuration Number	Object Detection Model	Motion Tracking Model	Pose Estimation Model	Gait Recognition Model
1	YOLOX	Centroid Tracking	Topdown Heatmap + Resnet50	GaitGraph
2	YOLOX	Centroid Tracking	Topdown Heatmap + Resnet50	GaitGraph
3	YOLOX	Centroid Tracking	–	GaitNet
4	YOLOX	Centroid Tracking	–	GaitNet
5	YOLOX	Centroid Tracking	–	GaitNet
6	YOLOX	Centroid Tracking	Topdown Heatmap + Resnet50	GaitGraph

Figure 3: Configuration possibilities (before considering policy and computation constraints) for a scene in which faces are not visible and people are walking parallel to the camera. The graph at the top represents the flow of models that any configuration will have. All rows in the table are based off of this flow and represent (theoretically) possible placements of models (edge or cloud). Red cells correspond to models being run on the camera (edge), and blue cells correspond to models being run in the cloud.

dataset's ground truth identities, we can quantify how well that configuration achieves the verification task. Our videos were sourced from five datasets: OutdoorGait [53], 3DPeS [7], MARS [54], Chokepoint [**?**], and Sarasota. Samples frames are shown in Figure 4. Video from each dataset is subject to different camera constraints, which impacts face visibility, walking manner, and color recognition feasibility. We briefly detail each below.

- OutdoorGait - We use the test set of the OutdoorGait dataset. This dataset was designed to test and train gait recognition models with video from realistic surveillance settings. Subjects were filmed walking parallel to the camera from a closer distance than MARS or 3DPeS but a higher distance than Sarasota. The comparatively low distance of the camera from subjcts is offset by low resolution.
- 3DPeS - 3DPeS was designed for pedestrian re-identification in multi-camera systems. We use a sample of 3DPeS in which the same individuals are seen from one camera overlooking a courtyard multiple times within one day.
- MARS - MARS was designed for motion analysis and pedestrian re-identification. Video was captured from five cameras positioned in a busy courtyard. We manually inspected MARS video and selected video clips where participants were walking (as opposed to sitting or standing in place). This is by far the most difficult dataset to identify subjects in, as there are many occlusions, walking is completely unconstrained and faces are rarely visible.
- Sarasota - We created this dataset for our initial gait recognition experiments. People walk parallel to the camera in a controlled setting. We blurred subject faces for privacy requirements.
- Chokepoint - Chokepoint consists of four sequences of people walking through a door, captured from 4 cameras. Black and white images of cropped faces are provided. Gait and shirt color are thus not observable. We sample five images from each sequence-camera pair, leading to over 30 test images per camera.

The characteristics of each dataset are summarized in Table 2.

4.3 Results

We analyze feasible configurations under the edge computing constraints established above and our three considered policies: none, no face recognition , no raw video to cloud. Note that camera constraints are implicitly encoded in the input videos, as discussed above.

For each configuration, embeddings of each subject's gait and/or face are generated using the appropriate model(s) (i.e., GaitNet, GaitGraph, or FaceNet). Optionally, the color of the person's shirt is predicted using color recognition. Color cannot be used alone

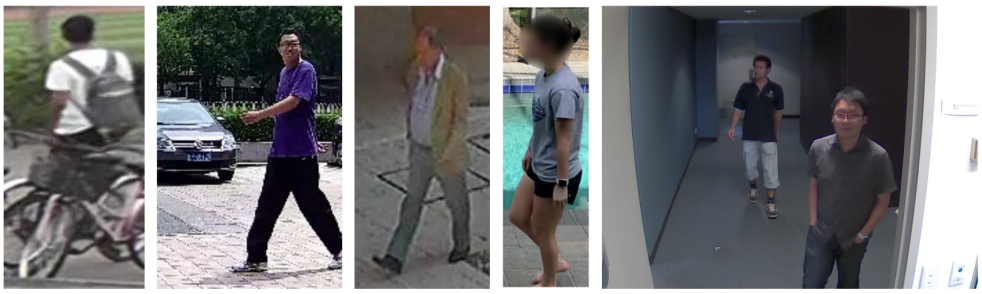


Figure 4: From left to right, sample frames from the MARS, OutdoorGait,3DPeS,Sarasota, and Chokepoint datasets.

Dataset	Number of subjects and videos/images	Face image quality	Gait Observable?	Walking Manner	Shirt color observable?	Samples collected on same day? (i.e., same shirt color?)
OutdoorGait Test	68 identities, across 223 videos	Medium	Yes	Parallel to camera	Yes	Yes
3DPeS	6 identities, across 13 videos	Low	Yes	Unconstrained	Yes	Yes
MARS	247 identities, across 736 videos	Low	Yes	Unconstrained	Yes	Yes
Sarasota	8 identities, across 24 videos	N/A (Blurred)	Yes	Parallel to camera	Yes	Yes
Chokepoint	25 identities, across 825 images	High	No	N/A	No	N/A

Table 2: Dataset Summary

to verify an individual, because the color of a person's clothes is not necessarily unique. However, it can be used in conjunction with other modalities to increase accuracy (e.g., by eliminating false positives). When combined, all modes of information form a "profile" of a subject that can be used to verify them across multiple sightings.

During our experiments, we generate a profile for each subject, and all possible pairs of profiles are compared to each other. If the distances of all modes of information in the two compared profiles are below their respective thresholds, we conclude that the profiles are of the same individual. For instance, if we wish to verify that two video samples collected within a short span of time (i.e., an hour) feature the same person, we might do so by verifying that the gait embeddings and colors of the shirts of the individuals in each video are sufficiently similar.

We analyze verification accuracy using two standard metrics: **(1) False rejection rate (FRR)**, which measures how often two sets of embeddings from the same individual are determined to be from different individuals, at a specific threshold; and **(2) False acceptance rate (FAR)**, which measures how often two sets of embeddings from different individuals are determined to be from the same individual, at a specific threshold.

Table 3 summarizes our results per dataset and combination of policy and edge computing constraint being considered. Camera constraints are implicitly encoded in the input videos, as discussed above. Note that, when listing configurations in Table 3, we only report the modality used (i.e., color, face, or gait) and the location that modality's recognition model was run at (edge or cloud). All models in the configurations except GaitNet were able to run on the edge.

4.4 Key Findings

Finding 1 *Edge-based verification methods (i.e., verification with GaitGraph or FaceNet) perform significantly worse than cloud-based methods (i.e., verification involving GaitNet).* The inability of edge-based verification methods to perform as well as cloud-based methods is best summarized by the large increases in FAR under Policy 3 (no video or raw image to the cloud), as shown in Table 3. In particular, GaitGraph and FaceNet perform poorly.

GaitGraph is an angle-variant gait recognition model capable of running on edge devices. Since it is angle-variant, we only test it on datasets in which people were walking at the same angle relative to the camera in every video. Only Sarasota and OutdoorGait meet this criteria. Figures 5 and 6 show that GaitGraph-based verification performs significantly worse than all other forms of verification on the Sarasota and OutdoorGait datasets.

Face recognition was tested with the ChokePoint and OutdoorGait datasets. Since the Chokepoint dataset consists of black-and-white images of faces captured from a surveillance camera, facial recognition is its only viable modality. Table 3 shows that, even on the relatively high quality face image in Chokepoint, FaceNet has

Dataset	Policy 1		Policy 2		Policy 3	
	FAR	Configuration	FAR	Configuration	FAR	Configuration
OutdoorGait Test	1.17 %	GaitNet (cloud) or Color (edge) + GaitNet (cloud)	1.17%	GaitNet (cloud) or Color (edge) + GaitNet (cloud)	75%	Color (edge) + FaceNet (edge))
3DPeS	23.3%	GaitNet (cloud) or Color (Edge) + GaitNet (Cloud)	23.3%	GaitNet (cloud) or Color (Edge) + GaitNet (Cloud)	N/A	N/A
MARS	54.4%	GaitNet (cloud)	54.4%	GaitNet (cloud)	N/A	N/A
Sarasota	0%	Color (edge) + GaitNet (cloud)	0%	Color (edge) + GaitNet (cloud)	47.3%	Color (edge) + GaitGraph (edge)
Chokepoint	49%	FaceNet (edge)	N/A	N/A	49%	FaceNet (edge)

Table 3: For each dataset, we report the configuration giving the lowest false accept rate (FAR) at false negative rate (FNR) = 5% under the given policy, assuming the edge is constrained. Note that Policy 3 prevents any configurations for the 3DPeS and MARS datasets from being possible.

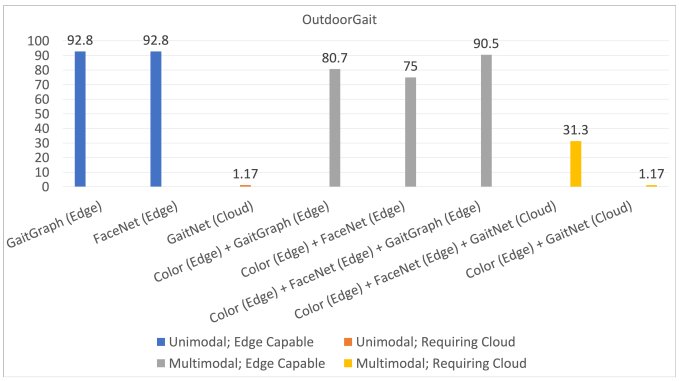


Figure 5: False accept rates (FARs) at false negative rate (FNR) = 5% with all computationally feasible possible configurations on the OutdoorGait Test dataset. Note that facial feature extraction succeeded on only 68 of the 223 videos in this dataset. We report the performance of FaceNet for just these 68 samples. Also note that policies may make some of these configurations infeasible. Edge-capable configurations have FARs close to 1, performing significantly worse than those requiring cloud computing. FaceNet, which performs poorly on this dataset as discussed in Finding 3, corrupts the performance of verification using Color, Gait, and Face.

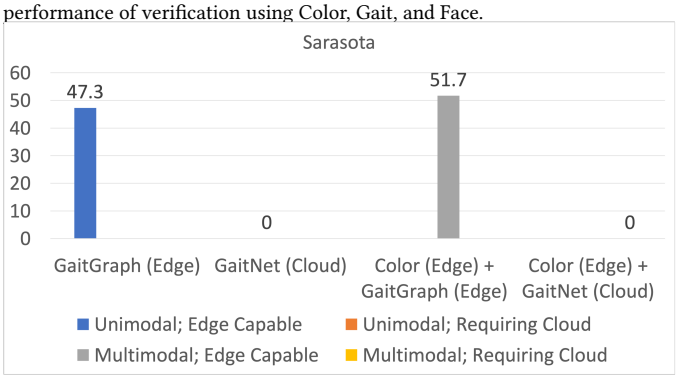


Figure 6: False accept rates (FARs) at false negative rate (FNR) = 5% with all computationally feasible possible configurations on the Sarasota dataset. Note that policies may make some of these configurations infeasible.

one of the highest FAR's at FRR = 5%. Furthermore, Figure 5 shows that FaceNet has almost zero precision on the OutdoorGait dataset.

Finding 2 *Gait recognition in the cloud (i.e., with GaitNet) is a significant privacy threat, potentially on par with facial recognition.* GaitNet is an angle-invariant gait recognition model that requires cloud computing in our testbed. We find that, across datasets, GaitNet has the highest verification accuracies.

GaitNet's accuracies were always comparable to (if not better than) the accuracy achieved by FaceNet on the Chokepoint dataset, in which high quality faces are visible. This indicates that gait recognition - not facial recognition - posed the largest privacy risk in our evaluation. Facial recognition makes up a large part of the public discourse on privacy risk, but our results show that gait recognition can be just as powerful. Perhaps we should elevate our concerns over gait recognition, developing policies that explicitly consider gait a sensitive biometric identifier.

Finding 3 *GaitNet performs significantly better in controlled settings and instances in which people are walking parallel to the camera.* In the OutdoorGait and Sarasota datasets, subjects walk parallel to the camera and the settings are controlled to prevent occlusions and irregular walking behaviors (e.g., holding bags or wearing baggy coats). On these datasets, GaitNet has almost perfect accuracy. However, in the MARS and 3DPeS datasets, in which individuals walk at any angle and often have coats or bags, performance significantly dropped (See Table 3).

Finding 4 *Video well-suited for gait recognition is often ill-suited for facial recognition.* OutdoorGait Test was the only dataset in which both gait and sufficiently high quality faces were observable. Yet face alignment and facial feature extraction only worked on 68 out of the 223 videos of OutdoorGait Test. This is in stark contrast with the almost perfect precision and recall achieved by GaitNet on this data. The disparity between gait recognition and facial recognition performances is because most gait recognition models achieve peak performance when subjects are walking parallel to the camera, which is the worst angle for facial recognition. Walking parallel to the camera allows walking manner to be observed most easily. However, this means that only the profile of a person's face is visible. Most facial recognition models are intended to take frontal images of a person's face as input. While we find that FaceNet can technically function on face profiles by applying an alignment step, this is not ideal. Figure 5 demonstrates that, on OutdoorGait Test data, FaceNet performs almost as poorly as a random guess

when used for verification, while GaitNet achieves almost perfect verification.

Finding 5 *In our evaluation, prohibiting video/raw image from being streamed to the cloud is more effective than prohibiting facial recognition.* In Findings 1, 2, and 4 we establish that GaitNet, which requires the cloud, is extremely powerful. However, GaitGraph and FaceNet, which are capable of running on the edge, perform significantly worse. Thus, in our evaluation, the largest threat is a cloud-based model. Furthermore, we see that gait recognition may be more well-suited to modern surveillance video than face recognition since it does not require that cameras be positioned close to individuals' faces, which may not always be possible. This is supported by the fact that only one of our datasets had high quality faces present. Policy 3, which prevents GaitNet from being used, is thus the most effective. Table 3 shows that Policy 3 significantly reduces the verification accuracies of permitted configurations.

As an aside, we note that GaitNet requires raw images as input. Contrastingly, GaitGraph and some other gait recognition models take as input joint keypoints aggregated across frames, which are generated through pose estimation. We show that pose estimation can be run on the edge. If GaitNet instead took as input joint keypoints, we would only have to send joint keypoints off the edge and GaitNet would be unaffected by Policy 3. This raises the question of whether aggregated joint keypoints should also be considered biometric identifiers.

Finding 6 *Color recognition can sometimes be used for preliminary processing on the edge to reduce cloud processing workload and increase verification accuracy.* Color recognition, which is capable of running on the edge, strengthened verification in the case of the Sarasota dataset, but contributed no improvement and even small reductions in accuracy in the cases of the other tested datasets.

These reductions were due to the inability of the color recognition model to accurately predict color under some lighting conditions. The color recognition model uses heuristics to predict the color of the person's shirt, and in cases when lighting in two samples were significantly different, it predicted that the color of the same shirt in the two pictures were quite different.However, reductions in accuracy due to this were minimal when they occurred.

Color recognition thus potentially provides a way for the IoT video analytics system to use edge processing to reduce workload on the cloud and in the network. Color recognition can be used to perform an initial verification step when lighting conditions are similar. Specifically, color recognition can be run on the edge so that the edge only sends collected gait video to the cloud for further processing if the predicted shirt colors are similar.

4.5 Limitations

One key limitation of our blackbox evaluation is that we only consider publicly available models and datasets. Many significant privacy threats come from privately-owned models or models trained and finetuned on privately-owned video datasets. We thus do not claim that the results of our blackbox evaluation fully represent privacy risk today. Instead, we hope to demonstrate how our method provides a starting point for achieving a realistic understanding of risk. After all, models and datasets constantly evolve, necessitating new analysis. Perhaps a method for facilitating this analysis is equally as valuable as a definite yet temporary understanding of risk.

5 CONCLUSION

We propose and conduct a blackbox evaluation of privacy risk in IoT video analytics systems. Guided by policy constraints, edge computing constraints, and camera constraints, we explore configurations that leverage face, gait, and shirt color recognition for verification. Our results highlight the growing risk of cloud-based gait recognition, which is not addressed by many existing privacy policies. They also demonstrate that our method for conducting a blackbox evaluation is a practical approach to understanding privacy risks.

REFERENCES

[1] 2004. Real-Time Video Analysis on an Embedded Smart Camera for Traffic Surveillance. In *Proceedings of the 10th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS '04)*. IEEE Computer Society, USA, 174.

[2] 2022, howpublished=https://adipandas.github.io/multi-object-tracker/includeme/apidocuments.html#attribution. Multi-object trackers in Python. (2022, howpublished=https://adipandas.github.io/multi-object-tracker/includeme/apidocuments.html#attribution).

[3] 2022 howpublished=https://arrayofthings.github.io/privacypolicy.html. Array of Things Governance Privacy Policies. (2022 howpublished=https://arrayofthings.github.io/privacypolicy.html).

[4] 2022, howpublished=https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57s. Illinois Biometric Information Privacy Act. (2022, howpublished=https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57s).

[5] ,howpublished=https://www.clearview.ai/. ClearviewAI. (,howpublished=https://www.clearview.ai/).

[6] Hamid Aghajan, Juan Carlos Augusto, Chen Wu, Paul McCullagh, and Julie-Ann Walkden. 2007. Distributed Vision-Based Accident Management for Assisted Living. In *Proceedings of the 5th International Conference on Smart Homes and Health Telematics (ICOST'07)*. Springer-Verlag, Berlin, Heidelberg, 196–205.

[7] D. Baltieri, R. Vezzani, and R. Cucchiara. 2011. 3DPes: 3D People Dataset for Surveillance and Forensics. *Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding* (2011).

[8] Valentin Bazarevsky, Yury Kartynnik, Andrey Vakunov, Karthik Raveendran, and Matthias Grundmann. 2019. BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs. (2019). https://doi.org/10.48550/ARXIV.1907.05047

[9] Paula Carrillo, Hari Kalva, and Spyros Magliveras. 2010. Compression Independent Reversible Encryption for Privacy in Video Surveillance. *EURASIP Journal on Information Security* 2009, 1 (2010), 429581. https://doi.org/10.1155/2009/429581

[10] F. M. Castro, M. J. Marín-Jiménez, N. Guil, and R. Muñoz-Salinas. 2016. Fisher Motion Descriptor for Multiview Gait Recognition. (2016). https://doi.org/10.48550/ARXIV.1601.06931

[11] Charlie Catlett, Pete Beckman, Nicola Ferrier, Howard Nusbaum, Michael E. Papka, Marc G. Berman, and Rajesh Sankaran. 2020. Measuring Cities with Software-Defined Sensors. *Journal of Social Computing* 1, 1 (2020), 14–27. https://doi.org/10.23919/JSC.2020.0003

[12] Andrea Cavallaro. 2004. Adding Privacy Constraints to Video-Based Applications. In *EWIMT*.

[13] Hanqing Chao, Kun Wang, Yiwei He, Junping Zhang, and Jianfeng Feng. 2021. GaitSet: Cross-view Gait Recognition through Utilizing Gait as a Deep Set. (2021). https://doi.org/10.48550/ARXIV.2102.03247

[14] Datong Chen, Yi Chang, Rong Yan, and Jie Yang. 2007. Tools for Protecting the Privacy of Specific Individuals in Video. *EURASIP J. Adv. Signal Process* 2007, 1 (jan 2007), 107. https://doi.org/10.1155/2007/75427

[15] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tianheng Cheng, Qijie Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. 2019. MMDetection: Open MMLab Detection Toolbox and Benchmark. *arXiv preprint arXiv:1906.07155* (2019).

[16] Liudmila V. Chernenkaya, Elena N. Desyatirikova, and Alexander V. Rechinskii. 2021. Realization of Computer Vision System for Biometric Identification of Personality. In *2021 International Russian Automation Conference (RusAutoCon)*. 409–414. https://doi.org/10.1109/RusAutoCon52004.2021.9537374

[17] Sen-ching S. Cheung, Jithendra K. Paruchuri, and Thinh P. Nguyen. 2008. Managing privacy data in pervasive camera networks. In *2008 15th IEEE International Conference on Image Processing*. 1676–1679. https://doi.org/10.1109/ICIP.2008.4712095

[18] Sen-ching S. Cheung, Jithendra K. Paruchuri, and Thinh P. Nguyen. 2008. Managing privacy data in pervasive camera networks. In *2008 15th IEEE International Conference on Image Processing*. 1676–1679. https://doi.org/10.1109/ICIP.2008.4712095

[19] Sen-Ching S. Cheung, Jian Zhao, and M. Vijay Venkatesh. 2006. Efficient Object-Based Video Inpainting. In *2006 International Conference on Image Processing*. 705–708. https://doi.org/10.1109/ICIP.2006.312432

[20] Cloudflare. 2022, howpublished=https://cities-today.com/industry/new-trends-in-smart-city-video-analytics/. What is data privacy? (2022, howpublished=https://cities-today.com/industry/new-trends-in-smart-city-video-analytics/).

[21] MMPose Contributors. 2020. OpenMMLab Pose Estimation Toolbox and Benchmark. https://github.com/open-mmlab/mmpose. (2020).

[22] Antitza Dantcheva, Petros Elia, and Arun Ross. 2016. What Else Does Your Biometric Data Reveal? A Survey on Soft Biometrics. *IEEE Transactions on Information Forensics and Security* 11, 3 (2016), 441–467. https://doi.org/10.1109/TIFS.2015.2480381

[23] Anupam Das, Martin Degeling, Xiaoyou Wang, Junjue Wang, Norman Sadeh, and Mahadev Satyanarayanan. 2017. Assisting Users in a World Full of Cameras: A Privacy-Aware Infrastructure for Computer Vision Applications. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 1387–1396. https://doi.org/10.1109/CVPRW.2017.181

[24] Simon Denman, Michael Halstead, Clinton Fookes, and Sridha Sridharan. 2015. Searching for People Using Semantic Soft Biometric Descriptions. 68, P2 (2015). https://doi.org/10.1016/j.patrec.2015.06.015

[25] Tamara Denning, Zakariya Dehlawi, and Tadayoshi Kohno. 2014. In Situ with Bystanders of Augmented Reality Glasses: Perspectives on Recording and Privacy-Mediating Technologies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 2377–2386. https://doi.org/10.1145/2556288.2557352

[26] Chao Fan, Yunjie Peng, Chunshui Cao, Xu Liu, Saihui Hou, Jiannan Chi, Yongzhen Huang, Qing Li, and Zhiqiang He. 2020. GaitPart: Temporal Part-Based Model for Gait Recognition. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 14213–14221. https://doi.org/10.1109/CVPR42600.2020.01423

[27] Renan Fialho, Rayele Moreira, Thalyta C. P. Santos, Samila S. Vasconcelos, Silmar Teixeira, Francisco Silva, Joel J. P. C. Rodrigues, and Ariel S. Teles. 2021. Can computer vision be used for anthropometry? A feasibility study of a smart mobile application. In *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*. 119–124. https://doi.org/10.1109/CBMS52027.2021.00058

[28] Hiren Galiyawala, Kenil Shah, Vandit Gajjar, and Mehul S. Raval. 2018. Person Retrieval in Surveillance Video using Height, Color and Gender. *CoRR* abs/1810.05080 (2018). arXiv:1810.05080 http://arxiv.org/abs/1810.05080

[29] Hiren Galiyawala, Kenil Shah, Vandit Gajjar, and Mehul S. Raval. 2018. Person Retrieval in Surveillance Video using Height, Color and Gender. (2018). https://doi.org/10.48550/ARXIV.1810.05080

[30] Cristian González García, Daniel Meana-Llorián, B. Cristina Pelayo G-Bustelo, Juan Manuel Cueva Lovelle, and Nestor Garcia-Fernandez. 2017. Midgar: Detection of people through computer vision in the Internet of Things scenarios to improve the security in Smart Cities, Smart Towns, and Smart Homes. *Future Generation Computer Systems* 76 (2017), 301–313. https://doi.org/10.1016/j.future.2016.12.033

[31] Thales Group. 2021, howpublished=https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/biometrics/biometric-data. Biometric data and privacy laws (GDPR, CCPA/CPRA). (2021, howpublished=https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/biometrics/biometric-data).

[32] Kashmir Hill. 2020, howpublished=https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html. The Secretive Company That Might End Privacy as We Know It. (2020, howpublished=https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html).

[33] IBM. 2022, howpublished=https://www.cloudflare.com/learning/privacy/what-is-data-privacy/#:~:text=Data%20privacy%20generally%20means%20the,online%20or%20real%2Dworld%20behavior. IBM SVS Analytics 4 0 Plan Update for NYPD. (2022, howpublished=https://www.cloudflare.com/learning/privacy/what-is-data-privacy/#:~:text=Data%20privacy%20generally%20means%20the,online%20or%20real%2Dworld%20behavior).

[34] Tomohiro Inoue, Megumi Chikano, Shuji Awai, and Takeshi Konno. 2021. Gait Recognition with 2D Pose Infomation Using a Surveillance Camera. In *2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*. 351–355. https://doi.org/10.1109/LifeTech52111.2021.9391846

[35] Fortune Business Insights. 2022, howpublished=https://www.fortunebusinessinsights.com/industry-reports/video-analytics-market-101114. Video Analytics Market. (2022, howpublished=https://www.fortunebusinessinsights.com/industry-reports/video-analytics-market-101114).

[36] Amnesty International. 2021, howpublished=https://www.amnesty.org/en/latest/news/2021/06/scale-new-york-police-facial-recognition-revealed/. Surveillance city: NYPD can use more than 15,000 cameras to track people using facial recognition in Manhattan, Bronx and Brooklyn. (2021, howpublished=https://www.amnesty.org/en/latest/news/2021/06/scale-new-york-police-facial-recognition-revealed/).

[37] intersoft Consluting. , howpublished=https://gdpr-info.eu/art-4-gdpr/. Art. 4 GDPR Definitions. (, howpublished=https://gdpr-info.eu/art-4-gdpr/).

[38] Elvira Ismagilova, Laurie Hughes, Nripendra P. Rana, and Yogesh K. Dwivedi. 2020. Security, Privacy and Risks Within Smart Cities: Literature Review and Development of a Smart City Interaction Framework. *Information Systems Frontiers* (2020). https://doi.org/10.1007/s10796-020-10044-1

[39] Anil K. Jain, Sarat C. Dass, and Karthik Nandakumar. 2004. Soft Biometric Traits for Personal Recognition Systems. In *Biometric Authentication*, David Zhang and Anil K. Jain (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 731–738.

[40] Suman Sekhar Jana, Arvind Narayanan, and Vitaly Shmatikov. 2013. A Scanner Darkly: Protecting User Privacy from Perceptual Applications. *2013 IEEE Symposium on Security and Privacy* (2013), 349–363.

[41] Dangwei Li, Xiaotang Chen, and Kaiqi Huang. 2015. Multi-attribute learning for pedestrian attribute recognition in surveillance scenarios. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*. 111–115. https://doi.org/10.1109/ACPR.2015.7486476

[42] Xiang Li, Yasushi Makihara, Chi Xu, and Yasushi Yagi. 2021. End-to-end Model-based Gait Recognition using Synchronized Multi-view Pose Constraint. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. 4089–4098. https://doi.org/10.1109/ICCVW54120.2021.00456

[43] Xiang Li, Yasushi Makihara, Chi Xu, Yasushi Yagi, and Mingwu Ren. 2019. Joint Intensity Transformer Network for Gait Recognition Robust Against Clothing and Carrying Status. *IEEE Transactions on Information Forensics and Security* 14, 12 (2019), 3102–3115. https://doi.org/10.1109/TIFS.2019.2912577

[44] David H. Nguyen, Aurora Bedford, Alexander Gerard Bretana, and Gillian R. Hayes. 2011. Situating the Concern for Information Privacy through an Empirical Study of Responses to Video Recording *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 3207–3216. https://doi.org/10.1145/1978942.1979419

[45] Odyuzaki. 2019, howpublished=https://medium.com/@odyuzaki/white-box-vs-blackbox-testing-cef090a3b955. White Box Vs Black-box Testing. (2019, howpublished=https://medium.com/@odyuzaki/white-box-vs-blackbox-testing-cef090a3b955).

[46] JoséRamón Padilla-López, Alexandros Andre Chaaraoui, Feng Gu, and Francisco Flórez-Revuelta. 2015. Visual privacy by context: proposal and evaluation of a level-based visualisation scheme. *Sensors (Basel, Switzerland)* 15, 6 (06 2015), 12959–12982. https://doi.org/10.3390/s150612959

[47] José Ramón Padilla-López, Alexandros Andre Chaaraoui, and Francisco Flórez-Revuelta. 2015. Visual privacy protection methods: A survey. *Expert Systems with Applications* 42, 9 (2015), 4177–4195. https://doi.org/10.1016/j.eswa.2015.01.041

[48] Associated Press. 2022, howpublished=https://apnews.com/article/bf75dd1c26c947b7826d270a16e2658a. Chinese 'gait recognition' tech IDs people by how they walk. (2022, howpublished=https://apnews.com/article/bf75dd1c26c947b7826d270a16e2658a).

[49] Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. FaceNet: A unified embedding for face recognition and clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. https://doi.org/10.1109/cvpr.2015.7298682

[50] Alireza Sepas-Moghaddam and Ali Etemad. 2021. Deep Gait Recognition: A Survey. (2021). https://doi.org/10.48550/ARXIV.2102.09546

[51] Jiayu Shu. 2017. A Survey on Visual Privacy in Ubiquitous Computing. (2017).

[52] Jiayu Shu, Rui Zheng, and Pan Hui. 2018. Cardea: context-aware visual privacy protection for photo taking and sharing. *Proceedings of the 9th ACM Multimedia Systems Conference* (2018).

[53] Chunfeng Song, Yongzhen Huang, Yan Huang, Ning Jia, and Liang Wang. 2019. GaitNet: An end-to-end network for gait based human identification. *Pattern Recognition* 96 (2019), 106988. https://doi.org/10.1016/j.patcog.2019.106988

[54] Springer 2016. *MARS: A Video Benchmark for Large-Scale Person Re-identification*. Springer.

[55] Patrick Sudowe, Hannah Spitzer, and Bastian Leibe. 2015. Person Attribute Recognition with a Jointly-Trained Holistic CNN Model. In *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*. 329–337. https://doi.org/10.1109/ICCVW.2015.51

[56] Suriyon Tansuriyavong and Shin-ichi Hanaki. 2001. Privacy Protection by Concealing Persons in Circumstantial Video Image. In *Proceedings of the 2001 Workshop on Perceptive User Interfaces (PUI '01)*. Association for Computing Machinery, New York, NY, USA, 1–4. https://doi.org/10.1145/971478.971519

[57] World Wide Technology. 2020, howpublished=https://www.wwt.com/article/can-physical-distancemonitoring-and-smart-cameras-help-businesses-reopen-faster. Can physical distance monitoring smart cameras help businesses reopen faster? (2020, howpublished=https://www.wwt.com/article/

can-physical-distancemonitoring-and-smart-cameras-help-businesses-reopen-faster).

[58] Torben Teepe, Ali Khan, Johannes Gilg, Fabian Herzog, Stefan Hormann, and Gerhard Rigoll. 2021. Gaitgraph: Graph Convolutional Network for Skeleton-Based Gait Recognition. In *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE. https://doi.org/10.1109/icip42928.2021.9506717

[59] Cities Today. 2022, howpublished=https://cities-today.com/industry/new-trends-in-smart-city-video-analytics/. New trends in smart city video analytics. (2022, howpublished=https://cities-today.com/industry/new-trends-in-smart-city-video-analytics/).

[60] Hsin-Hsiang Tseng and Wen-Hsiang Tsai. 2013. Protection of Privacy-Sensitive Contents in Surveillance Videos Using WebM Video Features.

[61] Jaychand Upadhyay, Rohan Paranjpe, Hiralal Purohit, and Rohan Joshi. 2020. Biometric Identification using Gait Analysis by Deep Learning. In *2020 IEEE Pune Section International Conference (PuneCon)*. 152–156. https://doi.org/10.1109/PuneCon50868.2020.9362402

[62] Junjue Wang, Brandon Amos, Anupam Das, Padmanabhan Pillai, Norman Sadeh, and Mahadev Satyanarayanan. 2018. Enabling Live Video Analytics with a Scalable and Privacy-Aware Framework. *ACM Trans. Multimedia Comput. Commun. Appl.* 14, 3s, Article 64 (jun 2018), 24 pages. https://doi.org/10.1145/3209659

[63] Xiaoyi Yu, Kenta Chinomi, Takashi Koshimizu, Naoko Nitta, Yoshimichi Ito, and Noboru Babaguchi. 2008. Privacy protecting visual processing for secure video surveillance. *2008 15th IEEE International Conference on Image Processing* (2008), 1672–1675.

[64] Shikun Zhang, Anupam Das, Lujo Bauer, Lorrie Faith Cranor, and Norman M. Sadeh. 2020. Understanding People's Privacy Attitudes Towards Video Analytics Technologies.

[65] Wei Zhang, Sen-Ching Samson Cheung, and Minghua Chen. 2005. Hiding privacy information in video surveillance system. *IEEE International Conference on Image Processing 2005* 3 (2005), II–868.

[66] Jianqing Zhu, Shengcai Liao, Dong Yi, Zhen Lei, and Stan Z. Li. 2015. Multi-label CNN based pedestrian attribute learning for soft biometrics. In *2015 International Conference on Biometrics (ICB)*. 535–540. https://doi.org/10.1109/ICB.2015.7139070

[67] Shilin Zhu, Chi Zhang, and Xinyu Zhang. 2017. Automating Visual Privacy Protection Using a Smart LED. *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking* (2017).