Tampering Detection In Low-Power Smart Cameras

Adriano Gaibotti[1], Claudio Marchisio[1], Alexandro Sentinelli[1], and Giacomo Boracchi[2]

[1] STMicroelectronics, Advanced System Technology, Via Camillo Olivetti 2, 20864, Agrate Brianza (MB), Italy
{adriano.gaibotti, claudio.marchisio, alexandro.sentinelli}@st.com
[2] Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Via Ponzio 34/5, 20133, Milano (MI), Italy
giacomo.boracchi@polimi.it

Abstract. A desirable feature for smart cameras is the ability to autonomously detect any tampering event/attack that would prevents a clear view over the monitored scene. No matter whether tampering is due to atmospheric phenomena (e.g., few rain drops over the camera lens) or to malicious attack (e.g., the device displacement), these have to be promptly detected to possibly activate countermeasures. Tampering detection becomes particularly challenging in battery-powered cameras, where it is not possible to acquire images at video-like frame-rates, nor use sophisticated image-analysis algorithms.
We here introduce a tampering-detection algorithm that has been specifically designed for low-power smart cameras: the algorithm leverages very simple indicators that are then monitored by an outlier-detection scheme. Any frame yielding anomalous indicator is detected as a tampering attempt. Core of the algorithm is the partitioning of the scene into adaptively defined regions, that are preliminarily defined by segmenting the image during the algorithm-configuration phase, and which shows to substantially improve the detection of camera displacements. Our experiments show that the proposed algorithm can successfully operate on sequences acquired at very low-frame rate, such as one frame every minute, and at a very small computational complexity.

Keywords: tampering detection, defocus, displacement detection

1 Introduction

When cameras operate outdoor and in harsh environments, dust, rain drops or snow flakes might lie on the camera lens resulting in blurry pictures, as in Fig. 1(a), or in partial occlusions of the scene, as in Fig. 1(b). Similarly, other intentional attacks like displacing the camera, changing its focus, or spraying some opaque or glossy liquid over the lenses, would result in heavily compromised pictures, that would be surely useless for monitoring purposes. We refer to these events/attacks as tampering. In some cases, tampering is easy to detect, e.g.,

when the camera integrity is affected and the device goes out-of-order. However, in many other situations, namely when the device is not physically damaged, but the correct interpretation of the scene or of small important details is prevented (e.g., the identification of licence plates), tampering detection is not straightforward and requires image-analysis algorithms.

The early detection of tampering events/attacks is clearly essential in surveillance systems [**Giacomo**: Adriano, add REFERENCES] [1], where cameras are expected to autonomously detect any tampering event/attack and promptly report alerts. Surveillance cameras are typically connected to the power supply and acquire and process images at normal frame-rates (e.g. around few frames per second). In these conditions, several algorithms have been presented in the literature [**Giacomo**: Adriano add REFERENCES].

[**Giacomo**: Qui ST puó aggiungere qualche esempio di applicazione qui o menzionare dispositivi di riferimento (con tanto di link a datasheet). Magari aggiungendo che le batterie recenti permettono operativitá di due anni a questi regimi.][**Claudio**: Verifico cosa possiamo dire del SecSoc.] This work expressly targets low-power and ultra-low-power smart cameras. Such devices typically operate at very low frame-rates (e.g. possibly less than one frame every minute), with constrained computational power and memory, and are battery powered (typically lasting one or two years). As an example, consider Wireless Multimedia Sensor Networks (WMSN) [2] where nodes are wirelessy connected smart-cameras, that can acquire and transmit images at regular time interval or upon requests. Often, in WMSN, the units are not connected to the power supply and have to operate with batteries, possibly implementing energy harvesting mechanisms [**Giacomo**: REFERENCES, magari riprescando quelle del paper di TIM]. Even though these devices are not employed in critical surveillance applications, low-power smart cameras are becoming popular in distributed monitoring of wide environments due to their low cost and maintenance requirements [**Giacomo**: possiamo mettere qualche REFERENCE]. Tampering detection is thus very important in low-power smart cameras, also considering that tampered frames should be identified to avoid unnecessary energy-demanding operations like the local processing or radio activation for transmission over the network.

Tampering detection in low-power smart cameras is much more challenging than in conventional surveillance cameras [3]. Beside computational aspects – such as the number of operations per pixels allowed – the big issue is that low-power smart cameras typically operate at very low-frame rates (e.g., less than one frame per minute), and the acquired sequence does not evolve smoothly. This prevents the use of learned background models and the analysis of foreground variations [4]. When dynamic environments are acquired at low frame rates, like the example depicted in Fig. **??**, two consecutive frames might be very different because of changes in the scene and in the light conditions: smart cameras have to correctly distinguish between these *normal* changes and changes due to camera displacement or blur/defocus. Moreover, in low-power smart cameras, tampering has to be reliably detected because false alerts would lead to unnecessary energy waste.

(a) (b)

Fig. 1. Examples of tampering events due to atmospheric phenomena. **(a)** Blur due to rain drops on the camera lens. **(b)** Occlusion due to some snow on the camera lens.

This work presents an algorithm to detect both camera defocus and displacement, as stated in Section 2. The algorithm relies on simple indicators that are computed with a low computational complexity (Section 3.2) and leverages an outlier detection technique to detect frames yielding anomalous indicators as related to a tampering event (Section 3.3). In particular, we perform a preliminarily partitioning of the scene into nonoverlapping regions, and separately compute and monitor indicators over different regions. Scene partitioning is defined by clustering feature vectors gathering the indicators in different pixels, thus performing a very coarse segmentation of the image (Section 3.1). Regions are computed during the initial configuration of the algorithm and as such do not implies any computational overhead w.r.t. operating on the whole image. Our experiments in Section 4 show that regions substantially improve the detection of camera displacements, while when monitoring indicators meant to detect blur/defocus, it is more convenient to consider the whole scene at once. Concluding remarks and discussions are given in Section 5.

1.1 Related Works

The literature concerning tampering detection is mostly focused on video surveillance applications and operates at few frames per second [**Giacomo**: Adriano, é vero?]. Background models are typically leveraged to identify defocus and occlusions; in particular [5] performs detect defocus by analyzing the wavelet domain of each frame and performs histogram comparison for detecting occlusions, while camera displacements are not considered. A background-subtraction technique is employed in [6] to identify defocus, occlusions, and displacements. [**Giacomo**: FIXME Comparison in the Fourier domain for defocus detection, histogram comparison for occlusion detection, comparison between current background and delayed background for displacement detection]. Background subtraction is also used in [7] to detect defocus, occlusions, and displacements. [**Giacomo**: FIXME Comparison of edges pixels count for defocus detection, entropy comparison for occlusion detection, block matching algorithm for displacement detection]. In contrast, no back-

ground models are used in [8], where a sequential monitoring scheme based on a change-detection test is employed to detect changes in the average gradient energy of each frame to detect defocus due to external disturbances on the camera lens. [9]: comparison between frames belonging to a buffer in order to find high values of dissimilarity, associated to tampering. [10]: implementation in a FPGA of a solution based on background modeling, histograms comparisons, edges comparisons. [11]: tampering detection inside a moving vehicle; uses background subtraction methods in order to identify defocus, occlusions, and displacements. Comparison of edges pixels count for defocus detection, entropy comparison for occlusion detection, block matching algorithm for displacement detection. [12]: monitoring of the number of key points extracted by SURF in order to detect defocus events, partition in blocks and HOG descriptors matching for each block in order to detect occlusions. These types of solutions requires a lot of computations

2 Problem Formulation

Let z_t be the frame acquired at time t, which we model as

$$z_t(x) = \mathcal{D}_t[y_t](x) \quad \forall x \in X \tag{1}$$

where $\mathcal{D}_t$ denotes the degradation operator that transforms the original image y_t in the frame z_t; $X \subset \mathbb{Z}^2$ denotes the regular pixels grid and $x \in \mathbb{Z}^2$ indicates the pixel coordinates. As far as there is no tampering attacks/events,

$$\mathcal{D}_t[y_t](x) = y_t(x) + \eta_t(x) \quad \forall x \in X \tag{2}$$

where η_t is a random variable accounting for different noise sources (e.g., thermal, quantization, photon-counting). In normal conditions, all the images y_t (thus also the frames z_t) might show different content but are acquired from the same viewpoint and the same camera orientation.

When at time τ^* an external disturbance introduces *blur/defocus*, the image y_t is degraded by a spatially variant blur operator, and z_t becomes

$$\mathcal{D}_t[y_t](x) = \int_{\mathcal{X}} y(s) h_t(x,s) ds \; + \eta_t(x) \quad \forall x \in X, t \geq \tau^* \tag{3}$$

where $h_t(x,\cdot) > 0$ is the point-spread function at pixel $x \in X$. A *camera displacement* at frame τ^* is instead modeled as

$$z_t(x) = \begin{cases} y_t(x) + \eta(x) \text{ per } t < T^* \\ w_t(x) + \eta(x) \text{ per } t \geqslant T^* \end{cases}, \tag{4}$$

where w_t relates to a different viewpoint and/or camera orientation than y_t.

The proposed tampering-detection algorithm analyzes a sequence of frames $\{z_t\}_t$ to detect time τ^* when any tampering like (3) or (4) occurs.

3 Proposed Solution

3.1 Scene Segmentation

Consideriamo la sequenza $\{z_t\}$ di frame acquisiti dalla camera, con $t = 1, \ldots, T_c$. Per ciascun pixel $x \in \mathcal{X}$ calcoliamo un vettore $\mathbf{d}(x)$ di 5 elementi

$$\mathbf{d}(x) = [r(x); c(x); \mu_\nabla(x); \sigma_\nabla(x); \bar{z}(x)]\,, \mathbf{d}(x) \in \mathbb{R}^5 \tag{5}$$

dove:

- $r(x)$ rappresenta il numero di riga del pixel x.
- $c(x)$ rappresenta il numero di colonna del pixel x.
- $\mu_\nabla(x)$ rappresenta il valore del gradiente nel pixel x mediato nel tempo:

$$\mu_\nabla(x) = \frac{\sum_{t=1}^{T_c} (\|\nabla z_t\|_2^2 \circledast f)(x)}{T_c}, \tag{6}$$

  dove abbiamo indicato con $\|\nabla z_t\|_2^2$ la norma del gradiente per l'immagine z_t, definita in (11), e con f il filtro gaussiano discreto derivato dal campionamento di (10).
- $\sigma_\nabla(x)$ rappresenta la deviazione standard nel tempo del gradiente nel pixel x:

$$\sigma_\nabla(x) = \sqrt{\frac{1}{T_c - 1} \sum_{t=1}^{T_c} \left(\left(\|\nabla z_t\|_2^2 \circledast f \right)(x) - \mu_\nabla(x) \right)^2}. \tag{7}$$

- $\bar{z}(x)$ rappresenta il valore della luma del pixel x mediato nel tempo:

$$\bar{z}(x) = \frac{\sum_{t=1}^{T_c} (z_t \circledast f)(x)}{T_c}. \tag{8}$$

3.2 Indicators

[**Giacomo**: Adriano: mettere formule degli indicatori e anche del frame difference qua]

$$\begin{aligned} g^k(t) &= \mathcal{G}^k[z_t] = \frac{\sum_{R_k} \|\nabla z_t(x)\|_2^2}{|R_k|} \\ \partial g^k(t) &= g^k(t) - g^k(t-1) \end{aligned}, \tag{9}$$

In particolare, per il calcolo delle derivate orizzontali abbiamo utilizzato il seguente filtro f_h:

$$f_h = f \circledast \begin{bmatrix} 1 & 0 & -1 \end{bmatrix},$$

mentre per il calcolo delle derivate verticali abbiamo utilizzato il seguente filtro f_v:

$$f_v = f \circledast \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix},$$

dove abbiamo indicato con $\circledast$ l'operatore di convoluzione. Il filtro f, invece, è ottenuto tramite un campionamento della *funzione gaussiana* h, con media 0 e deviazione standard σ

$$h(i,j) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{i^2+j^2}{2\sigma^2}\right), \tag{10}$$

e ponendo il valore massimo di questa funzione nel centro del filtro. Con questi filtri è possibile calcolare la *norma del gradiente* nel seguente modo:

$$\|\nabla z_t(x)\|_2^2 = (z_t \circledast f_h)(x)^2 + (z_t \circledast f_v)(x)^2. \tag{11}$$

Una volta calcolata la norma del gradiente è possibile farne la media come specificato in (**??**). Il risultato finale è un indicatore *scalare* per ciascun frame acquisito, che può essere monitorato per individuare eventi di sfocature. In particolare ci aspettiamo che l'evento di sfocatura provochi un abbattimento del valore di g.

$$\begin{array}{rl} l^k(t) & = \mathcal{L}^k[z_t] = \frac{\sum_{R_k} z_t(x)}{|R_k|} \\ \partial l^k(t) = & l^k(t) - l^k(t-1) \end{array}, \tag{12}$$

$$l(t) = \mathcal{L}[z_t] = \frac{\sum_{\mathcal{X}} z_t(x)}{|\mathcal{X}|}, \tag{13}$$

$$\frac{\partial g}{\partial t}(t) = g(t) - g(t-1), \tag{14}$$

$$\frac{\partial l}{\partial t}(t) = l(t) - l(t-1). \tag{15}$$

FRAME DIFFERENCE:

$$\varphi_k(t) = \frac{\sum_{x \in R_k} (z_t(x) - z_{t-1}(x))^2}{|R_k|}, k = 1, \ldots, K. \tag{16}$$

3.3 Outlier Detection

$$\begin{array}{l} \Gamma^k_{min} = \hat{\mu}^k_g - \gamma \hat{\sigma}^k_g \\ \Gamma^k_{max} = \hat{\mu}^k_g + \gamma \hat{\sigma}^k_g \end{array}, \tag{17}$$

dove $\hat{\mu}^k_g$ indica il valore medio delle osservazioni del training set

$$\hat{\mu}^k_g = \frac{\sum_{\tau=1}^{T_o} \frac{\partial g^k}{\partial t}(\tau)}{T_o},$$

$\hat{\sigma}^k_g$ indica la deviazione standard delle osservazioni del training set

$$\hat{\sigma}^k_g = \sqrt{\frac{1}{T_o - 1} \sum_{\tau=1}^{T_o} \left(\frac{\partial g^k}{\partial t}(\tau) - \hat{\mu}^k_g(\tau)\right)^2}$$

e $\gamma > 1$ è un parametro moltiplicativo ottenuto sperimentalmente.

$$\begin{aligned} \Gamma^k_{min} &= \hat{\mu}^k_l - \gamma \hat{\sigma}^k_l \\ \Gamma^k_{max} &= \hat{\mu}^k_l + \gamma \hat{\sigma}^k_l \end{aligned}, \tag{18}$$

dove $\hat{\mu}^k_l$ indica il valore medio delle osservazioni del training set

$$\hat{\mu}^k_l = \frac{\sum_{\tau=1}^{T_o} \frac{\partial l^k}{\partial t}(\tau)}{T_o},$$

$\hat{\sigma}^k_l$ indica la deviazione standard delle osservazioni del training set

$$\hat{\sigma}^k_l = \sqrt{\frac{1}{T_o - 1} \sum_{\tau=1}^{T_o} \left(\frac{\partial l^k}{\partial t}(\tau) - \hat{\mu}^k_l(\tau) \right)^2}$$

e $\gamma > 1$ è un parametro moltiplicativo ottenuto sperimentalmente.

3.4 Algorithm Summary

4 Experiments

The proposed algorithm has been implemented in MATLAB, and has been tested on two datasets. The first one refers to frame sequences taken from webcams recording some parts of cities (as in Figure 2(a)), where we have introduced some synthetically generated tampering events: defocus has been simulated using gaussian filters (2(b)), while displacement has been created with concatenation of similar frame sequences (Figure 2(c)). In some cases these sequences contained real tampering events, as in Figure 2(d).

The second dataset was taken using a Raspberry Pi Model B+, with its camera module, and the tampering was introduced by moving the device or putting water on the camera.

Four figures of merit have been suggested to assess the performance of the proposed algorithm:

TP True positives. It measures the number of tampering events correctly detected by the algorithm.

TN True negatives. It measures the number of frames without tampering that are not detected by the algorithm.

FP False positives. It measures the number of tampering events erroneously detected by the algorithm.

Algorithm 1: Tampering detection algorithm

Input: γ, T_o, $\{R_k\}, k = 1, \ldots, K$
Configuration:
1. **for** $t = 1, \ldots, T_o$ **do**
2. Get frame z_t
3. **for** $k = 1, \ldots, K$ **do**
4. Compute $l^k(t)$, $\partial l^k(t)$ for the region R_k
end
end
5. **for** $k = 1, \ldots, K$ **do**
6. Compute σ_l^k
end
Operational phase:
7. **for** $t = T_o, \ldots, \infty$ **do**
8. Get frame z_t
9. $n_l = 0$
10. **for** $k = 1, \ldots, K$ **do**
11. Compute $l^k(t)$, $\partial l^k(t)$ for the region R_k
12. **if** $\partial l^k(t) < -\gamma\sigma_l^k \vee \partial l^k(t) > \gamma\sigma_l^k$ **then**
13. $n_l = n_l + 1$
end
end
14. **if** $n_l \geq K - 1$ **then**
15. z_t is a tampered frame
end
end

Fig. 2. Sequences taken from webcams: **(a)** no tampering events; **(b)** defocus event on 4-th and 5-th frames, created using gaussian filtering; **(c)** displacement event on 4-th and 5-th frames, created using concatenation between similar sequences; **(d)** real displacement event on 4-th frame.

FN False negatives. It measures the number of tampering events not detected by the algorithm.

These indicators are computed varying the parameter γ, which defines the thresholds for the one-shot monitoring. This permits us to generate *ROC curves*, where on the x-axis there is:

$$1 - \text{SPECIFICITY}_\gamma = 1 - \frac{\text{TN}_\gamma}{\text{TN}_\gamma + \text{FP}_\gamma} = \frac{\text{FP}_\gamma}{\text{TN}_\gamma + \text{FP}_\gamma},$$

while on th y-axis there is:

$$\text{RECALL}_\gamma = \frac{\text{TP}_\gamma}{\text{TP}_\gamma + \text{FN}_\gamma}.$$

The construction of these curves permits us to make a comparison with respect to other methods. In particular the alternative approaches that we have considered working:

- considering the whole scene for the features computation;
- considering adaptive region, as described in Section 3, for the feature computation;
- considering voronoi regions [13], which are easier to compute with respect to our solution but don't consider the scene content, for the feature computation.

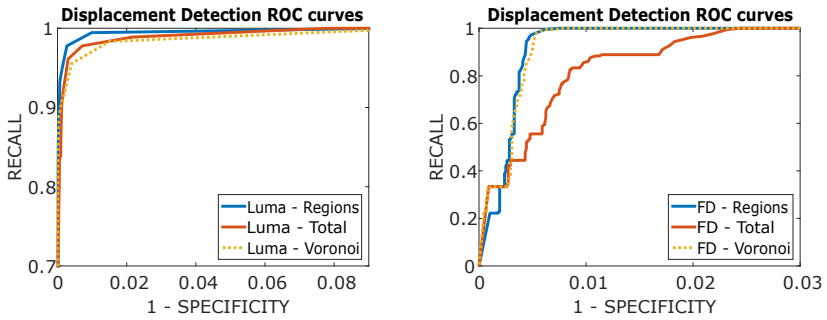

Fig. 3. ROC curves for displacement detection, considering three alternative approaches. **(a)** Analysis of the mean luma energy. **(b)** Analysis of the frame differencing.

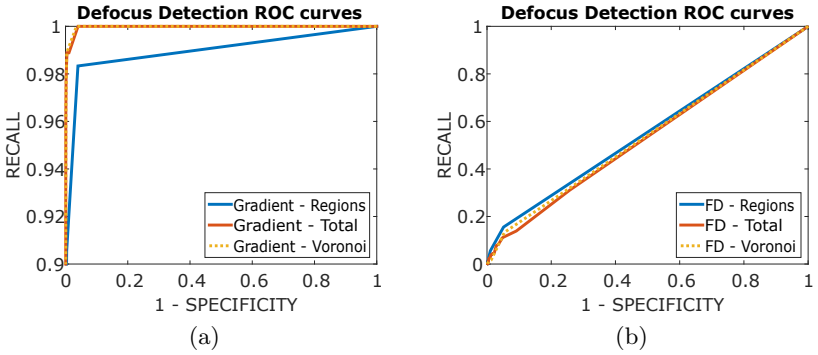

Fig. 4. ROC curves for defocus detection, considering three alternative approaches. **(a)** Analysis of the mean gradient energy. **(b)** Analysis of the frame differencing.

Experimental results are shown in Figures 4 and 4. As we could see there is an improvement, in the detection of camera displacements, with our approach which separately analizes the behavior of indicators in adaptive regions. On the other side, as illustrated in Figure 4, when monitoring indicators meant to detect blur/defocus, it is more convenient to consider the whole scene at once.

4.1 Discussion

[**Giacomo**: Adriano: Aggiungi qua la complessitá computazionale]

5 Conclusion

[**Giacomo**: Adriano: butta in inglese gli ongoing works (come ultima cosa)] .
Random Toughs:

- The problem of false alarms, radio module activation
- Other tampering attacks like obfuscation (??) which might be due to environmental phenomena such as rain, fog and mist over the camera lenses have to be detected by image analysis methods
- Displacement can be perceived by MEMS as well but these device alone are prone to false alarms. Visual inspection is necessary to reduce false alarms

Ongoing work regards the extension of our solution to other types of tampering, as occlusions or imaging sensor degradations. Furthermore, we are investigating strategies to improve the detection performance and reduce the number of *FP*s by combination of our solution with other techniques: for example we could use sequential techniques on the features, or integrate the frame analysis with the data extracted from MEMS inertial sensors.

Acknowledgments

Authors would like to thank ST for supporting Adriano Gaibotti.

References

1. Arun Hampapur, Lisa Brown, Jonathan Connell, Ahmet Ekin, Norman Haas, Max Lu, Hans Merkl, and Sharath Pankanti. Smart video surveillance: exploring the concept of multiscale spatiotemporal tracking. *Signal Processing Magazine, IEEE*, 22(2):38–51, 2005.
2. Ian F Akyildiz, Tommaso Melodia, and Kaushik R Chowdhury. A survey on wireless multimedia sensor networks. *Computer networks*, 51(4):921–960, 2007.
3. Adrian Perrig, John Stankovic, and David Wagner. Security in wireless sensor networks. *Communications of the ACM*, 47(6):53–57, 2004.
4. Massimo Piccardi. Background subtraction techniques: a review. In *Systems, man and cybernetics, 2004 IEEE international conference on*, volume 4, pages 3099–3104. IEEE, 2004.

5. Anil Aksay, Alptekin Temizel, and A. Enis Cetin. Camera tamper detection using wavelet analysis for video surveillance. In *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*, pages 558–562. IEEE, 2007.
6. Ali Saglam and Alptekin Temizel. Real-time adaptive camera tamper detection for video surveillance. In *Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference on*, pages 430–435. IEEE, 2009.
7. Pedro Gil-Jiménez, R. López-Sastre, Philip Siegmann, Javier Acevedo-Rodríguez, and Saturnino Maldonado-Bascón. Automatic control of video surveillance camera sabotage. *Nature Inspired Problem-Solving Methods in Knowledge Engineering*, pages 222–231, 2007.
8. Cesare Alippi, Giacomo Boracchi, Romolo Camplani, and Manuel Roveri. Detecting external disturbances on the camera lens in wireless multimedia sensor networks. *Instrumentation and Measurement, IEEE Transactions on*, 59(11):2982–2990, 2010.
9. Evan Ribnick, Stefan Atev, Osama Masoud, Nikolaos Papanikolopoulos, and Richard Voyles. Real-time detection of camera tampering. In *Video and Signal Based Surveillance, 2006. AVSS'06. IEEE International Conference on*, pages 10–10. IEEE, 2006.
10. T Kryjak, M Komorkiewicz, and M Gorgon. Fpga implementation of camera tamper detection in real-time. In *Design and Architectures for Signal and Image Processing (DASIP), 2012 Conference on*, pages 1–8. IEEE, 2012.
11. Sebastien Harasse, Laurent Bonnaud, Alice Caplier, and Michel Desvignes. Automated camera dysfunctions detection. In *Image Analysis and Interpretation, 2004. 6th IEEE Southwest Symposium on*, pages 36–40. IEEE, 2004.
12. Theodore Tsesmelis, Lars Christensen, Preben Fihl, and Thomas B Moeslund. Tamper detection for active surveillance systems. In *Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on*, pages 57–62. IEEE, 2013.
13. Franz Aurenhammer. Voronoi diagrams – a survey of a fundamental geometric data structure. *ACM Computing Surveys (CSUR)*, 23(3):345–405, 1991.