Who Clicks There!: Anonymising the Photographer in a Camera Saturated Society

Shishir Nagaraja
IIIT Delhi
nagaraja@iiitd.ac.in

Peter Schaffer
SnT, University of Luxembourg
peter.schaffer@uni.lu

Djamila Aouada
SnT, University of Luxembourg
djamila.aouada@uni.lu

ABSTRACT

In recent years, social media has played an increasingly important role in reporting world events. The publication of crowd-sourced photographs is one of the reasons behind the high impact. However, the use of a camera can draw the photographer into a situation of conflict. Examples include citizen journalists posting photographs of incidents of human rights violations on the Internet. The published images contain unambiguous clues about the location of the photographer such as the angle from which the scene was captured. Further, in the context of adversary operated CCTV systems, knowledge of the photographer’s potential location allows the adversary to identify the photographer by reviewing relevant footage. This is the camera location detection attack — a novel privacy threat against photographers seeking anonymity while posting images. In order to resist such powerful attacks, we introduce the notion of camera location anonymisation; by combining multiple input images captured from different viewpoints we produce a single image that appears to have been shot from a randomly chosen angle. To this end, we examine the use of view synthesis algorithms from computer vision literature as concrete defences. We show (both analytically and experimentally) that such defences could be a promising step in the direction of providing probabilistic anonymity guarantees. We analyse the extent of anonymity such techniques can provide in various scenarios and describe the challenges posed by scene geometry.

Categories and Subject Descriptors


General Terms

Security

Keywords

Privacy, Anonymity, Images, Photographer, View synthesis

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WPES’11, October 17, 2011, Chicago, Illinois, USA.
Copyright 2011 ACM 978-1-4503-1002-4/11/10 ...$10.00.

1. INTRODUCTION

Cameras are becoming ubiquitous. In the near future almost everyone will carry a personal high quality camera included in a mobile computing device. Furthermore, surveillance cameras are also increasing in spaces frequented by the general public. In the last few years, the quality of such cameras has significantly increased along with an increasing number of deployments on the streets.

Little attention has been paid to a more important threat in the context of camera ubiquity: the anonymity of the photographer capturing an image. Indeed, photographers widely expect to remain anonymous when publishing images via anonymous channels. However, this is not the case because images can leak information that can lead to the de-anonymisation of the photographer. In the following paragraphs we highlight a novel threat to privacy when adversary operated cameras point at photographers.

During the Burmese emergency of 2007 [13], protestors capturing photographs at anti-government rallies were identified and hunted down by crews of government secret police collecting footage of amateur journalists in attendance using mobile video cameras in a bid to prevent ‘sensitive’ footage from being shared on the Internet. Hidden cameras would not help the photographers because the published images would still reveal the location from which they were taken. When combined with images from adversary operated (CCTV) cameras, the leaked information can substantially reduce the anonymity of the photographer — we term this process as the camera-location detection attack. Indeed the adversary can use the published images to reduce the effort of monitoring the scene using his own cameras.

It is worth noting that the camera-location detection attack can still be a threat even in the absence of adversarial security cameras. We are aware of NGOs who distribute cameras to civilian populations to improve the accountability of armed groups in regions involved in conflict. Unfortunately, much of the digital imagery is not used since the users self-censor the output. Often because the cameras are installed at fixed positions and publishing the images would reveal the camera location. Or worse still, the user shot the footage from home during a disturbance and would risk having his/her property attacked, should the image be published.

These incidents inspired the authors to come up with the idea of combining input images from multiple scene viewpoints (by one or more photographers) to synthesise an image that appears to have been captured from a location where none of the photographers’ cameras were positioned. It turns out that the idea of combining multiple images to synthesise a new image from a virtual camera position is not entirely a new idea; there is existing work on synthesising views from novel viewpoints. Our contribution is to take the first steps in measuring the extent of input image information leaked by the output of current view synthesis algorithms. How-
ever ours is the first piece of work that revisits view synthesis from a security perspective\(^1\).

We foresee the usage of adversarial cameras along with camera location detection as a serious threat to citizen journalists among other categories of photographers. Our goal is to develop methods and techniques to minimise the leakage of camera location information from published images and the anonymity of photographers capturing images whilst maintaining maximal image quality. Essentially, we aim to develop a framework for analysing view synthesis algorithms and analyse existing techniques from computer vision literature. We propose a privacy metric for photographer anonymity, and measure the extent to which photographer anonymity is preserved by existing view synthesis algorithms both analytically and by experiments.

2. PROBLEM OVERVIEW

In this paper, we discuss solutions to the camera-location detection attack, i.e., given a scene and a camera, we wish to anonymise the location from which the photograph was captured. With the assumption that the camera and the photographer are at the same location, this translates to the problem of photographer-location anonymity.

Camera location anonymity problem.

Given a set of people \( P \) within a scene where \( k \leq |P| \) people are recording (not necessarily from the same location) and publishing images to the Internet. We aim to maximise the unlinkability between a published image (output) and the location from which it was captured (input). Anonymity is achieved by reducing the correlation between raw input and published images. This is similar to the function of a mix in anonymous communications, where the correlation between input and output traffic streams is minimised.

Attack.

Given a published image, an attacker can infer on the set of potential locations from which an image could have been captured by examining the scene viewpoint. This is because a given camera at a scene will record the image of the scene from a certain viewpoint. In this way, the recorded image will reveal some information about the location of the camera. An adversary can mount a camera location detection attack by combining information from a leaked (or published) image and footage from adversary cameras operating within the scene.

The image may also reveal information specific to the camera itself such as aberrations in the lens or the CCD which can be used to specifically identify the camera used in capturing a scene \([2, 8, 7]\). Such side channel information can subsequently be used to identify the photographer. However, we do not seek to defend against such attacks in this work.

Threat model.

Our threat model is that of a global passive adversary who deploys surveillance probes (video cameras). However, we assume that the adversary does not have the resources to analyse camera footage in real-time but rather that all video analysis is post-event.

Approach.

Our approach towards maximising camera location is the following. Instead of publishing an image downloaded from the camera’s memory card, the photographer records multiple images (or frames) of the scene of interest from different viewpoints. The photographer then generates a new image from a randomly chosen viewpoint using the input images and publishes the result. Since the scene viewpoint could potentially point anywhere within the physical space of interest within limits, the photographer’s anonymity may be maximised to the entire set of people present in that space. A view synthesis algorithm takes two images and an input viewpoint and generates a synthesised image from that viewpoint using the input images. Our paper is the first piece of work that identifies the camera-location detection attack as a novel privacy threat to camera users, proposes view synthesis as a solution approach, and analyses the current state-of-the-art from the perspective of anonymity.

3. BACKGROUND

There is a large body of literature related to view synthesis techniques in the area of computer vision and computer graphics. These techniques generate images at new viewpoints from different input views. View synthesis depends on a stereo correspondence algorithm applied on the input images followed by forward or inverse warping steps and a hole-handling step \([9]\). Each of these steps corresponds to a sub-algorithm. For a detailed review of view synthesis the reader is referred to the work of Scharstein and Szeliski \([12]\). In this section, we provide a brief introduction to view synthesis.

Early view synthesis techniques proceeded by using a 3D reference model provided as an input. This would then be used along with input images to generate a synthetic image from a chosen viewpoint by projecting the camera rays on the scene and then rendering the image. The synthetic images generated were unrealistic since scene points absent from the reference model geometry would appear as black.

Scharstein was the first to revisit the problem of stereo correspondence from the context of view synthesis. Given two input images, Scharstein \([10]\) proposed a technique of view morphing to reconstruct any viewpoint on the plane connecting the left, right and synthetic camera coordinates. His technique addresses the problems (of unrealistic output) mentioned in the previous paragraph. However, he assumed knowledge of camera parameters such as aperture size and shutter speed.

Fusiléo and Colombari \([5]\) proposed a technique that can synthesise new views without knowledge of camera parameters, i.e., from uncalibrated images. Avidan and Shashua \([1]\) presented another such technique with the improvement that they only require a small number of images. Given three images represented as a \(3 \times 3 \times 3\) matrix (trilinear tensor), they point out that if we know a pair of corresponding points in any two images then the corresponding point in the third image can be generated without resorting to projection within a known 3D model, to generate novel views.

All the view synthesis techniques we have discussed so far primarily attempt to identify the depth of each pixel in the scene. By simultaneously solving for both colour and depth, better results can be achieved. The implementation used in the experiments in this paper are based on one such algorithm. The Woodford algorithm \([14]\), in addition, also incorporates improvements in correspondence algorithms made using techniques from graph theory (graph cut methods). However this improved algorithm still faces non-trivial complexities such as visual artefacts. Therefore artefact removal is an important requirement for realistic images. In a recent work, Devernay et al. \([3]\) proposed techniques to detect and remove artefacts from synthetic images. Such techniques are also important to design acceptability-tests that can serve as a quantitative metric on the acceptability of synthetic images in real-world situations.

\(^1\)An earlier version of our work was published as a technical preprint at \url{http://arxiv.org/abs/1106.2696}
Finally, Zitnick et al. [15] applied concepts from still-camera based view synthesis to develop video techniques that allow a viewer to dynamically modify the viewing point of a video using a small number of video cameras.

4. PRIVACY ANALYSIS

Let \( I_L \) and \( I_R \) be two images of the same scene but taken from two different viewpoints, respectively left and right, and corresponding to the two camera positions \( L = (x_L, y_L, z_L) \) and \( R = (x_R, y_R, z_R) \). The objective of a view synthesis algorithm is to compute a synthesised image \( I_S \) at a new viewpoint or camera position \( S = (x_S, y_S, z_S) \) (see Figure 1). This new image aims at protecting the anonymity of the photographers. In what follows, we investigate to which extent this anonymity can be preserved. Concretely, given a synthesised image \( I_S \), how much can be inferred about the photographers’ positions \( L \) and \( R \)?

To answer this question, we start by identifying the extent of information leaked in a classical stereo-based view synthesis algorithm. Such an algorithm starts by establishing a stereo correspondence between \( I_L \) and \( I_R \). This means establishing a mapping of every pixel on the left image to its counterpart pixel in the right image. In computer vision literature, this is known as a disparity map. A disparity map \( D \) is a mapping\(^3\) of every pixel \( p_L^x = (u_L, v_L) \) on \( I_L \) to its corresponding pixel \( p_R^x = (u_R, v_R) \) on \( I_R \), so that,

\[
(u_R, v_R) \mapsto (u_L + D(u_L, v_L), v_L)
\]

(1)

For a realistic rendering of the synthesised image \( I_S \), a dense disparity map that maps all pairs of points \((p_L, p_R)\) is required. To that end, it is necessary to find the most optimal match between \( I_L \) and \( I_R \). If this correspondence is perfectly achieved, the only possible errors can only be due to geometrical constraints such as occlusions – physical obstructions in the left (\( L \)) and right (\( R \)) camera viewpoints which prevent capture of information visible from the synthetic viewpoint. These errors are contained in \( D \), and often translated on \( I_S \) as holes. The location of these holes may strongly infer on the relative geometry of the scene as seen from \( L \) and \( R \); which implies inferring on \( L \) and \( R \) if the scene geometry is known or can be discovered.

Once \( D \) is estimated, the synthetic image \( I_S \) can be computed as the sum of scaled versions of the matched images \( I_R \) and \( I_L \), defined at a given pixel location \((u, v)\) as:

\[
K_R(u, v) = \left(\frac{x_R - x_S}{b}\right) \cdot I_R(u + D(u, v), v)
\]

and

\[
K_L(u, v) = \left(\frac{x_S - x_L}{b}\right) \cdot I_L(u, v).
\]

The scales are the normalised positions of the synthetic camera relative to the photographers’ positions. The normalisation is with respect to the baseline \( b = (x_R - x_L) \). The synthetic image \( I_S \) is therefore written as:

\[
I_S(u, v) = K_L(u, v) + K_R(u, v),
\]

(2)

Depending on whether \( x_S \) (\( x \)-coordinate of the synthetic camera location) is chosen to be inside or outside the interval \([x_L, x_R]\), we distinguish the two cases known as interpolation and extrapolation, respectively. In what follows, we investigate the nature and the extent of the occlusions that cause privacy leakage for these two cases.

For simplicity, we assume that the cameras are operated at the same height, in other words, disparity changes along a single axis (\( z \)) only.

4.1 Privacy Leakage on Extrapolated Records

This section is based upon the observation that independently of the object recorded, there may be occlusion beside the object on the extrapolated record. The reason for this occlusion is that, in specific setups, neither of the two original cameras can record what is behind the object as the object itself occludes that part of the scene. Now, if one is trying to establish an extrapolated record based on such two original records, then the final record will contain the projection of the occluded scene part. The projection of the occluded scene part is called a hole. The existence of holes is independent of the effectiveness of the stereo correspondence subroutine of the view synthesis algorithm. Even in the case of perfect matching (pixel correspondence across input images), the occluded scene parts cannot be mitigated from the synthesised record. Of course, view synthesis algorithms incorporate a hole filling step, but this is usually achieved by replacing the hole pixels by an interpolation of the pixel values in the neighbourhood of the hole, and as such, filled holes can be observed and measured by an attacker (see, for example, Figure 1. in [5]). In the following, we assume that the attacker is given an extrapolated picture containing the holes. The attacker is trying to recover the position of the original cameras using information given by the position and size of the holes.

In fact, there are two slightly different setups when holes appear on the synthesised image. These setups are characterised by the

- Position of the left camera \( L \), position of the right camera \( R \) and the distance between them \( b \).
- Position \((x_S, y_S, z_S)\) of the synthesised viewpoint \( S \), and its distance \( s \) from \( R \).
- Object size \( l \) and its position, particularly its distance \( Z_o \) from the focal plane and size \( \ell \) of its projection on the image plane.
- Distance \( Z_o \) of the background from the cameras, and the focal plane.
- Position of the planes (i.e., focal plane, image plane), particularly the focal length \( f \).
- Hole size \( h \) on the image plane.

Figure 1.a and Figure 1.b depict the two slightly different setups when holes appear. In the former setup, a part of the occluded background from \( L \) and \( R \) is within the synthetic viewpoint. While in the latter setup the whole of the occluded background is within the synthetic viewpoint. In both cases the focal plane, the image plane and the background are parallel, and the cameras’ fields of view are assumed to be 180°. The object to record (depicted by the thick black line in the middle) is also parallel to the planes. The grey regions in Figure 1.a and Figure 1.b show the part of the scene that is not seen by either of the cameras \( L \) and \( R \). In other words, these are the occluded regions. As long as there is no object to record in these regions, no information will be lost because of the occlusion. However, as the background also has an occluded part, a hole will appear as the projection of these occluded parts (which are highlighted in red on the background planes) on the image plane. These holes are denoted by \( h'' \) and \( h'\) respectively, highlighted in red on the image planes (see Figure 1.a and Figure 1.b).

Note that, in this example we only consider the case when \( S \) is to the right from \( R \). We do not lose on generality with this assumption, as \( S \) being to the left from \( L \) would result in the same derivation. Note that there is also a third possible setup, namely
Figure 1: The two different setups that result in holes on the synthesised picture

when the left ray of camera $L$ and the right ray of camera $R$ are intersecting each other before the background. In this case, however, the background has no occluded part and therefore no hole will appear on the image plane. We will return to this case later on, after having investigated the cases when holes appear.

$h'$ and $h''$ can be calculated with the help of coordinate geometry. When we take coordinates of the points defining the projection lines of the cameras into consideration, $h'$ and $h''$ take the following form (note that $h', h'' \geq 0$ when holes appear):

$$h' = f \left( \frac{s}{Z_o} - \frac{s}{Z_b} \right) = s \hat{l} \left( 1 - \frac{Z_o}{Z_b} \right), \quad (3)$$

$$h'' = f \left( \frac{l}{Z_o} - \frac{b}{Z_o} + \frac{b}{Z_b} \right) = l - b \hat{l} \left( 1 - \frac{Z_o}{Z_b} \right) \quad (4)$$

Based on $h'$ and $h''$, we can already formulate the general equation
for $h$, namely,

$$h = \begin{cases} \min(h', h'') & \text{if } h'' > 0 \\ 0 & \text{if } h'' \leq 0 \end{cases} \quad (5)$$

By introducing

$$\alpha = \frac{\hat{l}}{l} \left(1 - \frac{Z_b}{Z_o}\right), \quad (6)$$

$h$ can be rewritten as

$$h = \begin{cases} \min(\alpha s, \hat{l} - \alpha b) & \text{if } h'' > 0 \\ 0 & \text{if } h'' \leq 0 \end{cases} \quad (7)$$

After this derivation, the question is what effect does the above have on the anonymity of the left and right camera locations. In other words, we are interested in the extent to which $h$ and $b$ can be inferred on from Equation (7). The knowledge of $s$ and $b$ together would reveal the camera locations completely because, obviously, the virtual coordinates of the synthesised record’s focal point are known (from the synthesised image itself) and $s$ and $b$ would give the locations of $L$ and $R$. Naturally, this would mean that the anonymity of the camera locations is zero. If the attacker is uncertain about $s$ or $b$, then the anonymity is higher. If both $s$ and $b$ are unknown, then the anonymity is maximal. In the following, we will quantify the anonymity of the photographers more precisely based on information leaked by Equation (7). For the further analysis, we assume that $h$ and $l$ are known. The magnification of both $h$ and $l$ are measurable on the synthesised picture, therefore the knowledge of their original value depends only on the knowledge of the CCD/CMOS size of the original cameras (as $h$ and $l$ are measured on the image plane, which is the CCD/CMOS in this case). There are only a few different CCD/CMOS sizes; the typical CCD/CMOS size for professional cameras is $36 \times 24$ mm. We note, however, we do not know whether the value of $h$ is the representation of $h'$ or $h''$. We further assume that $l$ and $f$ can be guessed by the observer. $l$ is the length of the object which, in most of the cases, has well-known dimensions. For example, if a speaker is recorded then $l$ is about 45-55 cm (measured at the shoulders), depending on the gender. The value of $f$ (i.e., the focal length) is guessable by knowing that specific scenes require specific $f$ values. Still considering the example with the speaker, the optimal focal length for capturing him/her would be between 85 and 100 mm, as this is the focal length interval most suitable for (full-length) portraits. Finally, we also assume that $Z_b$ and $Z_o$ are guessable as well. The guessability of the latter two parameters is an implication of the previous assumptions, namely of the guessability of $l$, $f$, and $s$. We note, however, that $Z_o$ can only be guessed if the background is not textureless.

Now, from Equation (7) we know that

$$i) \quad h = \alpha s \quad \text{or} \quad ii) \quad h = \hat{l} - \alpha b, \quad (8)$$

but we do not know which of the two cases holds. If the first case prevails then

$$s = \frac{h}{\alpha}, \quad (9)$$

$$b < \frac{\hat{l} - h}{\alpha} \quad (10)$$

which means that the observer knows the position of $R$ precisely and has an upper bound for $b$ (i.e., for the distance between $L$ and $R$). Otherwise, if the second case prevails then

$$s > \frac{h}{\alpha}, \quad (11)$$

$$b = \frac{\hat{l} - h}{\alpha} \quad (12)$$

meaning that we have a lower bound on the distance of $R$ from $S$ and we know $b$ precisely.

In the following, we assume that $n$ photographers are positioned with their cameras along a section of the focal plane. We refer to the two ends of this section as $P$ and $Q$, left to right, with coordinates $(x_p, z_p)$ and $(x_Q, z_Q)$, respectively. Note that the worst case from the observer’s point of view who is aiming at de-anonymisation is when $s$ does not restrict the anonymity set of the photographers, i.e., when the knowledge of $s$ does not exclude any of the suspected photographers. This happens if, in the first case, $x_S - s \geq x_P + b$, and if, in the second case, $x_S - x_Q > s$. In the further analysis we will assume this worst case, i.e., the results at the end will be conservative from the observer’s point of view.

We refer to the two ends of this section (denoted by $P$ and $Q$, left to right, with coordinates $(x_p, z_p)$ and $(x_Q, z_Q)$, respectively) as walls, in accordance with a possible scenario of a press conference. By not assuming a single mandatory position for the cameras relative to the photographers body (i.e., the camera is not necessarily located at the centerline of the torso), there are $\left(\frac{4}{2}\right)$ possible pairs of journalists who are suspicious for being the original recorders in the first case. This is because, in the first case, $R$ is known and $L$ is within distance $b$ to $R$. If we consider that the journalists are standing shoulder-to-shoulder then width $l$, the above statement becomes clear. In the second case, there are $(n - \left(\frac{4}{2}\right)) + (n - \left(\frac{4}{2}\right))$ possible pairs of suspicious journalists.

We can now quantify the anonymity of the photographers following [4] as

$$A = \frac{H(X)}{H_{\text{max}}}, \quad (13)$$

where $H$ stands for entropy. In our case, i.e., when all the pairs within the anonymity set are equally suspicious, the latter equation can be simplified as

$$A = -\frac{\sum_{i=1}^{N_{\text{sus}}} \frac{1}{N_{\text{sus}}} \log \frac{1}{N_{\text{sus}}}}{\sum_{i=1}^{N_{\text{tot}}} \frac{1}{N_{\text{tot}}} \log \frac{1}{N_{\text{tot}}}} = \log \frac{1}{N_{\text{sus}}}, \quad (14)$$

where $N = \frac{n(n-1)}{2}$ is the number of possible pairs of photographers.

When calculating anonymity, the two cases in Equation (8) have to be considered together but without their intersection calculated twice. Therefore, having $n$ photographers results in

$$N_{\text{sus}}^{(h>0)} = \left[\frac{\hat{b}}{\frac{4}{3}}\right] + \left(n - \left[\frac{\hat{b}}{\frac{4}{3}}\right]\right) + \left(n - \left[\frac{\hat{b}}{\frac{4}{3}}\right]\right) - 1 = 2n - \left[\frac{\hat{b}}{\frac{4}{3}}\right], \quad (15)$$

suspicious pairs (in other words, the size of the anonymity set is $N_{\text{sus}}^{(h>0)}$) in case when $h > 0$, where $b$ can be calculated as

$$b = \frac{\hat{l} - h}{\alpha} \quad (16)$$

As the components of these equations are known or guessable, the anonymity of the photographers can be evaluated using Equation (15) in case $b > 0$.

In fact, even the non-existence of holes reveals private information. In order to mitigate holes on the synthesised picture, $L$ and $R$ must reside on one of the specific constellations resulting in no
occlusion; and the number of such constellations is geometrically limited. Such constellations are characterised by the fact that the left ray of camera \( L \) and the right ray of camera \( R \) have to intersect each other before the background plane in order to avoid meaningful occlusions. In such cases \( h' \) can be negative.

In the latter case, i.e., when \( h = 0 \), then either \( h' \leq 0 \) or \( h'' \leq 0 \). Since \( h' \) is always positive, the former implies that

\[
b \geq \frac{l}{\alpha},
\]

(17)

which is the straightforward opposite of the previous cases when \( h'' \) was positive. Relying on the previous calculations, the number of suspicious pairs in this case is as follows:

\[
N_{\text{susp}}^{(h=0)} = \sum_{k=0}^{n-\left\lfloor \frac{b}{l} \right\rfloor} \left( n - \left\lfloor \frac{b}{l} \right\rfloor - k \right).
\]

(18)

The summation in Equation (18) is justified by the observation that the \( h = 0 \) case is similar to the above second case when \( b \) was known and \( s \) did not convey information in the worst case to the observer. Equation (17) does not reveal any hint on \( s \), and it can be rewritten as a sum of specific \( b \) values, expressed in a discrete way with the help of \( k \) when formulating anonymity. Finally, by simplifying Equation (18), one gets the following expression for the number of suspicious pairs when \( h = 0 \):

\[
N_{\text{susp}}^{(h=0)} = \left( n - \left\lfloor \frac{b}{l} \right\rfloor + 2 \right) \left( n - \left\lceil \frac{b}{l} \right\rceil + 1 \right) = N - N_{\text{susp}}^{(h>0)},
\]

(19)

where \( b \) can be calculated with the help of Equation (16) considering that \( h = 0 \). Anonymity can be further evaluated using Equation (16).

To give an example, let us assume a press conference scenario with a speaker and \( n \) journalists. The speaker is facing the journalists who are aligned beside each other, two of them secretly recording the speech with hidden cameras. Later these two will collaboratively establish a synthesised, extrapolated record with or without holes beside the speaker depending on the geometrical setup. In this scenario, one can calculate the anonymity of the photographers using Eqs. (14), (15) and (19). Some typical values of parameters required for the evaluation could be as follows:

- \( n = 20 \) (number of journalists)
- \( l = 0.5 \text{ m} \) (length of object, in this case shoulder width)
- \( \hat{l} = 5 \text{ mm} \) (length of projection of the object on the CCD/CMOS)
- \( Z_o = 5 \text{ m} \) (distance between the focal plane (where the journalists are standing) and the speaker)
- \( Z_b = 7 \text{ m} \) (distance between the focal plane and the background)
- \( h = 1 \text{ mm} \) (size of the hole beside the speaker measured on the CCD/CMOS if there is any, otherwise \( h = 0 \))

With these values, the anonymity of the photographers is

\[
A = \begin{cases} 
0.688 & \text{if } h > 0 \\ 
0.958 & \text{if } h = 0 
\end{cases}
\]

(20)

This tells us that the existence of holes reveals a large amount of position information (the anonymity of the photographers is reduced from 1 to 0.688), but even the non-existence of holes is transpiring some private information. Hence we conclude that the level of anonymity provided by the synthetic image can be improved much more.

### 4.2 Privacy Leakage on Interpolated Records

In the same way as in the case of extrapolated view synthesis, we assume that the observer is given an interpolated picture \( I_s \). The difference this time is that \( x_S \in [x_L, x_R] \). The observer tries to recover the position of the original cameras \( L \) and \( R \) from the information contained in the holes on \( I_s \). We consider the simplified setup depicted in Figure 2, with the same assumptions on the cameras as in Section 4.1. We model the object to record with the profile depicted by the thick black line, and defined by the function \( z = F(x) \), where the \( x \) axis is chosen for simplicity to overlap with the background, and the origin \( o = (0, 0) \) of the \( x \) and \( z \) axes approximately coincides with the center of the support of \( F \), i.e., \( F : x \in [-X_F, X_F] \rightarrow \mathbb{R} \).

The occlusions causing errors on the disparity map \( D \), or in the worst case causing holes, are shown on Figure 2 as \( H_L \) and \( H_R \). \( H_L \) corresponds to the occluded region on the object relative to the left camera, while \( H_R \) is the one relative to the right camera. To find \( H_L \), for instance\(^3\), we need to find the internal tangent line on \( F \) that intersects with \( L \). We assume that this line intersects \( F \) at a single point \( P_L = (x_{P_L}, z_{P_L}) \). Thus, we may define the line \((LP_L)\) as:

\[
(LP_L) : z = F'(x_{P_L})(x - x_{P_L}) + F(x_{P_L}) \quad \text{and} \quad F \cap (LP_L) = \{P_L\}
\]

(21)

We use polar coordinates \((l_T, r_T)\) to define the point \( P_T \) such that:

\[
x_{P_T} = r_T \cdot \cos(l_T) \quad \text{and} \quad z_{P_T} = r_T \cdot \sin(l_T)
\]

(22)

We approximate the expression of \( F \) around \( P_T \) by a circle centered at the origin \( o \), and with a known radius \( r_T = r \). Finding \( H_L \) becomes equivalent to finding \( t_L \). Given \( Z_o \), the distance of the object plane from the cameras, we replace the coordinates of \( L = (x_L, Z_o) \) in the equation of \((LP_L)\), and solve for \( \sin(t_L) \). We find:

\[
\sin(t_L) = \frac{2s^2 Z_o \pm s x_L \sqrt{4Z_o^2 + x_L^2 - r^4}}{4Z_o^2 + x_L^2}
\]

(23)

Therefore:

\[
H_L = X_F - r | \cos(t_L) |
\]

(24)

Similarly for the right occlusions, we define \( t_R \) using (23), with \( x_L \) replaced by \( x_R \). We find:

\[
H_R = X_F - r | \cos(t_R) |
\]

(25)

The resulting \( H_L \) and \( H_R \) correspond to the real sizes of the occlusion regions on the object. The corresponding hole sizes on the image planes are:

\[
h_L = \frac{f}{Z_o} \cdot (X_F - r \cos(t_L)) \quad \text{and} \quad h_R = \frac{f}{Z_o} \cdot (X_F - r \cos(t_R))
\]

(26)

To evaluate the privacy leakage, one needs to relate the size of the holes on \( I_s \) to \( x_L \) and \( x_R \). We assume that these holes/errrors, that originate from errors on \( D \), do no get further altered by Equations 2. Based on the position of these holes on \( I_s \), an observer can guess three distinct cases:

1. \( (x_R \times x_L) < 0 \): The two cameras are on different sides of the object. The total hole size is consequently \( h = (h_L + h_R) \). In this case, \( x_L \) and \( x_R \) are defined precisely from \( t_L \) and \( t_R \), respectively. Indeed, by plugging \((x_L, Z_o)\) and \((x_R, Z_o)\) in the line equations of \((LP_L)\) and \((RP_R)\), we find:

\[
x_L = 2 \cdot r_L \sin(t_L) - Z_o \tan(t_L) \quad \text{and} \quad x_R = 2 \cdot r_R \sin(t_R) - Z_o \tan(t_R)
\]

This means that positions of the photographers

\(^3\)The same approach applies to \( H_R \).
are known, i.e., the number of suspicious pairs of positions $(L, R)$ is $N_{	ext{susp}} = 1$, and consequently from Equation (14), anonymity $A = 0$.

2. $(x_R \times x_L) > 0$ and $0 < x_L < x_R$: The two cameras are on the right side of the object. In this case $h = \max(h_L, h_R) = h_R$. This means that only the right camera position $R$ can be found precisely, and for the considered profile $F$, we have

$$N_{	ext{susp}} = \left\lceil \frac{x_R}{t} \right\rceil,$$

$$= \left\lceil \frac{1}{t} \left(2 \cdot r_R \sin(t_R) - Z_o \tan(t_R)\right) \right\rceil$$  \hspace{1cm} (27)

3. $(x_R \times x_L) > 0$ and $x_L < x_R < 0$: The two cameras are on the left side of the object. In this case $h = \max(h_L, h_R) = h_L$. This time only the left camera position $L$ can be found precisely, and similarly to Equation (27), for the considered profile $F$, we find:

$$N_{	ext{susp}} = \left\lceil \frac{-x_L}{t} \right\rceil,$$

$$= \left\lceil \frac{-1}{t} \left(2 \cdot r_L \sin(t_L) - Z_o \tan(t_L)\right) \right\rceil$$  \hspace{1cm} (28)

5. EXPERIMENTAL EVALUATION

In this section, we apply view synthesis algorithms on input images taken from two viewpoints to generate images from a third (synthetic) viewpoint. We concentrate on the interpolation case here — the synthetic camera is located in between the two camera locations (to the right of the left camera and to the left of the right camera). We then compare the synthetic image with the ground truth (a real image captured by the same camera from the synthetic viewpoint). The view-synthesis algorithm and the implementation used for generating the synthetic images is the work of Woodford [14].

For our experiments, we used images from the 2006 Middlebury dataset [11, 6] which has been widely used in the computer vision literature. We experimented with two scenes, namely the bowling-scene (with lighting and shadows) and the monopoly-scene (containing an obstruction). For each scene, the dataset contains input images from slightly shifted camera positions that are rectified and the radial distortion has been removed. The baseline between neighbouring pairs is 16cm.

Closely spaced cameras or Narrow baseline: We start with two images taken by cameras located at $L_1$ and $R_1$. In Fig. 3.a and Fig. 3.b we show the input images of the bowling-scene. Fig. 3.c shows the synthetic image generated from a synthetic (virtual) camera position $S_1$ which is placed between $L_1$ and $R_1$, and somewhat closer to $R_1$. Fig. 3.d shows the corresponding ground truth. Similarly Fig. 4.a and Fig. 4.b are input images captured from identical camera locations $L_1$ and $R_1$ of the monopoly-scene. Fig. 4.c is the synthetic image at the same synthetic camera position $S_1$ as for the bowling-scene, while Fig. 4.d is the corresponding ground truth as before.

Wide spaced cameras or Wide baseline: We now move the left and right cameras away from each other resulting in a wider angle between the two input camera viewpoints located at $L_2$ and $R_2$. Fig. 5.a and Fig. 5.b show the two input images of the bowling-scene taken from $L_2$ and $R_2$ respectively. We also choose a different synthetic camera location $S_2$ than before; $S_2$ is roughly in the middle of $L_2$ and $R_2$. Fig. 5.c shows the synthetic image generated while Fig. 5.d shows the corresponding ground truth. Similarly, we have input images (Fig. 6.a and Fig. 6.b), synthetic image (Fig. 6.c) and ground truth (Fig. 6.d) corresponding to $L_2$, $R_2$ and $S_2$ for the monopoly-scene.

In comparison with the closely spaced experiment shown in Fig. 3 and Fig. 4, we observe that view-synthesis in the widely spaced experiment we considered next (Fig. 5 and Fig. 6), shows artifacts and looks less realistic. This supports our analytical result that wider placement leads to more artifacts. As we have argued earlier, artifacts leak information about input camera locations and hence result in lesser anonymity.

Our experiments demonstrate the quality of the synthetic images that can be generated. Even for wider baselines, the images are of

Figure 2: View synthesis setup in the case of interpolation
acceptable visual quality. However, closer inspection reveals artefacts. The most obvious artefacts, in case of the monopoly-scene, appear in the lower-left corner of the synthetic views. These tender spots are magnified in the right column of Fig. 7 and the artefacts are marked. We note that artefacts grow in size and range as the baseline gets wider, which implies a higher leakage of camera-location specific information. This is in accord with Section 4.2 as the total hole size $h$ can be rewritten as:

$$
 h = \frac{f}{Z_0} \cdot \left(2 \cdot X_F - r \cos \left(\frac{\Delta t}{2}\right) \cos \left(\frac{\Sigma t}{2}\right)\right), \tag{29}
$$

where $0 \leq \Delta t = t_L - t_R \leq \pi$ and $\frac{\pi}{2} \leq \Sigma t = t_L + t_R \leq \pi$. We thus find:

$$
 b \nearrow \implies \cos \left(\frac{\Delta t}{2}\right) \to 0 \implies h \nearrow \tag{30}
$$
Figure 7: View synthesis results using different input view pairs
6. CONCLUSIONS AND FUTURE WORK

As social media coverage increases and exceeds the coverage of traditional media, privacy of citizen journalists is an increasingly important problem. In this paper, we have highlighted the threat of camera-location detection attacks mounted by an adversary that combines location clues from published photographs with information from adversary operated cameras. In a world that is increasingly getting saturated with cameras, this is an important privacy problem.

Preliminary investigations on analysing current view synthesis algorithms indicate reasonable anonymity gains. However, given the complexity posed by geometric and algorithmic constraints, there is much potential for improvement of current algorithms from an anonymity and privacy perspective. Our immediate future work will, among others, consist of investigating how sacrificing some image quality could enhance privacy protection; developing security protocols for anonymous collaborative view-synthesis; adding artifacts to maximise photographer anonymity; and, evaluating technical and social aspects of making synthetic images usable in a court of law including resistance to hostile forensic investigations, attribution support and coercion resistance.

The method proposed in this paper indicates a fruitful line of enquiry in developing defence techniques against camera location detection attacks and, in turn, defend against the larger class of photographer de-anonymisation attacks.

7. REFERENCES