Technical Challenges in Location-Aware Video Surveillance Privacy

Jack Brassil

HP Laboratories
HPL-2008-213

Keyword(s):
GPS, CCTV, Camera, phone, privacy law

Abstract:

Though designing, deploying and operating a video surveillance system in a public place is a relatively simple engineering task, equipping operational systems with privacy enhancing technology presents extraordinarily difficult technical challenges. We explore using mobile communications and location tracking to enable individuals to assert a preference for privacy from video surveillance. Rather than prohibit or defeat surveillance, our system - Cloak - seeks to discourage surveillers from distributing video without the authorization of the surveilled. We review the system architecture and operation, and demonstrate how privacy can be enhanced while requiring no change to existing surveillance technology. We use analysis and simulation to explore the solution's feasibility, and show that an individual's video privacy can be protected even in the presence of the many sources of error (e.g., dense crowds, unsynchronized clocks, unreliable communications, location error, location signal loss) we anticipate in a deployed system. Finally, we discuss the key technical, social, and legal barriers to Cloak's large-scale deployment, and argue that the pervasive use of camera phones requires the focus of efforts on surveillance privacy technology to shift to limiting dissemination rather than limiting video capture.

External Posting Date: December 18, 2008 [Fulltext] Approved for External Publication
Internal Posting Date: December 18, 2008 [Fulltext]

To be published in Protecting Privacy in Video Surveillance, A Senior, Ed. Springer-Verlag, Computer Imaging, June 2009

Technical Challenges in Location-Aware Video Surveillance Privacy

Jack Brassil

HP Laboratories
Princeton, NJ 08540
jtb@hpl.hp.com

Abstract. Though designing, deploying and operating a video surveillance system in a public place is a relatively simple engineering task, equipping operational systems with privacy enhancing technology presents extraordinarily difficult technical challenges. We explore using mobile communications and location tracking to enable individuals to assert a preference for privacy from video surveillance. Rather than prohibit or defeat surveillance, our system – *Cloak* – seeks to discourage surveillers from distributing video without the authorization of the surveilled. We review the system architecture and operation, and demonstrate how privacy can be enhanced while requiring no change to existing surveillance technology. We use analysis and simulation to explore the solution's feasibility, and show that an individual's video privacy can be protected even in the presence of the many sources of error (e.g., dense crowds, unsynchronized clocks, unreliable communications, location error, location signal loss) we anticipate in a deployed system. Finally, we discuss the key technical, social, and legal barriers to *Cloak's* large-scale deployment, and argue that the pervasive use of camera phones requires the focus of efforts on surveillance privacy technology to shift to limiting dissemination rather than limiting video capture.

1 Introduction

We have become accustomed to the alarm bells sounding with the release of each new report specifying the growth in the number of fixed video surveillance cameras and associated closed circuit television systems operating in public spaces [1], [2]. Many of the public surveillance systems are operated by municipal agencies and are sanctioned to enhance routine operations including automobile and pedestrian traffic monitoring, or permit relatively inexpensive remote policing in an effort to deter crime. Operators of such systems frequently adhere to an acceptable use policy – often published – for their CCTV system, addressing matters such as the proper storage and dissemination of captured video content.

A large number of other video surveillance systems are privately owned and operated, though in many cases these systems capture images of people and objects in public spaces (e.g., building mounted perimeter security systems, sidewalk Automatic Teller Machines). The intended purpose of these systems is typically the protection of private property, and restrictions on the use of video captured by such systems tends to be at the discretion of the system's owner. Canada, however, has recently introduced new guidelines to assist private organizations in maintaining surveillance privacy [3].

Concern about the threat to privacy associated with dedicated surveillance systems seems warranted. Cases of misuse or unauthorized video dissemination are reported with regularity. Many systems are far from secure, often relying on low cost analog video technology and electronic transmission that is easily tapped. Worries about 'alternative use' of recorded content remain unabated, and public discourse on privacy issues seems slow to move forward [4].

But the threat to privacy of these conventional fixed position camera systems pales in comparison to the threat posed by the proliferation of camera phones; approximately 370 million camera phones were sold in 2005 [5]. Such phones are not only pervasive in both public and private spaces, but the combination of their mobility and operator (i.e., owner) control permits them uninhibited access to locations and targets unavailable to conventional surveillance cameras. Though people are often the targets of surveillance, we are equally concerned about photographing objects; an image of a sensitive document bearing confidential information may require greater protection than an image of its owner. Though in this chapter we will routinely refer to a person seeking privacy while moving through a public space, the technology we discuss applies to arbitrary mobile objects (e.g., automobile license tags, credit card transactions).

Though the primary purpose of a camera phone is not to serve as a privately owned and operated surveillance device, it is quite capable of that role. The sophistication of both cameras and video signal processing has grown enormously as computing power has increased exponentially over time. And disseminating content has never been easier or more accessible; internet content hosting sites such as *Flickr* and *YouTube* have grown enormously popular.

To address these concerns this chapter presents technology that permits a person (or object) to unambiguously assert a preference for video privacy [7]. A key insight which forms the basis of this solution is that it is crucial to distinguish between two activities – being surveilled, and subsequently having the captured video or images distributed. That is, we make the following distinction; while being surveilled may present a potential *threat* of privacy loss, the widescale public disclosure [8] or dissemination of tangible evidence of one's presence at a particular place and time is an *actual* loss of privacy.

We take the presence of ubiquitous 'dumb' video surveillance systems to be immutable. We believe that in many cases those concerned for personal privacy would be satisfied with the assurance that their image is simply not distributed without their permission. Our technology does *not* seek to prevent image acquisition, nor present any technical barriers to captured image distribution. Instead the system seeks to permit captured images and video segments to be reviewed and, if necessary, 'sanitized' prior to public release, if this is desired by any person or object in the video itself. In effect, we seek to extend to a very large scale the common courtesy of asking a party for permission to take photographs. We believe that in many – but not all – cases this approach will jointly satisfy the needs of both the surveillers and the surveilled.

The remainder of this chapter is organized as follows. Section 2 describes the operation of our proposed 'privacy-enhanced surveillance' system, and establishes the technical feasibility of the system for the case of stationary surveillance cameras. Section 3 introduces some basic detection schemes to identify privacy seekers in the presence

of crowds in surveillance footage, and presents simulation results that measure how reliably privacy seekers can be identified given the many potential sources of error (e.g., location uncertainty, clock error) that will arise in an actual deployment. We also discuss how our system can be extended to provide privacy from mobile camera phone surveillance. Section 4 examines some closely related technologies to support surveillance privacy, and the next section looks at the ever widening gap between technology and law. In the final section we close with a summary of our investigation.

2 System Architecture and Operation

Our system, called *Cloak*, places the burden of sanitization on video owners seeking public dissemination of content they own. It also places the burden of *asserting* the preference for privacy on the individual interested in retaining privacy. We believe that this division of labor is appropriate, based on the following assumptions:

- While many of those traversing public spaces might choose to cloak on an occasional basis, relatively few will choose to cloak on a frequent or continuous basis. Hence we propose a system where those seeking privacy must 'opt-in'.
- The vast majority of captured surveillance video has little obvious value and will not be disseminated. Hence, our system is designed to require no change to existing infrastructure, nor require any action unless content is to be disseminated.
- A video segment selected for dissemination that contains an image of a person seeking privacy and requiring sanitization will be an uncommon event; most individuals captured in a video will not be seeking privacy protection.
- Even when sanitization is necessary it will often not interfere with the specific content valued by the content owner. A privacy seeker in an image simply might not be central to the reason an owner seeks dissemination. In this case both the privacy objective of the surveilled and the dissemination objective of the surveiller can be jointly met.

Beginning with this set of assumptions certainly seems non-intuitive. For instance, why design a system for relatively light use? First, we suspect that while many people are interested in obtaining increased privacy, few will be willing to bear any associated cost or inconvenience, no matter how minimal, for an unfamiliar service that is difficult to value. As we will see, Cloak does not depend on any of these assumptions, but certain aspects of system design would likely change if alternative assumptions were made.

Our proposed privacy-enhanced surveillance system works as follows. A person entering a public space potentially subject to surveillance elects to be in a Privacy Assertion (PA) state by carrying and activating a Privacy-Enabling Device (PED). A PED is a mobile navigation and communications device containing

- a clock,
- a unique identifier,
- one or more location tracking devices, such as a Global Positioning System (GPS) receiver,
- a wireless, mobile data communication link, and

– a non-volatile data storage device.

The bandwidth required for the communication link (e.g., $<$ 1 kilobit/second) is a small fraction of that offered by current wireless, broadband, wide-area data services (e.g., EV-DO). A PED should be equipped with a small amount of storage (e.g., $<$ 32 MB flash memory) to preserve location information during wireless link dropout periods. Tight synchronization between the PED clock and surveillance video recording device clocks is not required. A PED should be inexpensive, as much of this functionality is present in most mobile phones in use today. While certain existing cellular telephone based location services are not sufficiently precise for this application (e.g., 50 meter accuracy in Enhanced 911 Phase II handset-based solutions), they can serve as useful redundant location systems when a primary GPS system loses signal acquisition.

Indeed, dedicated object location tracking devices and associated web-based infrastructure are already in use. One example is the Zoombak A-GPS Locator [11], which provides 'continuous' location updates to a central location tracking site at a specified interval between 5 minutes and 1 hour. Integrating both cellular and GPS location service permits the device to continue to acquire location information in environments where signal information is unavailable to one or the other location tracking technologies.

From a user's perspective entering a PA state is as easy as toggling a 'privacy switch' on one's PED. A PED in PA state periodically captures and timestamps its location coordinates (e.g., 1 location per second) and either transmits them immediately or stores them locally (for later transmission). A PED requires no manual intervention when active and communicating in real-time. Note, however, that real-time transmission might not be possible if the communications link connectivity is intermittent or the channel is unreliable. Non real-time operation might require the PED operator to initiate uploading of locally stored location information when reliable communications become available. Of course, timely delivery reduces the risk that surveillance video or images are disseminated prior to upload.

A location *clearinghouse* receives communications from each active PED. The clearinghouse indexes and stores received information by location. Note that maintaining the complete uncompressed trajectory of a single PED consumes $<$ 500 KB of memory per PA hour; cleverly compressed trajectories would consume considerably less. Though it might seem that recording paths would be storage-intensive, a typical user only requires roughly the same amount of storage as would be consumed by taking a single digital photo each day.

Suppose a surveiller (i.e., video content owner) wishes to disseminate a captured video. The surveiller queries the clearinghouse to determine if *any* PED was active in the field-of-view of its camera during the time interval the video was captured. For now let's limit our attention to fixed surveillance cameras rather than mobile camera phones; we will return to a discussion of how the system can be extended to mobile cameras later. For stationary-mount cameras, finding active PEDs in a camera's field-of-view requires that the surveiller know the camera's location, and also be able to calculate how each location in the camera's 3-dimension field-of-view corresponds to its position in the acquired 2-dimensional image; calculating such projections can be a one-time

event. We will discuss the effect of subsequent camera misalignment and other sources of error later in this chapter.

A surveiller queries the clearinghouse via a communication protocol. For example, a sequence of queries can be used to determine if *any* PED traversed the camera's field of view during the time interval of the video segment. An initial query may use only coarse-grain location information (e.g, "Were any PEDs within distance d of location (x, y, z) during time interval $[t_1, t_2]$?"). We will later argue that even under a scenario where the density of PEDs traversing surveilled spaces is high, the probability that a randomly chosen video segment requires sanitization prior to dissemination (i.e., a *hit*) is low.

In the event of a hit both the video segment from the surveiller and the associated PED paths or trajectories from the clearinghouse are forwarded to a video *sanitizer*. The sanitizer strives to modify the video segment to respect the privacy asserted by the photographed PED users. We ideally envision this video processing to be entirely automated, though the required level of sanitization and the sophistication of that process depends on the specifics of the case, and can vary from trivial to implement to extremely complex. We will discuss sanitization technologies again in Section 4. If the sanitizer can modify the video clip to jointly meet the privacy requirements of the photographed PED carrier and the video owner, the resulting clip is returned to the owner for dissemination. Figure 1 presents a sample of how a single image might appear after sanitation.

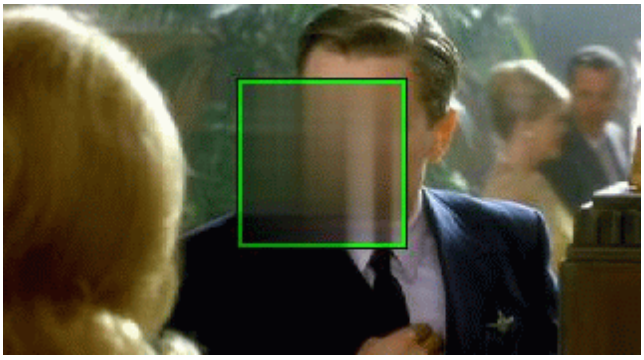

Fig. 1. The privacy objective of the surveilled and the image dissemination objective of the surveillers can often be jointly satisfied. In this image the identity of the man in the foreground is protected, though the image's commercial value might center on the second conversing couple (background, right).

Before we begin a more detailed look at the range of technical, social and legal issues associated with deploying Cloak, let's take a moment to establish what Cloak is *not*. Cloak does not seek to prevent or deter anyone from taking photographs or creating videos in public spaces. In particular, the system does not interfere with surveillance systems engaging in authorized remote policing. Cloak does no 'smart' video processing at video creation time; all video information is captured. Cloak places no burden on those not seeking surveillance privacy. Cloak requires no change to existing cameras or surveillance systems, save a requirement to time-stamp videos to facilitate off-line processing. Finally, Cloak does not alter the video ownership rights of a surveiller.

2.1 Discussion

Is building Cloak even technically feasible? To answer this question, let's begin with a very simple example. Suppose we have a single fixed camera primarily trained on a public space. We are given a sufficiently long duration video clip (e.g., 60 seconds) which shows 2 people traversing the space, and for simplicity let's say the 2 arrive and depart the video at different times. Next suppose that one is a PED carrier, and the other is the party of interest to the surveiller (i.e., the reason for dissemination), and we are given the location information reported by the PED.

Even assuming the many potential sources of error, including poor clock synchronization, considerable error in location reported by the PED, camera location error and subsequent misalignment, etc, the likelihood of being unable to determine which of two individuals is the privacy-seeker in the video is slight. It is often the case that a crude estimate of the arrival and departure times to and from the scene are enough to distinguish between two independent travelers moving through a scene. Upon identifying the privacy seeker, manually removing that party via video editing would often not interfere with the video owner's desire to preserve the presence of the other person. Indeed, examples of such editing are often presented on broadcast television, where surveillance videos of criminal activity are post-processed to remove the identities of other present parties (e.g., to protect potential witnesses).

While the above construction assures us that the *Cloak* system can be successfully implemented for very simple use scenarios, the next section focuses on far more complex and interesting scenarios. Some of the questions we seek to answer with our investigation are:

- Can we overcome the many potential sources of location tracking error to correctly identify a PED carrier?
- Can we pinpoint a PED carrier within a large, dense crowd?
- How much time must a PED carrier be present in a video to be reliably identified?

Before turning to these technically challenging questions, let's first take a closer look at several of the principal objections that are commonly raised by those skeptical of deploying the Cloak system.

- Voluntary Compliance
  Cloak's most dubious property is that the system relies on voluntary compliance by video owners to scrub their content prior to dissemination. Technology alone does

not appear to be capable of protecting privacy; only societal pressure or law can enforce compliance with the unambiguous desire of the privacy seeker. However, technology *can* be used to assert one's preference for video privacy, and to provide proof of that assertion should one's preference be ignored. A person whose privacy is asserted but ignored and violated can use *Cloak* to demonstrate this fact. Clearly, ignoring a person's privacy preference, particularly one going to the effort and expense of enabling technology to make the assertion, will earn little public good will. Public scorn can be an effective deterrent to violations by institutions owning content (e.g., municipalities) as well as broadcasters (e.g, television, newspaper, web hosting sites). Indeed, it is possible to imagine a sanitized video bearing a visible logo indicating its 'privacy protected' state; such an indication could conceivably *earn* good will for the video owner.

– Trusted Third Parties
  Privacy seekers must ultimately place trust in both the location clearinghouse and the sanitizer, even if some basic level of anonymity is afforded using one-time identifiers (e.g., the location clearinghouse need not know the identity of an individual whose location is being recorded). Surveillers receive no information on the surveilled except for their preference for privacy as ultimately indicated by their scrubbed image. Sanitation, of course, brings together images and trajectories, but no additional explicit identifying information is necessarily required.
– Location Privacy
  Individuals have a visceral objection to having their location tracked. The paradox inherent in *Cloak* is that individuals perceive that they must risk sacrificing location privacy to achieve video surveillance privacy.
  But do they? To begin with, if you are carrying a cell phone, how much location privacy do you have? By entering a public space a person implicitly accepts that he or she might lose anonymity at any moment and their presence will become known. But a PED need only be enabled when the possibility of surveillance is real, not at all times. Without it, if one is indeed surveilled, then not only is location revealed, but it is now potentially revealed to a much wider audience in a potent, tangible form (i.e., video).
– Misuse by Authorities
  In principle a law enforcement agency could mine the location clearinghouse to track a subscriber's movements. But the same is true of other location-sensing systems (e.g., Enhanced 911, Loopt) or automobile toll or credit card transactions, though perhaps to a lesser extent.

Cloak can be enhanced in a variety of ways, and some of these approaches can mitigate these concerns. For the sake of brevity we will omit an exhaustive discussion of the many possible system embellishments. To consider one example, however, a PED device can potentially specify a preferred Quality of Privacy (QoP). Examples of parameters that can determine service quality include retention time of location data, type of entity (e.g., person, automobile) to cloak, and the degree of desired video 'scrubbing.' Interestingly, the communication loop can be closed; QoP requirements can conceivably require the surveiller to obtain the explicit consent of the surveilled prior to video dissemination. Such an interaction might require the surveilled to examine and authorize

scrubbed content prior to dissemination, or even involve a negotiation of distribution rights.

3 Analysis and Simulation Results

Suppose there are P people (or objects) traversing the field-of-view of a single camera in some time interval $[t_0, t_0+D]$, and exactly one of those people carries an active PED. The PED reports its location each second, though both its estimate of time and location may be inaccurate; neither are synchronized with the surveillance system. Suppose a sanitizer is independently given the PED carrier's recorded, timestamped path and the recorded, timestamped surveillance video. Given a large value of P, how should the sanitizer correctly pinpoint the PED carrier in the video, and how likely is the sanitizer able to do so?

Let $\boldsymbol{s}(t)$ be the *actual* location of the PED at time t with $\boldsymbol{s} = \{x, y, z\}$. Let $\hat{\boldsymbol{s}}(t)$ be the PED's own estimate of its location. We will assume that the error in a PED's location estimate is bounded at each time instant, i.e.

$$|\boldsymbol{s}(t) - \hat{\boldsymbol{s}}(t)| < \epsilon. \tag{1}$$

For an inexpensive Garmin GPS 18 USB receiver [12], the advertised bound for location error is $\epsilon < 3$ meters (Wide Area Augmentation Service enabled) and $\epsilon < 15$ meters (Standard Positioning Service), for 95% of measurements. For simplicity let us first consider a *time synchronous* system; we assume that the PED's clock and surveillance system clocks are perfectly time-synchronized. We will revisit this assumption later.

We will also assume that image analysis enables us to discern the distinct people (or objects) in a scene, though in general this is not required for sanitization. Let $\boldsymbol{v_i}(t) : i \in P$ be the chosen 'center' of the i^{th} discernible object at time t in the captured video, where $\boldsymbol{v} = \{l, m\}$ is a point in the 2-dimensional image coordinate system. The sanitizer determines the path (i.e., trajectories) of each discernible object in the video segment, i.e., $\boldsymbol{v_i}(t) : i \in P, t \in [t_0, t_0 + D]$. The sanitizer also receives the *estimated* trajectory of the PED $\hat{\boldsymbol{s}}(t)$ as recorded while traversing the 3-dimensional space in the camera's field of view. Using knowledge of the camera location and perspective, the sanitizer projects that trajectory onto the 2-dimensional image coordinate space, i.e.,

$$\tilde{\boldsymbol{s}}(t) = T[\hat{\boldsymbol{s}}(t)]. \tag{2}$$

Hence, $\tilde{\boldsymbol{s}}(t)$ represents an approximate trajectory for one of the P objects in the video. Our objective is to determine to which object the approximate trajectory most closely corresponds.

To find that corresponding object (i.e., PED carrier) we compare the given approximate trajectory to each of the P object trajectories discerned from image analysis. To do so we choose to use one of the following two heuristics; either minimize the Mean Square Error (MSE) or (absolute) Linear Error (LE). That is, we find the object i such that

MSE:

$$\int_t^{t+D} |\boldsymbol{v_i}(t) - \tilde{\boldsymbol{s}}(t)|^2 dt < \int_t^{t+D} |\boldsymbol{v_j}(t) - \tilde{\boldsymbol{s}}(t)|^2 dt \tag{3}$$

or

LE:

$$\int_t^{t+D} |\boldsymbol{v_i}(t) - \tilde{\boldsymbol{s}}(t)|dt < \int_t^{t+D} |\boldsymbol{v_j}(t) - \tilde{\boldsymbol{s}}(t)|dt \tag{4}$$

for $0 < i \leq P,\ 0 < j \leq P, i \neq j$.

We next turn to simulation to determine the probability of correctly identifying a single active PED among P discernible objects given a noisy representation of its path. Since we have assumed that a PED updates its location periodically we model its trajectory in discrete rather than continuous time. Further, since our location-sensing device is assumed imperfect, we can approximate location (and the projection of location to the image coordinate system as defined by the transformation T in Eq. 2) as discrete-valued variables.

Selecting the most appropriate mobility model for simulating the movement of the P objects in a video greatly depends on where a camera is trained, and the type of objects in that scene. For example, vehicles photographed passing through a highway toll booth have relatively predictable motion, while pedestrians in a shopping mall might appear to move randomly. Further, correlations between the movement of different objects in a scene can vary greatly. Consider a camera trained on a busy public escalator where several people in close proximity have roughly the same trajectory – which person is carrying an active PED, the one just stepping off the escalator or the one immediately behind?

Recall that the P objects traverse the 3-dimensional volume in the camera's field-of-view, and each object's actual trajectory has an associated 2-dimensional trajectory in the image coordinate system. Since the transformation from a 3-d trajectory to a 2-d trajectory is deterministic, we can model object mobility in either coordinate system with no loss of generality; we choose to model mobility in the simpler 2-d image coordinate system. Note, of course, that a trajectory in image space does not in general have a unique corresponding trajectory through the volume. A preferred approach would be to work with 3-d mobility models (or better, to measure actual paths empirically) and transform those into the 2-d image space.

Object motion can now be represented by a walk on an infinite 2-d grid. To make matters concrete let's say that grid points correspond to a separation of 1 meter in the camera field-of-view, and the camera field has approximately a 10,000 square meter window (corresponding to grid points [0-99, 0-99]). Such a large, open space might correspond to the size of an outdoor public park or square.

Now let's introduce a simple mobility model. Suppose that at time $t_0 = 0$ each of the P objects enters the grid at location $v_i(0) \equiv v(0) = (0, 0)$; this corresponds roughly to the entire group entering the field-of-view simultaneously at the same location, as if entering though a single portal (i.e., hallway or door). Let each subsequent object step be independent from step-to-step, and independent of every other object step. Suppose that each object path follows a *biased random walk* with the following non-zero transition probabilities:

$$p[l+2|l] = 0.25 \quad p[m+2|m] = 0.25 \tag{5}$$
$$p[l+1|l] = 0.25 \quad p[m+1|m] = 0.25 \tag{6}$$
$$p[l|l] = 0.25 \quad p[m|m] = 0.25 \tag{7}$$
$$p[l-1|l] = 0.25 \quad p[m-1|m] = 0.25 \tag{8}$$

That is, in each step each object moves with equal probability either ahead or backward 1 meter, or ahead 2 meters, or does not move at all, in each dimension l (horizontal) and m (vertical).

Observed from directly above, an object traversing a surface under this model could appear at times to stop, to reverse course, or to vary speeds. The group of P objects would appear to be heading very roughly to the same destination, as if the group was entering a square room at one corner and moving toward the diagonally opposite corner. Though in principle an object could leave and return to the camera's field-of view (possibly multiple times) we will observe objects for a sufficiently long period that taking this event into consideration is unnecessary. Suppose we consider a time interval of duration $D = 200$ seconds (i.e., our observation interval is $t \in [0, 199]$); on average an object will drift 100 meters both horizontally and vertically during that interval, so a typical object will appear to move across the camera's entire field-of-view.

Now let's explore the effect of a PED's error in estimating its location on our ability to correctly identify the PED carrier. We introduce two location error models, each intended to capture all potential sources of location error. Each model is applied independently to each coordinate $\{l, m\}$ in the 2-dimensional image coordinate system at each step in the object's path. For the l dimension we have

Biased Uniform error:

$$p[|l - \tilde{l}| = d] = 1/W \quad d \in [a - \frac{W-1}{2}, a + \frac{W-1}{2}] \tag{9}$$

where W is an odd-valued integer greater than 1 meter, and a is a non-negative integer, or

'Worst Case' error:

$$\begin{aligned} p[|l - \tilde{l}| = -\frac{W-1}{2}] &= 0.25 \\ p[|l - \tilde{l}| = +\frac{W-1}{2}] &= 0.25 \\ p[|l - \tilde{l}| = 0] &= 0.5 \end{aligned} \tag{10}$$

An identical model is separately written for the m coordinate. Our uniform error distribution has mean a (i.e., $E[|l - \tilde{l}|] = a$), permitting us to examine cases where a PED's observed location has a fixed offset error (typically of a few meters). We chose to consider the 'Worst Case' error distribution because we anticipated that MSE detection would perform relatively poorly when the maximum location error is realized

in half the location updates, on average. Unless noted, in each of our simulations we considered $P = 1000$ objects traversing the grid, and we present error rates averaged over $10,000$ iterations of 200 second walks.

Let's now determine how well we can correctly identify the single PED carrier as the location error distribution width W increases in the absence of bias. Figure 2 shows that under MSE detection the error distribution width would have to exceed 41 meters before we guessed incorrectly in 10,000 scenes. Similarly, the figure shows no errors under LE detection with $W \leq 27$ meters, though note that the LE detection error rate climbs much more rapidly than the MSE detection error rate under uniformly distributed location error. Nonetheless both of these results compare very favorably with worst case GPS location error (Eq. 1). As we predicted the error rate for the MSE algorithm under the 'Worst Case' error distribution climbs much more rapidly with increasing W than under the uniform error distribution.

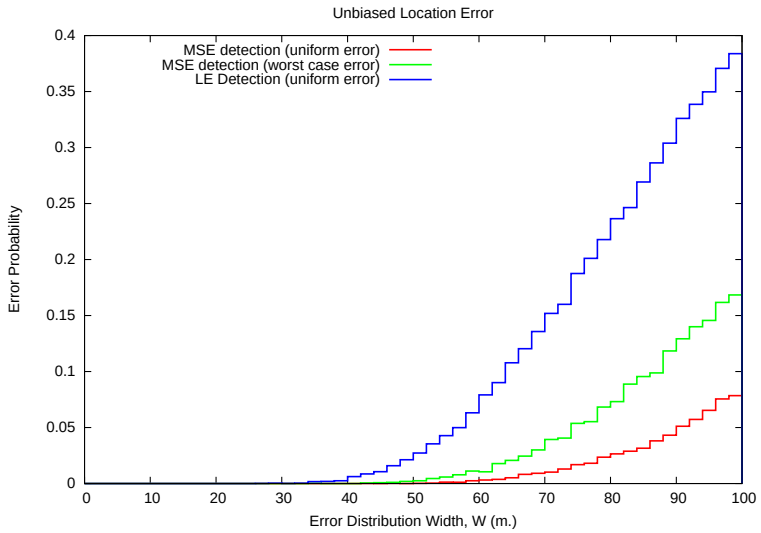

Fig. 2. Detection error probability vs. location error distribution width.

Figure 3 presents a set of sample paths for a typical iteration of our simulation. Each black dashed line depicts one of the P object trajectories. For clarity we show only a randomly selected 100 of the thousand object trajectories. The 200 green $+$ points indicated the erroneous sequence of locations reported by the PED carrier during the 200 second observation interval when moving under an unbiased, uniform error model with an error width of $W = 59$ meters. Given those locations, the MSE detection

scheme had an easy task of correctly identifying the path corresponding to the actual PED carrier (highlighted in red). The path highlighted in blue corresponds to the path among the 1000 that is furthest in distance from the reported trajectory. Remarkably, despite this noisy path report, the MSE algorithm is able to identify the PED carrier correctly in 99.75% of the 10000 iterations.

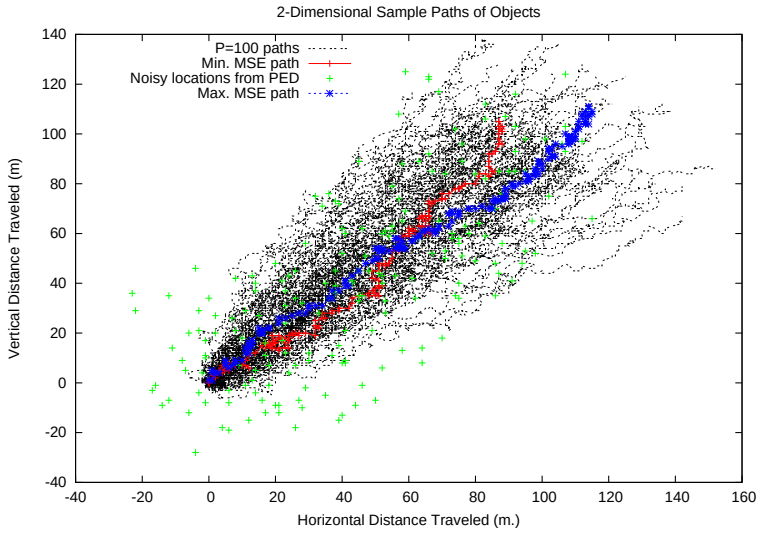

Fig. 3. Sample paths of 100 randomly select object trajectories among a group of 1000. The PED's actual trajectory (red) is selected correctly, despite the seeming randomness and wide range of the PED's reported positions (green $+$). The trajectory of the most distant object (in the mean square sense) is also shown (blue).

Next let's consider how error rates are affected by a fixed offset in location reported by the PED carrier. For a uniform distribution with a mean of a meters, we seek the value of the largest distribution width W that results in a detection error rate exceeding 1%. Table 1 shows that error rates can climb sharply as this fixed offset climbs; a fixed offset greatly reduces the location error variability that can be tolerated. In other words, if a PED carrier is always mis-reporting its location by a few meters, it is increasingly likely that the detection algorithm will select another trajectory as associated with the PED, even if the location error variability is shrinking.

Next we consider the effect of the absence of synchronization between a PED's clock and the surveillance system clock(s). Figure 4 shows how the detection error probability varies as we increase both the unbiased uniform distribution width W and

mean (a)	distribution width (W) LE	MSE
4	$W < 27$	$W < 41$
3	$27 \leq W \leq 35$	$41 \leq W \leq 57$
2	$37 \leq W \leq 41$	$61 \leq W \leq 67$
1	$43 \leq W < 45$	$69 \leq W < 73$
0	$45 \leq W$	$73 \leq W$

Table 1. The mean and width of a biased uniform location error distribution that result in an error rate higher than 1% for each detection.

the time difference between the surveillance system clock and the PED clock for a group of 100 objects traversing the grid.

Note how the the LE detection scheme degrades relatively slowly as asynchrony grows from 0 to 4 seconds, while the MSE scheme error rate climbs catastrophically after about 2 seconds. For very low timing error (≤ 2 seconds) detection performance is determined largely by the location error distribution width W. In a practical system it appears it would be necessary to ensure that clocks remained roughly in synchrony. In either case, these observations lead us to speculate that it might be beneficial to develop a multi-pass detection scheme. In such a system one would perform the first detection pass under the assumption of synchrony, then subsequent passes assuming growing asynchrony, searching for the absolute minimum error across all passes.

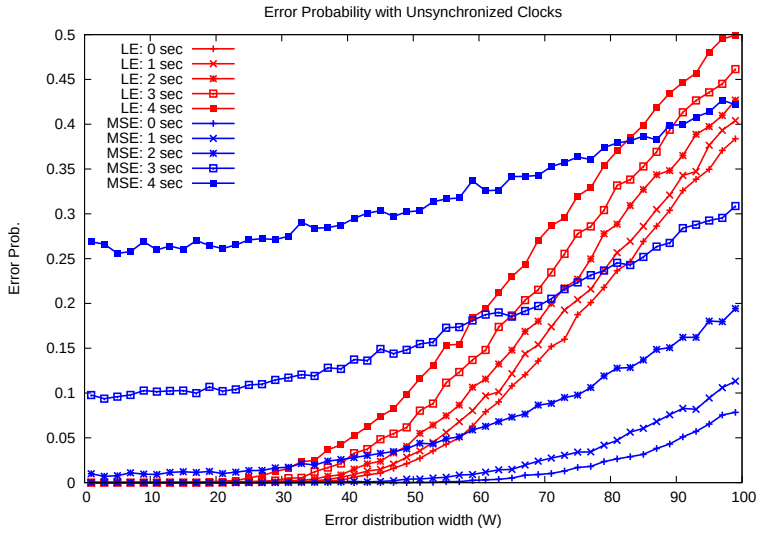

Fig. 4. Detection error rates increase with increasing timing error and location error.

What if we increase the number of objects P to choose from in a scene? Figure 5 shows that under the random walk model the detection performance decreases relatively slowly as we increase the object density in the presence of unbiased uniform location error. This too is unsurprising, for if we observe independent walks for a sufficiently long period k we expect to be able distinguish between many objects (i.e., $P \ll 2^k$).

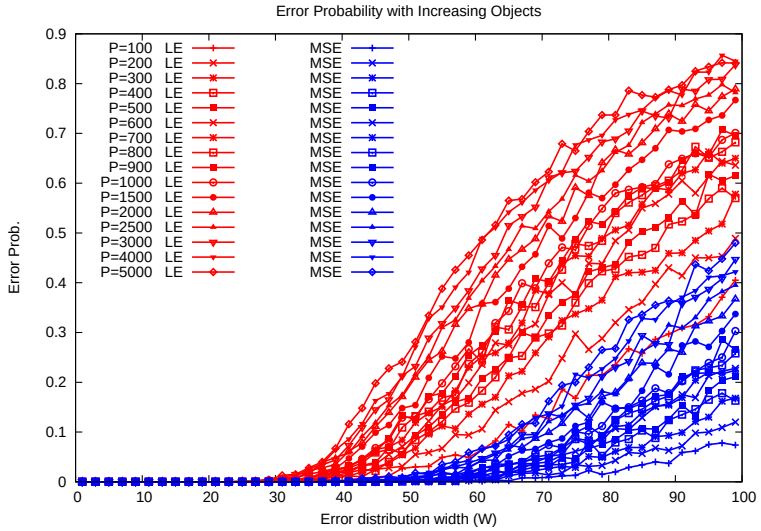

Fig. 5. Detection error rate increases relatively slowly with increased object density.

Dropouts, or the loss of location signal acquisition, is a familiar nuisance to GPS automobile navigation system owners. But dropouts can quickly become a significant source of detection error when dropouts coincide with surveillance of a PED carrier. We model this phenomenon as an on-off process where the probability of staying in the 'connected' state (i.e, location signal acquired) and 'disconnected' state are Bernoulli processes with probabilities α and β, respectively. Hence the duty cycle, or fraction of time in the connected state, is $D = \frac{\alpha}{\alpha+\beta}$. Let's make the rather severe assumption that during periods of disconnection from the location service the PED continues to report its last known location. Figure 6 shows how the detection error probability can increase dramatically with decreasing duty cycle. In this example, the mean time spent in the connected state is 8 seconds, and the mean in the disconnected state is $8 * (\frac{1-D}{D})$ seconds. There are, however, simple ways to make considerable improvements on this result. For example, one can imagine the PED delaying location reports until the location is re-acquired, and then sending a set of location updates estimating its previous positions (e.g., using linear interpolation between the last known location in the most recent connection state and the first location in the current connection state).

So far we have limited our attention to models of mobility in which each person's movement is independent of every other person's movement. But in many situations

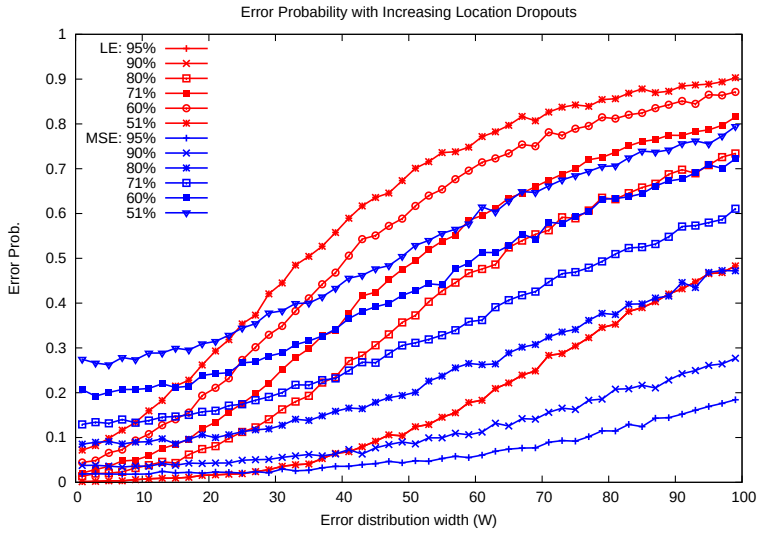

Fig. 6. Detection error rate increases as the location signal acquisition duty cycle decreases.

one's motion depends on the motion of others nearby; people traveling together move in tightly knit groups, and a chain of automobiles stops if one in front stops. Hence we developed a 'clustering' mobility model in which each person's movement depends in part on the movement of the PED carrier. The PED carrier effectively serves as the 'center' of a cluster of size P; those lagging behind the PED carrier increase their 'speed' and/or change their direction to catch up, and those ahead slow down. Figure 7 depicts the effects of both clustering and uniform location error as the error distribution width W varies. As expected the detection error rate increases when compared to independent motion. Nonetheless, despite the lower variability of distance between objects detection performance remains high.

Let's turn to the technical challenges faced by the sanitizer. Let's assume a sanitizer is given a video, a sanitization request, and access to the location database. The task faced is to jointly satisfy the privacy objective of the surveilled while preserving the image dissemination objective of the surveiller. Achieving such an outcome can be relatively trivial, or entirely infeasible.

As an example of the former case, a sanitizer might quickly determine that a video simply does not include images of the privacy seeker (i.e., a false alarm), or that the video can be readily 'cut' to remove such images. Basic video cuts are easily achieved with commonly available, open-source video processing tools. At the other extreme, the sanitizer might quickly determine that the privacy and dissemination objectives are at complete odds. This would ordinarily be the case for a video taken by a paparazzo of her targeted subject, if the subject was seeking privacy by carrying a PED.

Somewhere between these two extremes lies the challenge of video obfuscation. Here we typically wish the sanitizer to obscure only part of a video. Any number of

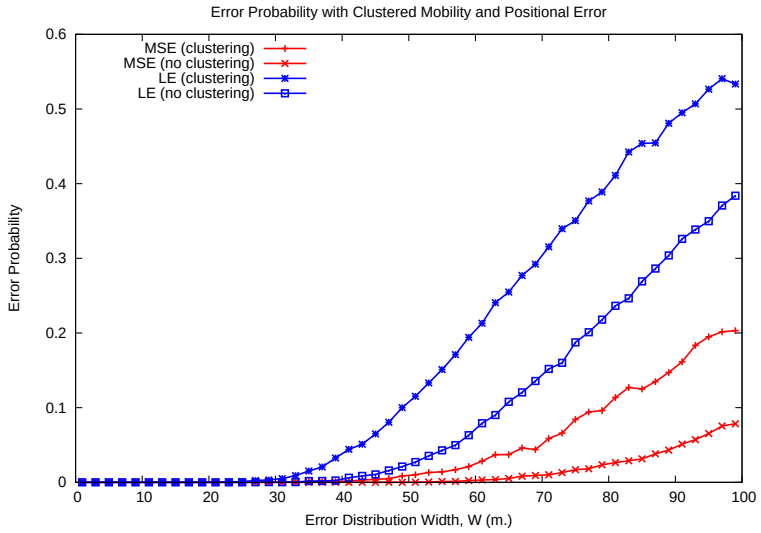

Fig. 7. The combined effect of clustered mobility and location error.

obfuscation techniques (e.g., blackout, pixellation) are easily available today in widely used software tools. Figure 1 provides a simple demonstration of a pixel blurring operation.

How does a sanitizer know when privacy and dissemination objectives are truly achieved? One approach might be based on an interactive process of scrubbed video review by the privacy seeker and video owner. In general, the more information shared between the parties the higher the likelihood of success. In some cases one imagines that the sanitizer could serve as a mediator between the parties.

To this point we have considered only fixed surveillance cameras, and we have not addressed the challenging special case of remotely controlled cameras that offer panning, tilting and zoom functions. But these cameras only begin to address the significant challenges we face in striving to support mobile surveillance cameras in the Cloak system. In the case of fixed surveillance cameras, we expected each camera to know its position, and to know the location of objects in its fixed field-of-view. While the former requirement can be reasonably applied to mobile cameras – at least to the same extent that location knowledge is expected of PEDs – it is difficult to imagine how the latter requirement could be realized. So instead, let us suppose that *every* PED within a radius r of an operating mobile camera at each time t is potentially in the camera's current field-of-view. That is, we will identify every PED carrier that could possibly have been photographed by the mobile camera at each instant during the video clip's duration.

To initiate this process, each surveillance video taken from a mobile camera to be disseminated is accompanied by the camera trajectory, indicating the camera's estimate of its location at 1 second intervals during the filming. Let $\hat{\boldsymbol{c}}(t)$ be the camera's own estimate of its location in the filming time interval $[t, t+k]$. The sanitizer queries the

location clearinghouse to determine if there exists a PED trajectory $\hat{\boldsymbol{s}}(t)$ satisfying

$$|\hat{\boldsymbol{s}}(t) - \hat{\boldsymbol{c}}(t)| < r \tag{11}$$

during the interval $[t, t + k]$.

In general this is a significant computational task. Suppose we define a region sufficiently large to enclose the mobile camera during the filming interval. If N PEDs are known to be present in that region for all or part of the interval, then a brute force algorithm to identify the subset of those PED carriers who came within distance r of the mobile camera is $O(Nk)$.

After completing such a task the location clearinghouse has now identified a set of PED carriers who *might* appear in the video. Of course, in many situations almost none of them will actually appear.

At this point, few courses of action remain open to a sanitizer to respect the privacy requested by the PED carriers. One approach is to initiate a negotiation with each PED carrier to privately review the video to confirm that she is indeed not a photographed party. In the event she is, a manual sanitization would be initiated with her assistance.

Of course, disclosure of the pre-sanitized video to a PED carrier who was not present in the video introduces a risk of subsequent disclosure (or even dissemination of the video itself) to a larger audience – exactly what the process is seeking to avoid. Here, however, we must ultimately rely on the benevolence of the community of PED carriers to respect each other's desire for privacy.

Before moving on let's take a moment to summarize what we have learned in this section. We have presented a collection of simulations to explore the feasibility of constructing *Cloak* using existing mobile communications, location technologies and video processing systems. We are keenly aware that each of these technologies is imperfect, and each can at times perform poorly in the challenging operational settings where we expect Cloak to be deployed.

Our first concern was maintaining a high PED detection rate given current technologies that imprecisely measure a carrier's location. But the results of Figure 2 suggest that we can use far more imprecise location measurement than is currently widely available with inexpensive GPS receiver technology. We also showed in Figure 6 that signal dropouts so common in GPS systems can be tolerated with modest reductions in detection rates, though applying simple location estimation techniques would increase performance during these dropout intervals.

Figure 3 highlights our considerable ability to correctly pick out a specific PED carrier even in a crowd of PED carriers moving along a roughly similar course, while Figure 5 shows that the detection rates remain high even as the size of the crowd grows to a large size. In both cases, of course, we are relying on the target carrier to remain in the surveillance camera's field-of-view for an extended time period, and for the target carrier to move through the scene largely independently of other carriers. To take matters a bit further, Figure 7 shows that a target PED carrier can also be readily detected when moving together in relatively tight clusters of other carriers.

An additional problem we anticipated that motivated our simulations is lowered detection rates caused by the lack of time synchronization between surveillance camera recording equipment and PED carriers. Figure 4 indicates that even a few seconds off

synchronization can degrade performance considerably, so this problem must be solved in an actual system. One approach is to enhance existing surveillance recording systems by providing them with the same clock source used by PED carriers. An alternate approach might involve using a different but relatively accurate time source, such as the *Network Time Protocol*. Both of these approaches suffer from the need to supplement existing surveillance recorders with new, more accurate time sources. In both cases to maintain the highest detection rates it might also be desirable for a sanitizer to use a more sophisticated, multi-pass detection algorithm which presumes imperfect time synchronization between recorder and PED carrier.

It is important to note that we have made a large number of assumptions in developing our simulations, and we have yet to address a number of easy-to-anticipate real-world considerations. As a simple example consider the variety of problems that we have yet to confront with even well-engineered, fixed-mount camera installations: scenes are typically viewed obliquely rather than from above, camera mounts might tilt and sag over time, scene lighting changes over time, objects occasionally block scene elements, etc.

In summary, our simulations give us confidence that a number of potentially significant barriers to *Cloak's* feasibility appear to be surmountable. While this is considerably short of an assurance of a practical system, it does encourage us to take the next step in our investigation – creation of a small-scale experimental prototype for laboratory and field evaluation.

4 Related Technologies

The automated analysis and processing of surveillance video is a central area of study for the computer vision and pattern recognition research community. IBM researchers have developed the *Privacy Camera* [16], an intelligent camera, or more precisely, a camera connected to a computer that performs real-time video analysis. The camera edits out certain pre-determined video content such as a person in a scene. Though both *Cloak* and the *Privacy Camera* effectively sanitize video, it is worth noting the differences between them. Our system does not edit video in real-time, yet it is designed to sanitize any video produced by any camera. Our video editing is driven by object trajectories; the objects themselves can be of arbitrary type and are not known a priori. We do not rely on information about a scene or objects in a scene. Indeed, in our work we have sought to deemphasize computationally intensive video processing, preferring to rely on positional information rather than information in the video itself. To our disappointment our research has uncovered relatively little work on mobility modeling for people or objects, and this appears to be a fertile area for future research [14], [15].

Intelligent mobile peer-to-peer systems are also an emerging technology that can either amplify or further threaten surveillance privacy. *Smart Camera Networks* such as *Facet* [27] create networks of cooperating mobile phone cameras. Though the intent of this system is to create composite images, they make use of location data of network participants, and may serve as a foundation to help Cloak address non-fixed surveillance systems. But these networks also bring the challenge of addressing the unanticipated problem that surveillance videos can be dynamically stitched together to

form composite surveillance video. This raises some very intriguing questions about ownership of derived and/or synthesized surveillance video which is well outside the scope of our study.

Technology useful in video sanitization is also advancing quickly. In particular, emerging *in-painting* [23] techniques promise to permit sanitized videos to appear unsanitized. This is achieved by not only removing objects but synthesizing perceptually plausible filler for the sanitized pixels.

Location-based services are also taking rapid hold, in part due to technological advances in GPS receivers and the application of wireless internet and cellular telephone technology to the development of 'social networks'. Personal location 'sharing' services such as *Loopt* go to some effort to distance their service offerings from 'tracking' services, as used to manage commercial vehicle fleets. Fortunately, as location services have evolved so have efforts by standards bodies to securely process networked location information to protect the privacy of participants; much of the work on protocols, privacy preference specifications, threat analysis and location data formats developed by the IETF *Geopriv* Working Group [24] would be directly applicable in the deployment of a location clearinghouse for Cloak.

Video privacy and surveillance is also a subject of considerable interest in the legal, political science, and social science research communities. Interested readers are strongly encouraged to examine the seminal privacy treatise by Warren and Brandeis [25], which serves as a foundation for current privacy law. This work anticipated the erosion of privacy due to advancing technology, correctly predicting the need for ongoing expansion in privacy law, and the article holds up astonishingly well more than a century after it was written. But despite their insight, the authors could not possibly have anticipated the degree to which technology has outpaced law, particularly in just the last decade. Fortunately, there are a growing number of research groups focusing on the intersection of technology, society and privacy. The *Surveillance Project* [22] at Queen's University is another resource available to those interested in studying various aspects of video privacy.

Cloak invites immediate analogies with the "National Do Not Call" registry [28] as implemented for telephone solicitation call blocking. The systems are similar in that users must opt-in to assert their privacy preference. Both systems are designed to discourage rather than impede an action by an unknown and frequently unwanted third party. Social and legal frameworks must be established to ensure compliance, which must be monitored and enforced. Of course, a surveilled party may be entirely unaware of surveillance and has no recourse to 'hang up' or 'not answer' or even review 'caller identification'.

An important distinction between Cloak and Do Not Call is that the former addresses privacy in public settings, while the latter was motivated in part due to what was viewed as an intrusion into the home. But mobile phones may be registered in Do Not Call. While this practice seems consistent with the mission of the registrar, the Federal Trade Commission, which is chartered to protect consumers (not homeowners), the registry can be arguably viewed as a precedent for a widely accepted [29] government-sponsored system to protect unwanted technological intrusions in both private *and* public places.

5 Technology and Privacy Law

The near absence of privacy rights for individuals in public spaces ensures that any primer on privacy law be brief. Unlike many countries – particularly in Europe – there is no explicit right to privacy guaranteed by the US constitution. Yet a vast incongruity exists between reality and perception; many citizens believe that legal protections to privacy exist that simply do not. Perhaps even more revealing is the lack of a high-level federal government privacy advocate. In Canada, for example, the mission of the Office of the Privacy Commissioner is to ”protect and promote the privacy rights of individuals.”

Video surveillance is closely tied with the concept of *territorial privacy*; that is, the fact that our assumption of privacy varies with place. Clearly, one expects a greater degree of privacy in one’s home than in more public settings. This distinction was upheld by US courts in the case of Katz vs. United States, 389 U.S. 347 (1967), which established that the Fourth Amendment provides privacy protections in places where a person has a reasonable expectation of privacy. At the same time, courts have supported public video surveillance by policing agencies as a legitimate means of protecting citizens. A person has no reasonable expectation of privacy when entering a crowded public space. Yet some people cherish the relative public anonymity that a dense urban setting can at times provide, suggesting that the notion of privacy in public might be desirable if not achievable.

Privacy rights also vary with the type of surveillance media. Audio communications are subject to Title 1 of the Electronic Communications Privacy Act (1996) [16 U.S.C. Section 2510] requiring warrants for audio ‘wiretapping’. Video surveilling is under no such constraint, and consequently existing public surveillance systems are video-only. We must await future court rulings to determine if objects containing sensitive information (e.g., credit card, driver’s license, medical report) are granted any photographic protections.

Perhaps the most shameful demonstration of one’s inability to assert one’s privacy in public is the street battle routinely waged between celebrities and paparazzi. In each encounter a photographer exploits the absence of their subject’s public privacy, and the subject at best feebly indicates their preference for no photography. In the absence of a clear and unambiguous ability to assert one’s preference and have it respected, the occasional result is a public skirmish. Both parties invariably believe that they are in the right. The subject believes in his or her ”right to be left alone”, while the photographer insists that any subject in public is ‘fair game’ and anything less is an assault on their livelihood.

The fundamental problem here, of course, is the inability (i.e., lack of protocol) for one to express one’s preference and have that preference respected. In the absence of such a protocol, certain lawmakers have begun to contemplate the flawed notion of a physical ‘privacy zone’ associated with individuals in public spaces [18].

Indeed, technology appears to be rapidly blurring what we traditionally think of as a well-defined line between public and private space. For example, use of a public sidewalk access Automatic Teller Machine (ATM) may arguably constitute a private transaction occurring in a public space, as may be a credit card exchange with a street vendor. Recent incidents have raised questions about the intrusiveness of Google’s Street View

video mapping service [19]; we are left to consider the possibility that the 'line' will be redrawn by technological advances, not by social consensus.

6 Conclusion

We have discussed a novel combination of mobile communications and navigation technology to permit individuals to assert a preference for privacy from video surveillance. We know of no other active research program that shares the ambitious agenda to universally supplement today's surveillance systems with privacy enhancing technology. Our simulations demonstrate that an individual's privacy can be protected even in the face of the many sources of error we expect to encounter in a deployed system. Though our investigation remains at an early stage – particularly in understanding its applicability to the rapidly growing number of mobile camera phones – our initial work suggests that there are no significant technical barriers to large-scale system deployment. However we have relied very heavily on simulation, and have made a large number of simplifying assumptions that we would not expect to hold in practical settings. Nonetheless, we are delighted by the promising early results, and the next step in our investigation will be to gather empirical evidence with a small-scale prototype.

Today's surveillance systems are rather primitive, technically unsophisticated tools. But with anticipated advances in technology we expect this to change. We believe that if individuals were aware of the extent that they are surveilled, and had the option to protect their images from distribution, that many would choose to do so.

References

1. New York Civil Liberties Union Report, "Who's Watching? Video Camera Surveillance in New York City and the Need for Public Oversight," http://www.nyclu.org/pdfs/surveillance_cams_report_121206.pdf, 2006.
2. Electronic Privacy Information Center, "Video Surveillance," http://www.epic.org/surveillance.
3. Government of Canada News Centre, "Privacy Commissioners Release New Video Surveillance Guidelines," http://news gc.ca/web/view/en/index.jsp?articleid=383707.
4. D.J. Solove "'I've Got Nothing to Hide' and Other Misunderstandings of Privacy," *San Diego Law Review*, Vol. 44, p. 745, 2007.
5. "InFoTrends/CAP Ventures Forecasts 2009 Camera Phones at 847 Million Units," http://www.digitalcamerainfo.com/content /InFoTrends-CAP-Ventures-Forecasts-2009-Camera-Phones-Sales-At-847-Million-Units.htm, 2005.
6. The Royal Academy of Engineering, "Dilemmas of Privacy and Surveillance: Challenges of Technological Change," London UK, 2007, http://www.raeng.org.uk/policy/reports/pdf/ dilemmas_of_privacy_and_surveillance_report.pdf.
7. J. Brassil, "Using Mobile Communications to Assert Privacy from Video Surveillance," *Proceedings of the First Workshop on Security in Systems and Networks (IPDPS 2005)*, Denver CO, April 2005.

8. D.J. Solove A Taxonomy of Privacy," *University of Pennsylvania Law Review*, Vol. 154, No. 3, p. 477, January 2006.
9. S.S. Hsu, "D.C. Forms Network of Surveillance," *Washington Post*, Feb. 17, 2002.
10. New York City Surveillance Camera Project, http://www.mediaeater.com/cameras.
11. Zoombak LLC, http://www.zoombak.com/.
12. Garmin GPS 18 USB Receiver Technical Specifications, http://www8.garmin.com/manuals/425_Technical Specification.pdf.
13. D. Donovan, "24-hour Camera Surveillance in City is Part of a Bigger Plan," *Baltimore Sun*, June 10, 2004.
14. A. Senior, "Tracking People with Probabilistic Appearance Models," *Workshop on Privacy Enhancing Technologies*, 2003.
15. I. Haritaoglu and M. Flickner, "Detection and Tracking of Shopping Groups in Stores," *Conference On Computer Vision and Pattern Recognition*, 2001.
16. Privacy Camera, IBM Research, http://www.research.ibm.com /peoplevision/videoprivacy.html.
17. D. Gillmor, "How do we adjust when cameras are everywhere?", *San Jose Mercury News*, June 20, 2004.
18. E. Ferkenhoff, "Lawmakers raise concerns over shadowing of Ill. candidate: Seek privacy zone to bar videotaping," *Boston Globe*, May 31, 2004.
19. M. Helft, "Google Zooms In Too Close for Some," *New York Times*, May 31, 2007.
20. J. Wozencraft and I. Jacobs, *Principles of Communications Engineering*, Wiley, New York, 1965.
21. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1965.
22. The Surveillance Project, Queens University, Canada, http://qsilver.queensu.ca/sociology/Surveillance/overview.htm.
23. Cheung, S-C., J. Zhao,. M.V. Venkatesh, "Efficient Object-based Video Inpainting," *IEEE International Conference on Image Processing (ICIP)*, 2006.
24. IETF Geopriv Working Group, http://www.ietf.org/. html.charters/geopriv-charter.html/.
25. Samuel Warren and Louis D. Brandeis, "The Right to Privacy," *Harvard Law Review* 193 (1890)
26. "Special Issue on Surveillance of Mobilities," *Surveillance & Society*, vol. 1 no. 4, Feb. 2004, http://www.surveillance-and-society.org/journalv1i4.htm.
27. P. Bolliger, M. Köhler, K. Römer, "Facet: Towards a Smart Camera Network of Mobile Phones," http://www.vs.inf.ethz.ch/pub/papers/ /bolligph-facet2007.pdf, 2007.
28. "The National Do Not Call Registry," http://www.donotcall.gov
29. The *Harris Poll*, "National Do-Not-Call Registry: Seven in Ten are Registered and All of Them Will Renew Their Registration," Poll #106, Oct. 31, 2007.