Article

Perception of Risks and Usefulness of Smart Video Surveillance Systems

Thomas Golda [1,2,*,†] 🔘**, Deborah Guaia** [3,†] **and Verena Wagner-Hartl** [3,*,†] 🔘

1 Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB,
   76131 Karlsruhe, Germany
2 Vision and Fusion Laboratory, Karlsruhe Institute of Technology KIT, 76131 Karlsruhe, Germany
3 Faculty Industrial Technologies, Furtwangen University, Campus Tuttlingen, 78532 Tuttlingen, Germany
* Correspondence: thomas.golda@iosb.fraunhofer.de (T.G.); verena.wagner-hartl@hs-furtwangen.de (V.W.-H.)
† These authors contributed equally to this work.

Abstract: The number of video cameras in public places increases due to different reasons such as detecting dangers (e.g., thefts, robberies, terrorist attacks) and security breaches in crowds. The application of video surveillance systems is sometimes evaluated ambivalently; therefore, the presented study focuses on factors influencing the acceptance of a privacy-friendly, smart video surveillance system. Overall, 216 persons aged between 18 and 81 years participated in an online survey. In terms of the perceived usefulness, there are significant interactions of public spaces × gender and public spaces × time of day. In addition, the assessment of different privacy levels of a video surveillance system differ significantly in terms of perceived risk. Interestingly, men rate the risk concerning their own privacy significantly higher than women do. Participants rate the presented system as fairly useful and slightly risky for their own privacy. The findings of the presented exploratory study provide insight into how people perceive smart video surveillance. These findings have the potential to support the conditions of the use of smart video surveillance systems and to address the possibly affected individuals.

Keywords: smart video surveillance; intelligent video surveillance; machine learning; human pose estimation; privacy protection; online survey; exploratory study; time of day; public areas; gender; perceived privacy risk

1. Introduction

In the past, the number of video cameras in public places increased due to reasons such as detecting dangers and security breaches in crowds [1–3]. Especially terror attacks, such as those that occurred in Hanau (Germany) and Metz (France) in 2020 [4], pose a huge threat to the safety and security of people. Therefore, it is important to be able to assess risky situations at large events directly when they occur [2]. However, this also brings challenges with it, as the amount of data is continuously increasing and a manual analysis of the image material requires a lot of effort.

In a study about crime anxiety, 70% of 3023 people mentioned that they perceived the duty of state authorities to take action against crime [5]. Among other steps, they therefore supported conventional video surveillance. For this reason, a smart solution, i.e., driven by machine learning methods and other algorithms, for the evaluation of the data is necessary [2]. Such technically advanced systems might confront people with their fear of artificial intelligence. Therefore, such algorithms should respect the acceptance and privacy of the population [6]. The results of a study on video surveillance showed that 40% of about 1000 European civilians perceived video surveillance as an invasion of their privacy [7].

When developing smart monitoring systems, the factors influencing acceptance such as usefulness, risks [6] and privacy [8] should be determined. In this way, the acceptance of these systems can be achieved by adapting it to the fears and doubts of those affected and

not vice versa. Possible ways to address this issue could be to censor, pixelate or to ensure an increased level of data-privacy by omitting information that allows the identification of people [9–11].

In the presented study, we focus on the third solution. It can be seen as an exemplary system for smart video surveillance that does not analyze faces [2,12,13]. Such solutions are typically designed to detect and evaluate abnormalities and salience. The focus is on movement and motion patterns by looking at the position of the body and its limbs. For this purpose, the algorithm of this system creates a digital skeleton of the human body. This type of smart video surveillance is examined in this paper with regard to its acceptance. For this reason, the Technology Acceptance Model for Video Surveillance (TAM-VS) proposed by Krempel and Beyerer [6,14] was mainly applied in the presented study. In the TAM-VS, the perceived usefulness and the risk of misuse were identified as factors influencing the acceptance (see Section 1.3).

*1.1. Smart Video Surveillance Using Digital Skeletons*

Conventional video surveillance is a topic that is already present in many cities. With the increasing number of installed video cameras [1], retaining an overview of data becomes increasingly difficult. At this point, smart video surveillance receives more attention, especially as assisting systems for video operators. In recent years, many technological advances have resulted in interesting application possibilities for smart video surveillance systems. Security services already started to rely on unmanned aerial vehicles (UAV) for short-term use cases such as recon flights at disaster scenarios, which opens up new possibilities, as initial work has already demonstrated their suitability for the detection and localization of persons and objects [15–17]. The use of facial recognition for tracking or recognition in classical video surveillance setups has also experienced significant improvement in the past [18–20] as well as attribute-based person re-identification [21,22]. Even the field of action recognition has also seen a variety of methods emerge in the near past, promising ever better accuracies and real-time processing, which is essential for serious applications [23,24]. The aforementioned aspects appear to be solutions for existing problems, such as the detection of salient behaving or hurt pedestrians, but the use of such methods is restricted by tough standards, especially in Europe or other Western countries. The General Data Protection Regulation (GDPR) [25] is a regulation in EU law on data protection and privacy in the European Union (EU) and the European Economic Area (EEA), which strongly limits the application of existing methods for smart video surveillance systems. In addition, the Regulation (EU) 2019/947 [26] restricts and regulates the application of UAVs for civil and public use in Europe [27]. Therefore, the urgent need for privacy-friendly methods and approaches in ground-related video surveillance emerges. There are different ways to deal with the legal limitations, such as the application of synthetic data for tracking or human pose estimation [28–30], and the utilization of human poses for behavior analysis of pedestrians, meaning that the analysis of pedestrians solely relies on motion and body-structure information. Especially the last one is an approach that has been successfully brought into practice in a project within the German city of Mannheim [2,3,10]. The presented work examines the factors influencing acceptance toward smart video surveillance, in particular that of using digital skeletons for behavior analysis [2,10]. The aim of such algorithms is to recognize crimes and atypical behaviors and movements of pedestrians (including tripping and collapsing) based on patterns in a live transmission of the video material. The typical setup proposed in the literature [2,12,13,31,32] shares the same structure: First, the system determines the spatial location of limbs for all human bodies within the given scene. This is completed either by locating the pedestrians first and then determining the position of the limbs (top–down) or by determining the position of the limbs first (bottom–up), and then associating them to corresponding persons. These information are then collected over time in order to extract spatio-temporal features for the following step. The collected information is used to create the digital skeletons, which are typically also referred to as

human poses. The most common way to extract such skeletons is to use machine learning methods.

This basic algorithmic approach can be included into a productive overall system, which could look as follows: The smart video analysis is part of a video management system (VMS), which is responsible for acquiring live video material from, e.g., IP cameras. The image material is then forwarded to the smart subsystem, processed, and the results are returned to the VMS. The VMS can then provide results, such as the location and the recognized behavior, such as punching, kicking or tripping and even falling, to a human operator. As mentioned earlier, biometric and soft biometric features are neglected and not analyzed [2]. This includes, in particular, the face or the identity of all visible people. With the given information and the original video material, the human operator checks whether a real critical situation can be seen on the recordings or not. If necessary, the operator will initiate an intervention by security staff or police officers. A particular example for such a system is given in [2,10,33], who follow the described approach.

1.2. Technical Principles of the Smart Subsystem

The description of the general workflow of the abnormality detection software is presented in Figure 1 [2,34]. The process is divided into three parts: The input of the neural network represents the used image material as well as sequences of images collected by video cameras. The first step of a general solution approach is the extraction of the posture of several human bodies simultaneously from this image material. This step is also called Multi-Person Pose Estimation [35,36] or just Human Pose Estimation [29,37]. The outputs are keypoints of the digital skeleton (see Figure 2), which relate to various body parts such as joints or some relevant body parts such as eyes or nose. That way, the resulting skeleton correlates with the posture of the corresponding human being, i.e., its body. In the next step, the movements of the body are identified by capturing the keypoints over several timesteps, which can be completed by using optical flow estimation or other tracking algorithms. This results in motion features.
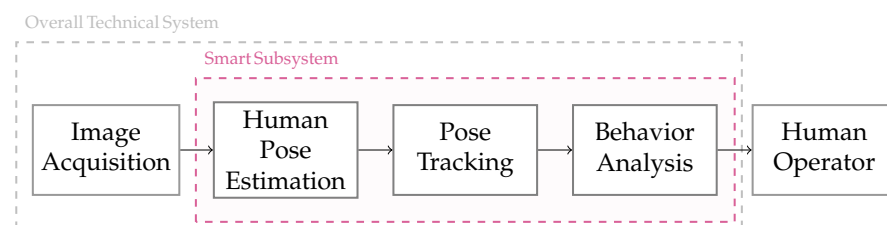
Figure 1. Simplified schematic block diagram showing a possible workflow in smart video surveillance as an aiding system. A video management system (VMS) is responsible for the data acquisition and provides the data to the smart subsystem. The output of the subsystem is then provided to the VMS, which presents it to the human operator. The human operator then takes these results and decides whether to perform additional steps (e.g., send a police officer to the location) or ignore them.

Based on this extracted motion information, a neural network is used to determine whether an abnormality is present or not. The achieved output of the generally represented procedure is an anomaly feature. The exact kind of feature depends on the chosen algorithm and can range from heatmaps [2] to classified bounding boxes [12,13,31,32].
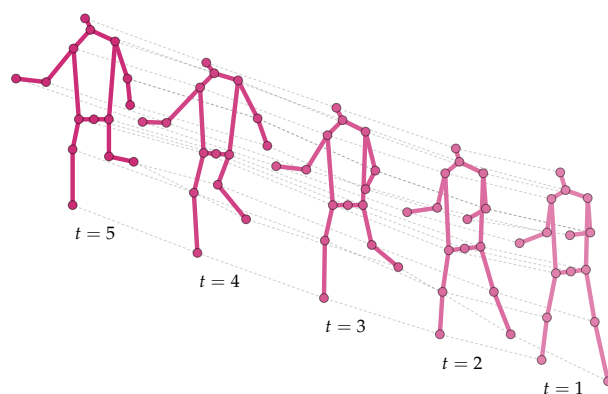
Figure 2. Exemplary skeleton sequence visualizing the movement of a person in five timesteps. The length of the sequence depends on the chosen algorithm. Each timestep consists of 15 keypoints, which again is just exemplary and depends on the chosen pose model. However, the basic shape of a person is recognizable.

1.3. Acceptance Model

The Technology Acceptance Model for Video Surveillance (TAM-VS) by Krempel and Beyerer [6,14] is a reinterpretation of Davis' [38] TAM model and focuses on video-based surveillance. TAM-VS [6,14] shows that the acceptance of such systems is influenced by the perceived usefulness and the emotional attitude toward an unknown technology. This feeling-based attitude is influenced by the risk of misuse posed by the application of video surveillance. Furthermore, the acceptance model used for the ubiquitous computing model mentions usage risk as a factor influencing acceptance [39]. The risk of abuse or usage risks was adapted for the present study as perceived privacy risk. Moreover, the presented research focuses on the direct influence of perceived usefulness on acceptance from the TAM-VS [6,14].

1.4. Perceived Usefulness

The perceived usefulness reflects the extent to which an increase in security through smart video surveillance is considered to be purposeful [6]. Video surveillance is usually an invasion of privacy. However, this invasion of privacy becomes acceptable for the affected people if the importance of the objective outweighs the invasion. There is almost no information on existing systems that work in the presented way despite a German project [10,40], although there are many cities all over the world using video surveillance and even smart systems, as presented in [1].

The authorities report that the use of smart video surveillance at different public areas in Mannheim produces different effects on crime frequencies [3]. Earlier, Hölscher [41] investigated various public areas, concerning their perception of usefulness using video surveillance. Therefore, the author chose areas such as train stations, playgrounds, parks, workplaces and residential areas for his study. In these public areas, the rated usefulness of traditional video surveillance differs descriptively. Klauser [42] asked participants whether they felt disturbed by conventional video surveillance or not. Related to this, he examined, e.g., residential areas, department stores and parks as well as the entrances of shops. Additionally, there was already an investigation by Kudlacek [43] regarding the acceptance of smart video surveillance in airports.

Furthermore, a study by Forster et al. [44] report an increase of the perception of security through video surveillance at large events as well as within crowds. Moreover, a descriptive difference between daytime hours in public areas such as train stations was also discovered [5].

In addition, a longitudinal study found a difference in approval of conventional video surveillance for different genders [45,46]. They indicate that men are generally more skeptical about video surveillance than women. In addition, Klauser [42] reports that

women show a stronger acceptance of video surveillance compared to men. However, previous studies regarding attitudes toward video surveillance systems have not dealt with interactions of different variables such as gender, public areas or the time of day when the different places or public areas are visited. Therefore, there is still uncertainly regarding whether such effects of interactions do exist and to what extend they are able to influence the perceived usefulness of video surveillance systems. Consequently, this will be explored in the presented study.

1.5. Perceived Privacy Risk

As mentioned previously, 40% of about 1000 people in a European study by Hempel and Töpfer [7], perceive video surveillance as a violation of their own privacy. In Klauser's [42] study, 21% perceive such a security measure as a threat to privacy. Furthermore, there seems to be a difference between genders [42,45,46].

In Hempel and Töpfer's [7] study, more than 50% of the 1000 respondents reported that a conventional video surveillance system could be misused. Consideration of privacy in the architecture of a technical system represents a basic principle [47]. In this context, the right of privacy should be preserved by submitting the data that can be traced back to an individual to a protection mechanism. There are also other research approaches for the development of privacy-compliant smart video surveillance than the analysis of digital skeletons [10,48]: for example, the possibility to identify crime weapons by such a smart system. Accordingly, the video material is analyzed only with regard to such objects. Video analysis based on the censoring of biometric features (e.g., faces) is also interesting for acceptance research [9,11]. Moreover, there are person-based smart video surveillance systems that analyze biometric features and can thus draw conclusions about the observed person [49]. Currently, the perceived privacy risk has not been assessed in terms of different privacy levels of smart video surveillance. Furthermore, as previously described for the acceptance and perceived usefulness of video surveillance systems, it also appears for the perceived privacy risk that interactions of different variables and their impact have not been considered in the literature to date. Therefore, this should also be emphasized in the presented study.

1.6. Aim of the Study

Based on the previously obtained findings, this study examines the following research questions: (1) Are there differences between gender, public areas, and the time of day when visiting certain public areas regarding the perceived usefulness of the smart video surveillance system? (2) Are there gender differences for different privacy levels of smart video surveillance, regarding the perceived privacy risk?

2. Method

2.1. Participants

In total, 217 people have participated of which 117 were men and 99 were women. Only a single non-binary person took part. Due to data protection reasons, this person was not included in the final sample ($N = 216$). The age of the participants ranged from 18 up to 81 years ($M = 35.65, SD = 13.60$). Overall, 212 participants were German, or if they had more than one nationality, they reported that they felt that they belonged to the German nationality. Regarding education, 3% of the participants had a middle school degree, 7% of the participants had completed an apprenticeship, and 27% of the participants reported an A-level or a specialized A-level as their highest degree. More than 50% of the participants reported that they had an academic education; of these, 21% of the participants had a bachelor's degree, 28% of them had a master's degree or a diploma, 11% had a doctorate, and 3% had another degree (five of them a state exam). Women and men are equally distributed with respect to their highest level of education, $\chi^2(6, N = 216) = 12.34, p = 0.055$. Significantly more men paid attention to whether video cameras were installed in their everyday environment than women did,

$\chi^2(1, N = 216) = 10.09, \rho = 0.001$. Sixteen percent of the participants described themselves as crime victims. Women and men did not differ significantly with respect to their past of victimization experience, $\chi^2(1, N = 216) = 0.29, \rho = 0.865$.

Concern about becoming a victim of crime in the future is not equally distributed across gender, $t(180.51) = 5.43, \rho \leq 0.0001$. The fear of becoming a victim of crime was rated significantly higher by women ($M = 2.83, SD = 0.78$) than by men ($M = 2.31, SD = 0.59$). In addition, the participants were asked to what extent they trusted that the government could manage potential privacy risks arising from smart video surveillance of digital skeletons. Men reported significantly less trust in government ($M = 2.68, SD = 1.06$) than women ($M = 3.16, SD = 0.99$), $t(214) = 3.42, \rho = 0.001$. Furthermore, women ($M = 3.62, SD = 0.90$) rated themselves as significantly less tech-savvy (excitement for new systems and functions, cf. [50]) than men ($M = 4.15, SD = 0.77$), $t(193.87) = -4.60$, $\rho \leq 0.0001$. On average, the participants' affinity for technology was rated fairly high ($M = 3.90, SD = 0.87$). Participants could answer on a 5-point scale from not at all (1) to exceptionally (5).

About 18% of the sample participants reported spending most of their time in rural areas (less than 5000 inhabitants), 12% in small towns (5000 inhabitants and more), 34% in medium-sized cities (10,000 inhabitants and more), 28% in large cities (100,000 inhabitants and more), and 8% of the participants in a metropolitan area (1,000,000 inhabitants and more). In order to connect to the usual privacy extent of the participants, household size was also surveyed. On average, participants reported living in a household with about two to three other people ($M = 2.64, SD = 1.23$), one to two of whom are children ($M = 1.42, SD = 0.81$). All participants provided their informed consent at the beginning of the online study. The study was approved by the ethics committee of the Furtwangen University.

2.2. Materials and Procedure

The exploratory study (repeated measurement design) was conducted using an online survey. On average, the participants needed about 19 min to complete the questionnaire. First, the participants were informed about the study and provided their informed consent. Subsequently, the subjects received an introduction to the smart video surveillance system, whereby no technical knowledge was required to answer the questions. Overall, the questionnaire consisted of three parts: the assessment of usefulness of smart video surveillance using digital skeletons, questions recording the perceived privacy risk of the system, as well as the perceived safety. After that, the participants were asked about socio-demographic information. Finally, there was a farewell.

The following independent variables (IVs) are used. IV1: gender (male, female), as well as the repeated measures IV2: time of day (day, night), IV3: public area (see Table 1) and IV4: data privacy levels of a smart video surveillance system (exclusive analysis of body movements using digital skeleton, exclusive analysis of possible crime weapons, analysis of video recordings, but censoring of biometric features (e.g., faces), exclusive analysis of biometric features (e.g., facial recognition)).

Table 1. Overview of public areas and their acronyms. The entries are shown column-wise in alphabetical order.

(AP) Amusement Parks	(LE) Large Events	(SM) Shopping Malls
(AR) Airports	(LS) Low-Traffic Sidewalks	(SS) Shopping Streets
(AS) Areas Surrounding Schools	(PG) Parking Garages	(TS) Train Stations
(AW) Areas Surrounding Workplace	(PL) Parking Lots	
(CS) City Squares	(PLG) Playgrounds	

The dependent variables (DVs) are described as follows. DV1: perceived usefulness. Participants indicated the extent to which they would perceive the smart video surveillance useful, using a 5-point scale ranging from not at all useful (1) to exceptionally useful (5). DV2: perceived privacy risk. The following scenario was described to the participants and rated for different data privacy levels (see IV4) with a 5-point scale, ranging from not at all risky (1) to exceptionally risky (5): The evaluated method (see IV4) detected a crime. Based on this suspicion, police officers shall be sent to the crime scene. Therefor, the data are only accessible at the time of video recording and are analyzed in real time.

2.3. Statistical Analyses

The software IBM SPSS Statistics was used for the statistical analysis. As statistical procedure, *t*-tests, $2 \times 2 \times 13$ and 2×4 ANOVA (analysis of variance) with repeated measures were used. The evaluation was based on a significance level of 5%.

3. Results

3.1. Perceived Usefulness

To answer the first research question if gender differences, differences between public areas and the time of day regarding perceived usefulness of the smart video surveillance system could be shown, and an ANOVA with repeated measures was analyzed. The results are shown in Table 2.

Table 2. Results of ANOVA with repeated measures.

	F	df	df_{error}	p	η^2_{part}
Public area × time of day	47.03	7.79	1666.25	≤0.0001	0.180
Public area × gender	2.67	12	2568	0.001	0.012
Public area	110.83	7.63	1631.83	≤0.0001	0.341
Time of day	5.73	1	214	0.018	0.026
Gender	18.46	1	214	≤0.0001	0.079

Note: If sphericity is not assumed, the correction by Greenhouse–Geisser is applied.

The following Figure 3 shows the significant interaction public area × time of day. The results of the post hoc analyses (Bonferroni) are presented in Figure 3, as well as in Table 3. The subsequent Figure 4 displays the significant interaction public area × gender. The results of the post hoc analyses (Bonferroni) are represented in Figure 4 and Table 4.
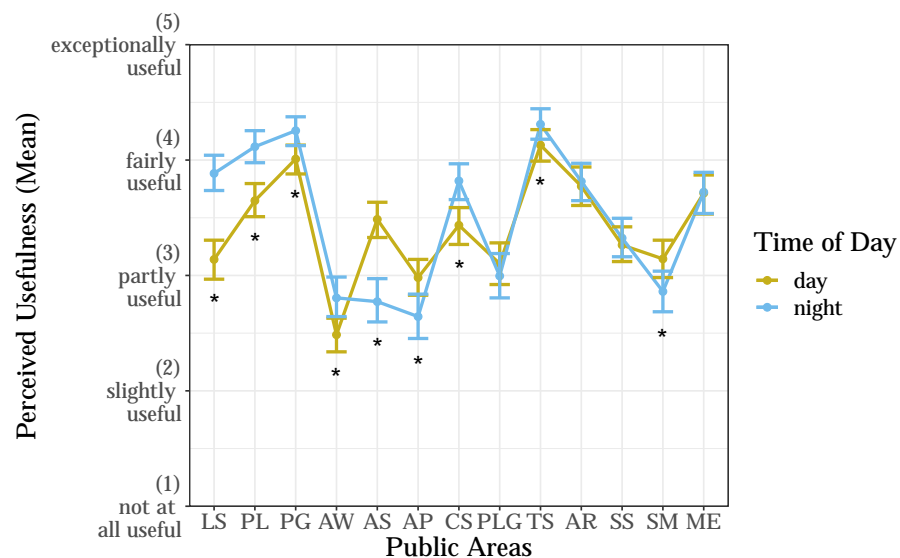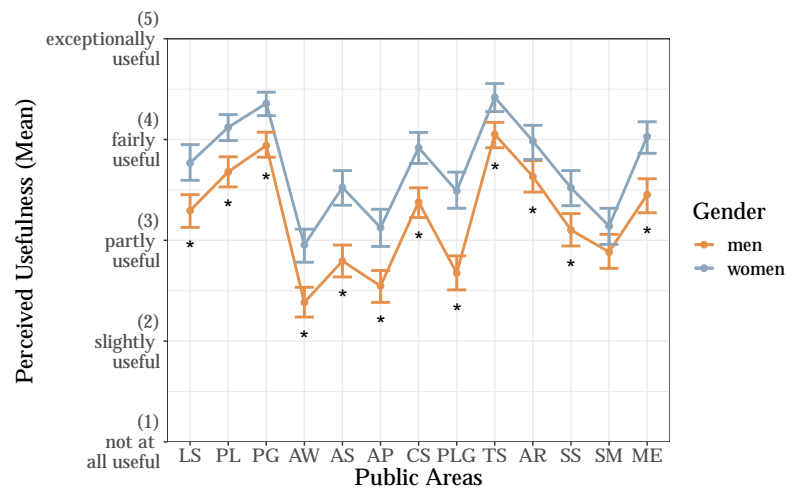

Figure 3. Interaction public area × time of day. Note: Significant results are marked with asterisks (*), $\rho \leq 0.05$. Standard error of means are represented.

Table 3. Results of interaction public area × time of day.

	PL	PG	AW	AS	AP	CS	PLG	TS	AR	SS	SM	ME
LS	3	3	3	3	1	2	2	3	3		2	1
PL		3	3	2	2	3	3	3	2	3	3	2
PG			3	3	3	3	3			3	3	3
AW				1	3	1	1	3	3	3	1	3
AS					2	1	1	3	2	2	1	2
AP						3	3	3	3		3	1
CS							2	3	3	3		3
PLG								3	3	2		3
TS									3	3	3	3
AR										3	3	
SS											2	3
SM												3

Note: Only significant ($\rho \leq 0.05$) results are presented for each group by mapping the following numbers: (1) only day, (2) only night, (3) both groups day and night.

Figure 4. Interaction public area × gender. Note: Significant results are marked with asterisks (*), $\rho \leq 0.05$. Standard error of means are represented.

Table 4. Results of interaction public area × gender.

|	PL	PG	AW	AS	AP	CS	PLG	TS	AR	SS	SM	ME
LS	3	3	3	1		3	1	3			3	
PL		3	3	3	1	3	3	3		3	3	
PG			3	3	3	3	3		2	3	3	
AW				3	2		2	3	3	3	1	3
AS					3	2		3	3	1	2	3
AP						3	3	3		3		
CS							2	3	3	3	1	3
PLG								3	3	1	2	3
TS									3	3	3	3
AR										3	3	
SS											3	3
SM												3

Note: Only significant ($\rho \leq 0.05$) results are presented for each group by mapping the following numbers: (1) only men, (2) only women, (3) both groups men and women.

3.2. Perceived Privacy Risk

An ANOVA with repeated measures was conducted to answer the second research question. This aimed to determine whether there are differences between gender and privacy levels of smart video surveillance systems in terms of perceived privacy risk.

The results show a significant effect of privacy levels of a smart video surveillance system, $F_{GG}(2.58, 551.48) = 216.17, \rho \leq 0.0001, \eta^2_{part} = 0.503$ (see Figure 5). No significant interaction gender × privacy levels can be shown, $F(3, 642) = 2.29, \rho = 0.077, \eta^2_{part} = 0.011$. Furthermore, a significant effect of gender can be shown, $F(1, 214) = 11.19, \rho = 0.001, \eta^2_{part} = 0.050$. According to the results of post hoc analyses (Bonferroni) the perceived privacy risk of the exclusive analysis of body movements using digital skeletons and of the exclusive analysis of possible crime weapons are estimated to be significantly lower than the perceived privacy risk of the analysis of biometric features (both: $\rho \leq 0.0001$). The privacy risks of the exclusive analysis of body movements using digital skeletons and the exclusive analysis of possible crime weapons are also perceived to be significantly lower than that of the analysis of video recordings but censoring of biometric features (both: $\rho \leq 0.0001$). Furthermore, the analysis of video recordings but censoring of biometric features is assessed to be significantly less risky than the analysis of biometric features ($\rho \leq 0.0001$). Other pairwise comparisons do not reach the level of significance of 5%. Additionally, women perceive the risk to their own privacy as significantly lower ($M = 3.45, SD = 0.88$) than men ($M = 3.05, SD = 0.81$).
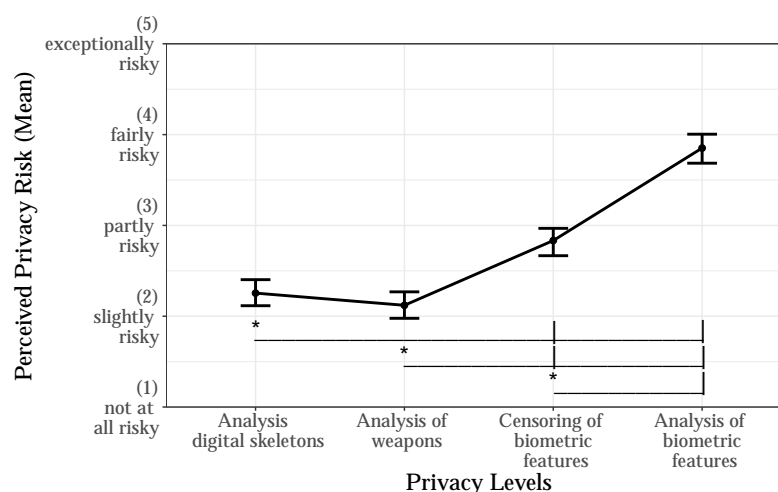
Figure 5. Privacy levels and perceived privacy risk. Note: Significant results are marked with asterisks (*), $\rho \leq 0.05$. Standard error of means are represented.

4. Discussion

The results of the first research question showed that there were significant gender differences as well as differences between public areas and the time of day in terms of perceived usefulness of smart video surveillance using digital skeletons. Furthermore, there were significant interactions for public area × time of day and public area × gender. The presented results of the second research question showed that different privacy levels of smart video surveillance differed significantly regarding the perceived privacy risk. In addition, a significant effect of gender was shown in terms of the perceived privacy risk.

The interactions public area × time of day and public area × gender were new results shown in the presented study. It can be mentioned that previous discoveries are essentially consistent with the presented results, such as the significant main effects. The perceived usefulness of the surveillance system for different public areas changed depending on the time of day. The controversial effect of time of day in the different public areas could be due to core opening or visiting hours (see Figure 3 and Table 3). For example, for the areas surrounding schools, amusement parks, as well as shopping malls, the usefulness was rated significantly lower at night than during the day. A possible explanation for this could be that these public areas are either not accessible at night or are not frequently visited at night by those potentially affected. Accordingly, the use of the system in these areas would not be appropriate. Therefore, the smart video surveillance appears to have a lower perceived usefulness at these locations at night than during the day. Krempels' [6]

suggestion that perceived usefulness has the purpose to increase the security of those that are potentially affected by the system can therefore be confirmed. Roose's [5] results from 2021, regarding the sense of safety in relation to public areas and time of day, are congruent with our results. In terms of feelings of safety, there were descriptive differences between areas and time of day (stops were rated most unsafe at night). In the presented study, the public areas where the usefulness of smart video surveillance was perceived to be high were also areas with less public traffic. This is consistent with the result of Forster et al. [44] that the feeling of safety can be increased in deserted, public areas. This includes areas where people are more likely to be alone.

With regard to the interaction public area × gender, it seems that women perceived smart video surveillance more useful in public areas where people are only occasionally present (see Figure 4 and Table 4). Men, on the other hand, apparently rated the usefulness of this digital skeleton surveillance higher in places associated with many people. For example, women rated the system's usefulness in parking lots and areas surrounding schools significantly higher than in shopping malls. Contrary to the assumption that women rate the usefulness of the system higher in less visited areas compared with more visited areas is, they perceived it to be significantly more useful, for example, in city squares than in playgrounds and areas surrounding schools. Men, in turn, perceived smart video surveillance using digital skeletons significantly more useful on shopping streets than in playgrounds and areas surrounding schools. Therefore, it seems that the higher the crowdedness in a certain area, the higher men rated the usefulness of the system. One possible explanation for why women would perceive smart video surveillance as more useful in less busy public areas could be that they may associate these locations with a higher incidence of crime against women or children. On the other hand, one reason why men would perceive video surveillance as more useful in busy public areas could be that group-based harassment, theft or fights are possibly considered as more likely to occur in these places. For example, in big cities on weekends, it could happen that groups of drunk people are hanging out on shopping streets or in city squares. This potential explanation would also be an interesting approach for future research. The discovery of the significant interactions opens a deeper understanding of how people perceive a smart video surveillance system in different public areas. With this understanding, the use of smart video surveillance could be adapted in a targeted manner, especially after further research of these interactions. An example for a targeted adaption of the system is to deploy the system in certain low-traffic as well as high-traffic public areas. In addition, education about its use in low-traffic public areas could be designed to be appealing to women, and in high-traffic public areas, the information could be designed to be appealing to men.

As already mentioned, it was shown that the effects of public areas and gender were also significant, regarding perceived usefulness. The context dependence of the acceptance of the use of conventional video surveillance reported by Klauser [42] is consistent with the results of the presented study that the usefulness of the smart video surveillance is perceived differently in different public areas. This also supports the results of Hölscher [41], which showed that the usefulness of video surveillance is reported higher in train stations and playgrounds than in parks and workplaces. The significant effect of gender in the presented study indicates that women generally perceived the usefulness to be higher than men did. This finding is also consistent with previous research in which females are more supportive of conventional video surveillance systems than males [42,45,46]. One possible explanation for this result could be that women also rate fear of crime in the future significantly higher than men (see Section 2.1). As mentioned above, the results of this study could be used for educational purposes. For example, certain information could be explicitly addressed to women. In the best case, this would make them feel more safe in public places.

The results of the perceived privacy risk of different privacy levels showed interesting effects. The assessment of privacy risk regarding the analysis of digital skeletons did not differ significantly from the perceived risk of analyzing crime weapons. Both were classified

from the participants as slightly risky. In addition, both showed significant differences in perceived privacy risk compared to the analysis of biometric features and to analysis that censored them. These results support the work of Krempel [6]. He argued that smart video surveillance provides a selective surveillance capability that limits privacy invasion. Within the presented study, men perceived the risk to their own privacy significantly higher than women, regardless of the privacy levels, which is consistent with previous findings [42,45,46]. These previous studies showed that men are generally more critical of video surveillance and more often express concerns related to their own privacy. The fact that men in this sample showed significantly lower trust in the government to avoid potential privacy risks than women could be a possible factor for this result (see Section 2.1). Interestingly, regarding the perceived privacy risk, no significant interaction gender x privacy levels can be shown. In addition, more men in the study sample paid attention to installed cameras in their environment. The fact that men were significantly more attentive to surveillance cameras could be interpreted in a way that they were more aware of them because of the higher perceived risk compared to women. Moreover, it should be noted that the variable affinity for technology is not distributed equally by gender. Maybe this could be seen as one explanation for gender differences regarding the perceived privacy risk. Men were significantly more tech-savvy than women. It is possible that they are thus accustomed to viewing the privacy risk more critically, e.g., regarding data access. Overall, this discovery could enrich educational efforts by addressing potential privacy risks and prevention measures. Because men perceive this risk to be higher, educational outreach could be designed and placed for men.

For further research, it would be interesting to investigate why there is a gender difference but no interaction with privacy levels. Furthermore, it should be investigated whether a correlation between affinity for technology or trust in (new) technologies and perceived privacy risk can be shown. It would also be interesting for further research to investigate whether the influence of time of day could be interpreted in more detail by measuring the public areas separately according to the times of visit. In addition, maybe more accurate conclusions about the effect of time of day could be made if a division were made not only by day and night but also by evening. Furthermore, other public areas such as green parks should be considered in future research. There is also the question for further analyses regarding if or to what extent age has an impact on the perceived usefulness of the system. Klauser [42] previously reported an effect of age with respect to advocacy of conventional video surveillance. In addition, another interesting research approach would be to compare multiple methods of smart video surveillance in terms of perception, not just perceived privacy risk.

Limitations

The study has some limitations. First, the sample studied was on average quite tech-savvy, and over 50% had an academic degree, so the sample may not represent the entire range of the population. Regarding the results concerning the public areas and the privacy levels, it should be taken into account that the participants did not assess the usefulness of the system in randomized order for the different public areas. Therefore, the answers might have been given in proportion to each other. Further research is needed to investigate whether this affects the perceived usefulness or privacy risk of smart video surveillance using digital skeletons. In addition, the participants did not experience the smart video surveillance directly but rather evaluated it as part of an online study.

5. Conclusions

The presented study examined two research questions. The first research question investigated whether there are differences between genders, public areas or time of day with regard to the perceived usefulness of smart video surveillance using digital skeletons. The second one investigated whether there are differences between genders and different privacy levels in terms of perceived privacy risk. Examining the first research question

showed the effects of gender, time of day and public areas. Furthermore, the interactions public area × time of day and public area × gender were also significant. Regarding the second research question, the effects of gender and privacy levels were shown. Previous research has only reported differences for the perceived acceptance or usefulness of video surveillance, regarding gender [45,46], times of day [5] or different areas [41]. The results of the presented study confirmed these effects; furthermore, they showed that the interactions public area × time of day and public area × gender were also significant in terms of perceived usefulness of smart video surveillance using digital skeletons. One interpretation for the first interaction could be seen in the common visiting hours of the analyzed areas. Perceived usefulness decreased at night in public areas which are not normally visited at night. A possible explanation of the second interaction could be that women perceived the investigated surveillance to be more useful in low-traffic places and men in high-traffic places. As mentioned earlier, this is the first time these interactions have been studied and found to be significant, so there is no previous literature to interpret them. Thus, further research is needed to provide more accurate conclusions about them. Overall, smart video surveillance of the digital skeletons can be seen as a promising solution for increasing security while reducing privacy risk. Based on the presented results, it is possible to provide targeted education on potential applications of surveillance. Furthermore, these results can help to better understand the factors influencing the acceptance of smart video surveillance.

References

1. Comparitech. Surveillance Camera Statistics: Which Cities Have the Most CCTV Cameras? Available online: https://www.comparitech.com/vpn-privacy/the-worlds-most-surveilled-cities/ (accessed on 23 May 2022).
2. Golda, T. Image-based Anomaly Detection within Crowds. In *Proceedings of the 2018 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory*; Beyerer, J., Taphanel, M., Eds.; Karlsruher Schriften zur Anthropomatik/Lehrstuhl für Interaktive Echtzeitsysteme, Karlsruher Institut für Technologie; Fraunhofer-Inst. für Optronik, Systemtechnik und Bildauswertung IOSB Karlsruhe; KIT Scientific Publishing: Karlsruhe, Germany, 2019; Volume 40, pp. 11–24.
3. Weirauch, B. Stellungnahme des Minesteriums für Inneres, Digitalisierung und Migration. Einjährige Bilanz zur Novellierung des Polizeigesetzes Baden-Württemberg. Drucksache 16 / 8128 [Statement of the Ministry for Internal Affairs, Digitization and Migration. Annual Balance on the Amendment of the Police Law in Baden-Württemberg. document 16 / 8128]. Available online: https://www.landtag-bw.de/files/live/sites/LTBW/files/dokumente/WP16/Drucksachen/8000/16_8128_D.pdf (accessed on 23 January 2022).
4. Commission, T.E. Migration and Home Affairs. Responding to Terror Attacks. Available online: https://ec.europa.eu/home-affairs/pages/page/responding-terror-attacks_en (accessed on 23 January 2022).
5. Roose, J. Wenn es Nacht wird in Deutschland. Ergebnisse einer repräsentativen Umfrage zu Kriminalitätsangst und der Akzeptanz von Maßnahmen gegen Kriminalität [When night falls in germany. Results from a representative Survey on Fear of Crime and the Acceptance of Actions against Crime]. *Monitor Sicherheit*, 25 October 2021.
6. Krempel, E.L. *Steigerung der Akzeptanz von Intelligenter Videoüberwachung in öffentlichen Räumen [Augmentation of Acceptance of intelligent Video Surveillance in Public Spaces]*; KIT Scientific Publishing: Karlsruhe, Germany, 2017; Volume 28.

7. Hempel, L.; Toepfer, E. *On the Threshold to Urban Panopticon? Cities and Assessing its Social and Political Impacts. Cctv in European Cities and Assessing Its Social and Political Impacts*; Working Paper no. 15; TU Berlin Centre for Technology and Society Technical: Berlin, Germany, 2004.

8. Senior, A.; Pankanti, S.; Hampapur, A.; Brown, L.; Tian, Y.L.; Ekin, A.; Connell, J.; Shu, C.F.; Lu, M. Enabling Video Privacy through Computer Vision. *IEEE Secur. Priv.* **2005**, *3*, 50–57. [CrossRef]

9. Birnstill, P.; Ren, D.; Beyerer, J. A user study on anonymization techniques for smart video surveillance. In Proceedings of the 12th International Conference on Advanced Video and Signal Based Surveillance (AVSS), Karlsruhe, Germany, 25–28 August 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–6. [CrossRef]

10. Fraunhofer IOSB. Intelligent Video Surveillance Enhances Safety and Privacy. Available online: https://www.iosb.fraunhofer. de/en/projects-and-products/intelligent-video-surveillance.html (accessed on 5 October 2022).

11. Senior, A.; Pankanti, S.; Hampapur, A.; Brown, L.; Tian, Y.L.; Ekin, A.; Connell, J.; Shu, C.F.; Lu, M. *Blinkering Surveillance: Enabling Video Privacy through Computer Vision*; Technical Report; IBM: Armonk, NY, USA, 2003.

12. Markovitz, A.; Sharir, G.; Friedman, I.; Zelnik-Manor, L.; Avidan, S. Graph Embedded Pose Clustering for Anomaly Detection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 10536–10544. [CrossRef]

13. Morais, R.; Le, V.; Tran, T.; Saha, B.; Mansour, M.; Venkatesh, S. Learning Regularity in Skeleton Trajectories for Anomaly Detection in Videos. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 11988–11996. [CrossRef]

14. Krempel, E.; Beyerer, J. TAM-VS: A Technology Acceptance Model for Video Surveillance. In *Privacy Technologies and Policy, Proceedings of the Second Annual Privacy Forum, APF 2014, Athens, Greece, 20–21 May 2014*; Preneel, B., Ikonomou, D., Eds.; Springer: Cham, Switzerland, 2014. [CrossRef]

15. Qi, Y.; Qin, L.; Zhang, S.; Huang, Q.; Yao, H. Robust visual tracking via scale-and-state-awareness. *Neurocomputing* **2019**, *329*, 75–85. [CrossRef]

16. Yu, H.; Li, G.; Zhang, W.; Huang, Q.; Du, D.; Tian, Q.; Sebe, N. The unmanned aerial vehicle benchmark: Object detection, tracking and baseline. *Int. J. Comput. Vis.* **2020**, *128*, 1141–1159. [CrossRef]

17. Wen, L.; Du, D.; Zhu, P.; Hu, Q.; Wang, Q.; Bo, L.; Lyu, S. Detection, tracking, and counting meets drones in crowds: A benchmark. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; IEEE: Piscataway, NJ, USA, 2021.

18. Qi, Y.; Wu, Q.; Anderson, P.; Wang, X.; Wang, W.Y.; Shen, C.; van den Hengel, A. REVERIE: Remote embodied visual referring expression in real indoor environments. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; IEEE: Piscataway, NJ, USA, 2020.

19. Qi, Y.; Zhang, S.; Jiang, F.; Zhou, H.; Tao, D.; Li, X. Siamese local and global networks for robust face tracking. *IEEE Trans. Image Process.* **2020**, *29*, 9152–9164. [CrossRef] [PubMed]

20. Meng, Q.; Zhao, S.; Huang, Z.; Zhou, F. MagFace: A universal representation for face recognition and quality assessment. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; IEEE: Piscataway, NJ, USA, 2021.

21. Specker, A.; Schumann, A.; Beyerer, J. An evaluation of design choices for pedestrian attribute recognition in video. In Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 25–28 October 2020; IEEE: Piscataway, NJ, USA, 2020.

22. Specker, A.; Stadler, D.; Florin, L.; Beyerer, J. An occlusion-aware multi-target multi-camera tracking system. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Virtual Conference, 19–25 June 2021; IEEE: Piscataway, NJ, USA, 2021.

23. Jiang, S.; Qi, Y.; Zhang, H.; Bai, Z.; Lu, X.; Wang, P. D3D: Dual 3-D Convolutional Network for Real-Time Action Recognition. *IEEE Trans. Industr. Inform.* **2021**, *17*, 4584–4593. [CrossRef]

24. Yang, F.; Wu, Y.; Sakti, S.; Nakamura, S. Make skeleton-based action recognition model smaller, faster and better. In *Proceedings of the ACM Multimedia Asia*; ACM: New York, NY, USA, 2019.

25. European Commission. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA Relevance), 2016. Available online: https://eur-lex.europa.eu/eli/reg/2016/679/oj (accessed on 5 October 2022).

26. European Commission. Commission Implementing Regulation (EU) 2019/947 of 24 May 2019 on the Rules and Procedures for the Operation of Unmanned Aircraft (Text with EEA Relevance.), 2019. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019R0947 (accessed on 5 October 2022).

27. Pūraitė, A.; Šilinskė, N. Privacy Protection in the New EU Regulations on the use of unmanned aerial systems. *Public Secur. Public Order* **2020**, *24*, 173–183.

28. Kohl, P.; Specker, A.; Schumann, A.; Beyerer, J. The MTA dataset for multi target multi camera pedestrian tracking by weighted distance aggregation. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; IEEE: Piscataway, NJ, USA, 2020.

29. Golda, T.; Kalb, T.; Schumann, A.; Beyerer, J. Human pose estimation for real-world crowded scenarios. In Proceedings of the 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Taipei, Taiwan, 18–21 September 2019; IEEE: Piscataway, NJ, USA, 2019.

30. Golda, T.; Blattmann, A.; Metzler, J.; Beyerer, J. Image domain adaption of simulated data for human pose estimation. In *Proceedings of the Artificial Intelligence and Machine Learning in Defense Applications II*; Dijk, J., Ed.; SPIE: Bellingham, WA, USA, 2020.

31. Liu, C.; Fu, R.; Li, Y.; Gao, Y.; Shi, L.; Li, W. A Self-Attention Augmented Graph Convolutional Clustering Networks for Skeleton-Based Video Anomaly Behavior Detection. *Appl. Sci.* **2022**, *12*, 4. [CrossRef]

32. Zeng, X.; Jiang, Y.; Ding, W.; Li, H.; Hao, Y.; Qiu, Z. A Hierarchical Spatio-Temporal Graph Convolutional Neural Network for Anomaly Detection in Videos. *IEEE Trans. Circ. Syst. Video Technol.* **2021**, *1*. [CrossRef]

33. IOSB, F. Öffentliche Sicherheit: Innenminister Besucht Fraunhofer IOSB [Public Security: Minister for Inner Affairs Visits Fraunhofer IOSB]. Available online: https://www.iosb.fraunhofer.de/de/presse/presseinformationen/2018/privatsphaere-und-datenschutz.html (accessed on 13 October 2022).

34. Golda, T. Part Affinity Field based Activity Recognition. In *Proceedings of the 2019 Joint Workshop of Fraunhofer IOSB and Institute for Anthropomatics, Vision and Fusion Laboratory*; Beyerer, J., Zander, T., Eds.; Karlsruher Schriften zur Anthropomatik/Lehrstuhl für Interaktive Echtzeitsysteme, Karlsruher Institut für Technologie; Fraunhofer-Inst. für Optronik, System und Bildauswertung IOSB Karlsruhe; KIT Scientific Publishing: Karlsruhe, Germany, 2020; Volume 45, pp. 53–65.

35. Fang, H.S.; Xie, S.; Tai, Y.W.; Lu, C. RMPE: Regional Multi-person Pose Estimation. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2353–2362. [CrossRef]

36. Iqbal, U.; Milan, A.; Gall, J. PoseTrack: Joint Multi-person Pose Estimation and Tracking. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4654–4663. [CrossRef]

37. Munea, T.L.; Jembre, Y.Z.; Weldegebriel, H.T.; Chen, L.; Huang, C.; Yang, C. The Progress of Human Pose Estimation: A Survey and Taxonomy of Models Applied in 2D Human Pose Estimation. *IEEE Access* **2020**, *8*, 133330–133348. [CrossRef]

38. Davis, F.D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* **1989**, *13*, 319–340. [CrossRef]

39. Spiekermann, S. *User Control in Ubiquitous Computing: Design Alternatives and User Acceptance*; Shaker Verlag: Herzogenrat, Germany, 2008.

40. Golda, T.; Cormier, M.; Beyerer, J. Intelligente Bild- und Videoauswertung für die Sicherheit [Intelligent Image and Video Analysis for Security Applications]. In *Handbuch Polizeimanagement: Polizeipolitik–Polizeiwissenschaft–Polizeipraxis*; Wehe, D., Siller, H., Eds.; Springer Fachmedien: Wiesbaden, Germany, 2022; pp. 1–21. [CrossRef]

41. Hoelscher, M. Sicherheitsgefühl und Überwachung. Eine empirische Studie zu Einstellungen der Bürger zur Videoüberwachung und ihrer Erklärung [Security Sense and Surveillance—An Empirical Study of Citizens' Attitudes Toward Video Surveillance and its Explanation]. *Kriminol. J.* **2003**, *35*, 42–56.

42. Klauser, F.R. *Die Videoüberwachung öffentlicher Räume: Zur Ambivalenz eines Instruments Sozialer Kontrolle [Video Surveillance of Public Spaces: On the Ambivalence of an Instrument of Social Control]*; Campus: Frankfurt am Main, Germany, 2006.

43. Kudlacek, D. *Akzeptanz von Videoüberwachung—Eine Sozialwissenschaftliche Untersuchung Technischer Sicherheitsmaßnahmen [Acceptance of Video Surveillance—A Social Scientific Investigation of Technical Security Measures]*; Springer: Wiesbaden, Germany, 2015.

44. Forster, M.; Huber, M.E.; Wüster, A. Subjektives Sicherheitsgefühl und Überwachung [Subjective sense of security and monitoring]. In *Komfortsignatur mit der Health Professional Card*; Donau Universität Krems: Krems, Austria, 2008

45. Bornewasser, M.; Kober, M. Videoüberwachung: Kriminalitätsreduzierung und gezielte Verdrängung. Bericht über eine Evaluationsmaßnahme in der Stadt Luxemburg [Video Surveillance: Crime Reduction and Targeted Displacement. Report on an Evaluation Measure in the City of Luxembourg]. *Forum Kriminalprävention* **2012**, *2*, 34–42.

46. Bornewasser, M.; Schulz, F. *Videoüberwachung öffentlicher Straßen und Plätze, Ergebnisse eines Pilotprojektes in Brandenburg. Ergebnisse der Evaluationsstudie im Land Brandenburg [Video Surveillance of Public Roads and Squares, Results of a Pilot Project in Brandenburg. Results of the Evaluation Study in the State of Brandenburg]*; Bornewasser, M., Classen, C.D., Stolpe, I., Eds.; Verlag für Polizeiwissenschaft: Frankfurt am Main, Germany, 2008.

47. Cavoukian, A. *Privacy by Design. The 7 Foundational Principles: Implementation and Mapping of Fair Information Practices*; Technical Report; Information and Privacy Commissioner of Ontario: Toronto, ON, Canada, 2009.

48. el den Mohamed, M.K.; Taha, A.; Zayed, H.H. Automatic gun detection approach for video surveillance. *Int. J. Sociotechnology Knowl. Dev.* **2020**, *12*, 49–66. [CrossRef]

49. Bretthauer, S. *Intelligente Videoüberwachung: Eine datenschutzrechtliche Analyse unter Berücksichtigung technischer Schutzmaßnahmen [Intelligent Video Surveillance: A Data Protection Analysis Considering Technical Protection Measures]*; Nomos: Grassshut, Germany, 2017; Volume 50.

50. Franke, T.; Attig, C.; Wessel, D. A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *Int. J. Hum.–Comput. Interact.* **2019**, *35*, 456–467. [CrossRef]