

Received January 23, 2020, accepted February 19, 2020, date of publication March 4, 2020, date of current version March 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2978247

Risk Entropy Modeling of Surveillance Camera for Public Security Application

HONGZHOU ZHANG¹, PEIYUE LI^{1,2}, ZHIGUO DU^{1,3}, AND WEI DOU¹

¹School of Police Information Engineering and Cyber Security, People's Public Security University of China, Beijing 100038, China

²School of Computer Science and Engineering, Beihang University, Beijing 100191, China

³School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

Corresponding author: Hongzhou Zhang (zhanghognzhou@ppsuc.edu.cn)

This work was supported in part by the Basic Research Project of Ministry of Public Security of PRC under Grant 2016GABJC01, and in part by the National Key Research and Development Program of China under Grant 2016YFC0801003.

ABSTRACT Surveillance cameras are widely installed at public places around the world, and the video surveillance system plays an un-substitutable role in police work, especially in case investigation. The problem regarding the effectiveness and rationality of the video surveillance system comes into being in terms of its high demand for investment and rising public concern of over-construction potentially. To answer the question, it ought to establish mode and metrics for measuring effectiveness in theory. This article argued that the police video surveillance system is preferably a sensor network than a Physical Protect System (PPS) because its main feature is to provide the police officers with the visual information they need. Once the police cannot receive sufficient information from the system, decisions of public security are given based on limited or misleading information, and there may be some potential risks remained. Such risks of public security are not directly relevant to the integrity and value of the assets but the uncertainty of decision-making, which is different from the one of traditional PPS. In this paper, we proposed an entropy model for measuring the uncertainty based on attributions of video surveillance for law enforcement. Public security risk was divided into three types within the model according to the source of the risk, such as fixed targets (or restricted areas), moving objects, and video information quality. We verified the validity of the model by the simulation experiment of camera field optimization and discussed further work.

INDEX TERMS Police video surveillance, public security, risk entropy, system effectiveness evaluation.

I. INTRODUCTION

Video surveillance system, which plays a vital role in the security area [1], is derived from Closed Circuit Television (CCTV), but the data stream mainly flows from the front-end camera to the control center. It is also called the CCTV system in some literature for this reason. Surveillance cameras were first introduced into Physical Protecting System (PPS) in the field of security to substitute the patrol guard for checking the alarm given by the intrusion detector [2]. Surveillance videos furnished the key clues to identify the suspects and expose their criminal behavior during the investigation process of the 2005 London bombings. It was the first time that governments realized the significance of the video surveillance system to the security of city life. From then on, video surveillance system becomes one of the essential components of

security infrastructures in urban [3]. It obtains a consensus that video surveillance is effective in crime prevention and also in reducing certain crimes to a great extent. According to statistics, robbery, serious assault, and motorcycle theft are the top three types of crime to be monitored and cracked down via video surveillance [3]. For instance, it is recorded that an around 51% reduction lay on the crimes after video surveillance equipped in public places, such as parking lot [4] and street [3], [5].

Governments and the public both pay much more attention to video surveillance system's input-output ratio and its rationality with the widespread application. It is no doubt that more surveillance cameras are expected at public places concerning public security, but the system scale, which is denoted by the number of front-end cameras in many cases, is constrained by limited investment. For another thing, public concern is sharply arousing related to the worries of personal privacy after a vast deployment of surveillance cameras.

The associate editor coordinating the review of this manuscript and approving it for publication was Guitao Cao¹.



FIGURE 1. An illustration of the surveillance system. The sub-system of the network, storage, and analysis systems is typically an IT system.

It is reasonable that the scale of a video surveillance system should be kept to a minimum and achieve an ideal balance between financial investment, public privacy, and public security purpose. In other words, the effectiveness of a video surveillance system should be carefully studied and weighed throughout every phase of the system's construction, operation, and maintenance.

Effectiveness displays a central character of any real application system, especially a video surveillance system that has a vital significance in public security. Generally speaking, system effectiveness refers to a measurement of how the system meets the application requirements under specific conditions.

Video surveillance system has the characteristics of both information system and sensor network, and its effectiveness problem should be explained from both point of views. As shown by Figure 1, Network, Storage, and Analysis sub-systems, which are cycled by dash-line, thus constitute a typical IT system. The hardware of the above portions are mainly universal IT component and equipment, their effectiveness problem has a long history of research and many mature achievements can be introduced into the video surveillance system.

In addition to powerful hardware devices, software, especially the data processing module, plays a crucial role in system performance. For the video monitoring system, it is computer vision (CV) algorithms that center on this task.

Scientists from the CV field make every effort to design algorithms to overcome poor light, object pose changes, and other adverse effects of imaging. The state of art CV algorithm can integrate multiple views image and multiple types of information for target detection and identity recognition, behavior understanding, and many other tasks [6]–[8]. Evaluation of CV algorithms is fruitful, and many academic conferences with significant impact have launched several regular competitions on image/video analysis such as PETS [9], TRECVID [10], PascalVOC [11] and ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [12]. These research works have shown how the image detail affects the performance of the CV algorithm. For example, the automatic face recognition system is widely applied to law enforcement and the business field, NIST has published series evaluation reports [13] to report the technology progress and explain the relationship between face recognition rates and image quality indexes, such as image size, clarity, and compression format. However, their evaluation works, like many other CV evaluations, are based on standard databases, did not take into account the on-site data collection process, although these images are mainly from the field of law enforcement.

There might be an inherent assumption that all video sources were well designed for the CV algorithm and the following applications. Nevertheless, it is not always the truth due to the fact that there are enormous challenges for the imaging procedure, such as the irregular shape of sites, growing plants, and diverse lighting conditions in urban areas. In addition, system construction and maintenance are complex system engineering involving human and social resources.

Wireless sensor network (WSN) is an important research area in recent years. Sensor coverage is a fundamental problem in this field [14], which, to a certain extent, reflects the consideration of effectiveness in the field data collection process. Cameras are one of the directional sensors, and various coverage models are developed for different security application scenarios. In full coverage problems, for example, an angular coverage model with a limited radius is employed to achieve coverage continuously over a given 2D ground plan [15], [16]. In the barrier coverage problem, co-operable cameras are designed to keep some specific connectivity to detect any intruder attempting to cross the border of sensor networks [17].

As mentioned above, it is definitely complex to promise an ideal image capture when cameras are deployed in open space. The requirement of a public security application is far beyond the focal depth and sharpness of the image, and the relationship between risk level and visual information needs to investigate. Moreover, the minimization of risk should be the goal of system optimization. In the above and many other WSN research works regard the object as a point, and image content is not the focus of system optimization, only a few parameters, such as focal length and depth of field are used to control the field of view (FoV) [14]–[19]. In addition, the goal of network optimization in current WSN works is always a certain coverage ratio, network lifetime, or a lower bound of data transfer rate.

From the perspective of systems engineering, the best system performance is only achieved with reasonable data collection and excellent data processing algorithms. As discussed above, the current CV and WSN research works did not consider the impact of the data collection factors on the effectiveness of video surveillance systems, especially the relationship between this impact and public security risk is not the central issue. As a result, the problem is left almost blank that to ensure the content of the collected image meets with the requirements of a specific public security task. Therefore, the primary topic of effectiveness is what is the ideal video data collection from a police officer's perspective and how to evaluate it.

Two things must be done in order to complete the task of public security. First, surveillance cameras, the sensor, are deployed in places where suspects may emerge, or where incidents may occur. Second, all technical details of each camera, such as installation parameters and performance parameters, are carefully considered. The former mainly responds to the distribution of crimes or public security

incidents, while the latter needs to consider both characteristics of criminal behavior or security event, and details of image information.

In this paper, we discuss the problem of the relationship between public security risk and detail of visual information and establish a quantitative model to evaluate whether the camera information collection does meet the requirements of public security applications, such as law enforcement. Within the knowledge of the author, it depicts a new picture of research in which only a few researchers have addressed attention [20], [21]. In the rest of this paper, the risk of police video surveillance system is analyzed in Section 2; visual surveillance risk entropy model is further proposed, and the calculation is given in Section 3 in detail; in Section 4, the proposed method is examined on camera FoV optimization, and some further work is discussed in the last section.

II. RISK OF POLICE VIDEO SURVEILLANCE SYSTEM

The typical PPS is designed to protect valuable assets from deliberately destroying or attacking and normally installed inside building or within a certain range out of the building but enclosed by the fence. A PPS with basic functions consists of intrusion detection, access control subsystem, and control center. Building boundaries or perimeter defines the control area or protected area of PPS. Any unauthorized enter or attack can be predefined and detected by the system. The Security risk is relevant to systematical effectiveness can be express as the following equation [2]

$$R = P_A * (1 - P_E) * C \quad (1)$$

where, R denotes risk has a range of 0 to 1.0, with 0 being no risk and 1.0 being a maximum risk. P_A is the probability of an adversary attack during a period. P_E represents the effectiveness of the PPS, and its range is 0 to 1.0. When operating systems, any alarm reported by intrusion detection or access control subsystem is transmitted to the control center and leads to a response of security guard. If the system can resist any attack successfully, P_E is set to 1.0, otherwise set zero. $(1 - P_E)$ represents the vulnerability of the PPS towards the defined threats, and then the product of P_A and $(1 - P_E)$ is the probability of a successful attack. C denotes consequence value. The value, which is from 0 to 1, relates to the severity of the occurrence of the event. It should be stressed that system effectiveness, P_E , is contributed by each subsystem, especially detector, responder, and their tactics. In the middle of the last century, CCTV technology was firstly embodied in PPS to check alarms before dispatching the security guard. From then on, video surveillance becomes a standard option of alarm assessment subsystems for the PPS. It can make the system more robust and reduce vulnerable value $(1 - P_E)$.

When the video surveillance system is brought into open areas, especially in public places, the borders of control and protection area disappeared. Consequently, it is hard to even impossible to define unauthorized access, abnormal events, or behavior. Based on more than ten years of applying history in public security, especially in China, some new

characteristics of video surveillance that is a dedicate system for police can be summarized as follows:

- 1) The video surveillance system is built and operated independently from the response subsystem; in contrast, the surveillance cameras of PPS are closely coupled with the subsystem of detectors and responders.
- 2) Control areas are sensitive to public security or have important political symbolism, but the boundary is ambiguous.
- 3) Pedestrian, vehicle, and security event are on the top of the surveillance list, but only very few of them is related to the crime.
- 4) Video data is mainly used to distill clues for law enforcement, such as case investigation.

The above characteristics are also consistent with previous findings in Britain, which is, the most valuable significance of video surveillance is the advantage or contribution to crime detection than deterrence [5]. It can be concluded that the video surveillance system in public places is rather a visual information sensor network than a PPS from the standpoint of public security. In order to distinguish the system from other applications, we call the system with the above characteristics police video surveillance system.

Therefore, there is an application risk if visual information was insufficient to support police decision-making, it is expressed as follows, in a similar form of Equation (1):

$$H = P * (1 - \tilde{p}) * W \quad (2)$$

In the above equation, each variable's range is 0 to 1.0. H is the remained risk of security decision-making after the application of video surveillance. Similar to Equation (1), P represents the probability of the target, such as a potential suspect, appearing in a certain area during the given period, or the probability of the public security incident. $P = 1$ means that the wanted person is bound to appear or that a certain incident must be happening in the space concerned.

\tilde{p} is the probability representing the situations that the system can provide the desired image data corresponding to certain requirements; it can be quantified mathematically as the ratio or correlation of some image indicators. \tilde{p} is further expressed as the product of two variables: $\tilde{p} = \tilde{p}_c \times \tilde{p}_e$. \tilde{p}_c is the probability that targets or incident locations is covered by the camera's FoV. Typical targets or incidents of police video surveillance are involved in human behavior, that means suspects try to avoid being photographed by surveillance cameras. From the view of whole city space, the high value of \tilde{p}_c means that distribution of camera deployment location holds a very close relevance with the temporal-spatial pattern of street crime what has been studied by experts in the crime prevention field, such as [22], [23]. In a given surveillance site, the camera's coverage determines the value of \tilde{p}_c . As for variable \tilde{p}_e , it describes the degree to which visual information meets police application requirements. Taking face recognition as an example, $\tilde{p}_e = 1$ is the situation that the clarity, size (or the other index) of the face area in the

image fully meets the requirements of a machine or human recognition.

The third factor in the right-side equation, i.e., W , is the weight of visual information in decision-making. Similar to C of Equation (1), W is a normalizing factor and relevant to the severity of the occurrence of wrong decisions. W is set to a high value if the video data is involved in decisions with significant impact. Public security application always needs multiple types of visual information; for example, suspect identification may require both facial and gait images. Normalizing factor W_i is attached to every i -th type visual information \tilde{p}_i , and $\sum W_i = 1$. W needs to be considered separately in different application scenarios.

III. VISUAL SURVEILLANCE RISK ENTROPY

A. THE PROPOSAL OF VISUAL SURVEILLANCE ENTROPY

Entropy is a widely-used measurement to examine the uncertainty of the system both in physics and information science. According to the definition of Shannon's information entropy, the probability of information source in state x_i is $P(x_i)$, and the information entropy, or we say the uncertainty of information source, is a mathematical expectation of self-information.

$$S = E(-\log P(x_i)) \quad (3)$$

Risk entropy of PPS is defined to describe the uncertainty in completing protection tasks. High uncertainty means that protection effectiveness is low. For a given protection node, protection effectiveness provided by protecting the system is affected by many factors. The extent to which each factor affects the protection effectiveness is measured by membership degrees of meeting protective task requirements. Therefore, the risk entropy of a protection node is calculated as the weighted sum of membership degrees of all the factors [24], [25].

For public security applications, the uncertainty in police decision-making is affected by technical details of camera deployment of the video surveillance system. Because the security situation of a specific space remains unchanged over time, and P is set to a constant value in Equation (2). Taking a similar method of [24], [25], we define visual surveillance risk entropy to measure the uncertainty of decision-making in police video surveillance as follows:

$$S = - \sum \omega_i \log R_i \quad (4)$$

In the above equation, R_i is the membership degree of visual information requirement. The bigger membership degree, the better it can meet the requirements of police and leads to less uncertainty in decision-making. In Equation (2), $\tilde{p} = \tilde{p}_c \times \tilde{p}_e$ is determined by the technical details of camera deployment. The former factor, \tilde{p}_c , is mainly relevant to the type of camera coverage, and the later factor, \tilde{p}_e , is mainly relevant to information detail.

On the whole, membership degree R_i is calculated based on the characters of \tilde{p} . Subscript, i , denotes the different

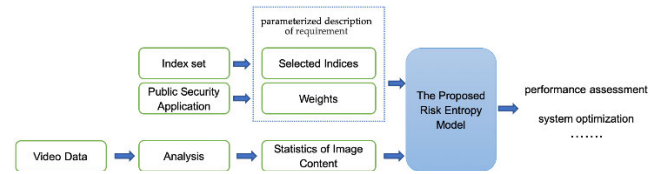


FIGURE 2. A framework of usage of the proposed entropy.

visual information requirements, and ω_i is the i -th weight corresponding to R_i . Wang *et al.* [26] proposed the concept of sensor's information coverage about which only useful information is concerned, and their research shows that WSN optimization benefits a lot from consideration of the coverage quality. Risk entropy defined above can exclude invalid image collection from assessment or optimization of police video surveillance system. The framework of usage of the proposed entropy is illustrated in Figure 2. Calculating the risk entropy for a given location, visual information requirement of police decision-making, and statistics of surveillance image content are inputted into the proposed model. Visual information requirement is comprised of visual information indices selected from the general image evaluation index set, desired values of each index, and corresponding weights. The statistics of image content can be achieved by using the latest state-of-art CV algorithm. The risk entropy value outputted by the proposed model reflects the effectiveness of a given system configuration. For video surveillance evaluation, the installation parameter or the other technical parameters are variable, different system configurations can be compared, and the rank of the system performance under such configurations can be decided according to the risk entropy values.

B. RISK ENTROPY RELATED WITH SURVEILLANCE OBJECTS

Public security incidents or criminal cases always involve three types of surveillance objects: fixed targets, moving targets, and control areas. Fixed targets typically are dangerous articles or symbolic buildings with political or historical significance. They generally have great significance for public security concerning. Moving targets typically are pedestrians and vehicles. Control area locates around the fixed targets or other public places such as plaza, road intersection, any access to the above locations should be taken notice.

Security control and case investigation are typical police business. Generally speaking, police commanders of security controlling need to monitor a particular area for security control, and the overall behavior of moving targets within the corresponding scope should be under their control. Case investigation needs to answer the question of 4-W with the help of video surveillance: who, when, where, and how something was done. To fulfill police business, target detection, tracking, recognition, and situational awareness are relying mainly on video data. Security control and case investigation require differently about the above tasks of video surveillance. Security control stress on situational awareness needs to know the subject's distribution in the whole,

coverage ratio, coverage degree, and density or the number of targets are on the top of the police officer's requirements list. Compared with the above, identification of suspects is the primary task of criminal investigators, detail of the image target and trajectory of the suspects in the city space is on the top of detective's requirements list.

Visual information of different types of surveillance objects involves decision-making with a different extent, which means the different contributions of uncertainty in decision-making.

1) RISK ENTROPY RELATED TO FIXED TARGETS AND AREAS

Coverage is the most critical performance index for sensor networks, and coverage degrees along with coverage ratio are two sensor coverage indices [19]. Coverage degree describes how many sensors cover a point of target, and coverage ratio measures how much area of a sensor field meets the application requirement.

For a given surveillance task, there are fixed targets T_i ($i=1, 2, \dots, n$) and control areas A_j ($j=1, 2, \dots, m$) need to be monitored. The fixed target is looked as a point here, and a certain number of coverage degrees express surveillance requirements. Under this consideration, only targets not covered by any camera contribute to the risk entropy of fixed targets, $S_{1,i}$. For a one-degree coverage request, risk entropy contributed by fixed targets T_i is:

$$S_{1,i} = -\log \frac{\sum R_{s,k} p_k}{\sum R_{s,i} p_i} \quad R_{s,i} \neq 0, \quad p_i \neq 0 \quad (5)$$

p_i ($i=1, 2, 3, \dots$) is the probability of T_i involved in a security incident, such as the dangerous articles explode with probability p_i . $R_{s,i}$ indicates the importance or influence of T_i from the perspective of applying in public security, if there was an incident happened. The subscript, $k = 1, 2, 3, \dots$, denotes targets covered by a surveillance camera in targets T_i .

Similarly, risk entropy contributed by control areas A_j can be expressed by the following:

$$S_{1,j} = -\log \frac{\sum R_{w,j} p_j a_j}{\sum R_{w,j} p_j A_j} \quad R_{w,j} \neq 0, \quad p_j \neq 0, \quad A_j \neq 0 \quad (6)$$

In the above equation, a_j is the size of the intersection of the camera FoV and the j -th control area, $A_j \neq 0$ is the size of the entire area. Typical control areas are the entrance of the square, area where pedestrians gathered as well as any other area where security events may occur because they are important for public security, as mentioned before. p_j is the probability that a security incident happens in area A_j . $R_{w,j}$ similar to $R_{s,i}$, denotes the importance of the control area A_j . If every covered area $a_j = A_j$, Equation (6) degenerates into Equation (5). Non-zero denominator guarantees mathematical significance of the above two equations, and only targets or fields involved in public security incidents are meaningful, that is $R \neq 0$ and $p \neq 0$.

The total risk entropy of video surveillance related to fixed targets and areas is the weighted sum of the above two

portions:

$$S_1 = \sum_i \omega_i S_{1,i} + \sum_j \omega_j S_{1,j} \quad (7)$$

2) RISK ENTROPY RELATED TO MOVING OBJECTS

Visual surveillance risk entropy contributed by moving targets denotes as S_2 . It consists of two parts, S^M and S^{MS} .

S^M is contributed solely by moving targets, such as street protest crowd or vehicles.

$$S^M = -R_w \log(p|P_M) \quad (8)$$

In the above equation, P_M is the probability of the security event, and $(p|P_M)$ is the conditional probability that a moving target is captured by camera FoV with probability p if there was a security incident. A vehicle incidence can explain such a conditional probability. There is a car and the probability of incident P_M ; if it moves into a video surveillance site, the surveillance camera can only capture it with the probability p because the camera coverage is finite. The object's intrinsic attributes (such as density, number, velocity, emergence time, duration time) decides the probability P_M . Generally speaking, p is decided by the distribution of FoV within a given surveillance space. There might be an interaction between p and P_M ; such interaction or dependence is a typical situation that suspect's anti-investigation consciousness or the influence of video surveillance on a person's sense of security. To simplify the problem, we here ignore the interaction between p and P_M , Equation (8) is expressed as:

$$S^M = -R_M(\log p + \log P_M) \quad (9)$$

S^{MS} describes the uncertainty about moving target interacted with a fixed target. An example is an attack on government buildings launched by terrorists or traffic incidents caused by vehicles. Similar to Equation (9), the influence of camera deployment is ignored, and S^{MS} is expressed as follows:

$$\begin{aligned} S^{MS} &= -R_{MS} \log(p|P_{MS}) \\ &= -R_{MS}(\log p + \log P_{MS}) \end{aligned} \quad (10)$$

In the above discussion, only the number of fixed and moving targets is supposed to one for simplicity. If there are several fixed and moving targets, the subscript p and q are used to denote each moving target. The total risk entropy S_2 is summarized as overall targets:

$$S_2 = \sum_{p,q} (S_p^M + S_q^{MS}) \quad (11)$$

3) TIME AND SPACE COVERAGE RATE

Given that the coverage degree and coverage ratio in sensor networks do not take into account the coverage attribute in the time dimension, we argue that it ought to be expanded to the time dimension for all kinds of targets. Pan-tilt-zoom (PTZ) cameras are widely used in video surveillance systems. The camera's FoV varies with the change of imaging directions and focus length. A security guard or automatic program

operates it to perform target inspecting, tracking, and site patrol. The police application always requires the camera to be able to gaze at the target or the region for a certain time duration to capture objects' trajectory or recognize it. Surveillance targets and fields are considered homogeneous in previous research [14], [15]. It is reasonable that take the probability and their influence on public security of each target or area into account. Based on the above facts, traditional coverage degrees and coverage ratios are expanded by the following definition.

Definition: weighted time-spatial coverage ratio is the product of weighted area coverage ratio and time coverage ratio as follows:

$$r = \begin{cases} \frac{\sum R_{s,j} p_j}{\sum R_{s,i} p_i} \times \frac{\Delta t}{T} & \text{for point targets} \\ \frac{\sum R_{w,j} p_j a_j}{\sum R_{w,i} p_i A_i} \times \frac{\Delta t}{T} & \text{for control area} \end{cases} \quad (12)$$

In the above definition, Δt is time duration if targets or control areas covered by FoV and T denote the total time when the target appeared in the surveillance site.

Because r is time and spatial variable, coverage ratio in Equation (6) can be replaced with its mathematical expectation or time-weighted average. Consequently, Equation (5) is rewritten as:

$$S_1 = - \sum \omega_L \log r \quad (13)$$

C. RISK ENTROPY RELATED WITH QUALITY OF VISUAL INFORMATION

Characteristics of video surveillance data quality is another consideration in risk entropy modeling. In the above discussion, both fixed targets and moving targets are treated as point targets, which means that detail of appearance is not considered. In other words, only quantitative characteristics of visual information are included in the model. One of the straightforward requirements of image quality is that the image, especially regions of interest, should be as larger and as clearer as possible. Research on image processing has also put forward a series of indices of image quality from many aspects, such as image size, sharpness, contrast, clarity, color accuracy, etc. Supposing that there were some indices D_k , $k=1, 2, \dots$, and the image quality requirement of a specific task is expressed by each satisfying degree of θ_k with weight ω_k . The third portion of video surveillance risk entropy S_{VQ} , which describes uncertainty in decision-making further decided by visual information quality, is expressed as:

$$S_{VQ} = - \sum_k \omega_k \log (\theta_k | p) = - \sum_k \omega_k (\log \theta_k + \log p) \quad (14)$$

In the above equation, p is the probability of an object captured by the camera's FoV. θ_k is a random variable. Generally, image quality is independent with probability p , and the conditional probability $(\theta_k | p)$ is equal to the product of probability p and satisfying degree θ_k .

D. WEIGHTS OF DIFFERENT PORTIONS OF RISK ENTROPY

So far, three portions of visual surveillance risk entropy are defined, and the total value is the weighted sum of S_1 , S_2 , and S_{VQ} :

$$S = \alpha S_1 + \beta S_2 + \gamma S_{VQ} \quad (15)$$

Weight factors α , β , and γ are relevant to the importance or influence of different aspects of visual information on decision-making. For example, if the surveillance task is just focusing on moving the target, the weight, α , which describes the importance of control areas and the fixed targets, should be set into a value of zero. Additionally, R_S , R_M , and R_{MS} within S_1 and S_2 are weight factors decided by the importance of each target in present applying in public security, ω_k tells the importance of each image quality index in consideration.

As mentioned at the beginning of Section 3.2, different types of police tasks request visual surveillance information differently. It is crucial to set the above weight factors and the other parameters, but this is not a trifle. The reason is that translation between security requirements and parameterized representation is difficult; researchers must handle the different language expressions of police and narrow the understanding gap between them and the police officers. All in all, there is a particular configuration of the above weight factors in computing visual surveillance risk entropy concerning the given task of police video surveillance.

E. A SHORT SUMMARY ON VISUAL SURVEILLANCE RISK ENTROPY

Expansion of filed coverage and image quality of targets always compete with each other in reality of police video surveillance.

There are problems of "cannot see" and "cannot see clearly" in police video surveillance practice. The former refers to some wanted targets that are unable to be covered by the camera, and the latter, or problem of "cannot see clearly" refers to the amount of the detail is insufficient in surveillance video. To cover more targets, the camera's FoV should expand as much as possible, but the targets' region in the image becomes smaller under a given TV resolution. If conversely, it is also true.

In Equation (5) to (12), all targets are looked at as point, more targets, and more portions of control areas covered lead to the small value of S_1 and S_2 . In other words, they represent a quantitative dimension in police application requirements. In Equation (13), S_{VQ} models whether the image quality is satisfied; it represents the quality dimension of police requirement.

The sum of all weights in the above equations are required to one unit to satisfy the mathematical nature of entropy:

$$\alpha \sum (\omega_i R_{s,j} + \omega_j R_{w,j}) + \beta \sum (R_M + R_{MS}) + \gamma \sum \omega_k = 1 \quad (16)$$

An ideal FoV configuration of a camera must balance the requirement in both quality and quantity to attain the

minimum value of the total risk entropy S :

$$\arg \min_{\{\text{FocLength, AzimuthAngle, ...}\}} (\alpha S_1 + \beta S_2 + \gamma S_{VQ}) \quad (17)$$

The subscript denotes parameters of FoV, such as focal length, azimuth angle as well as camera's resolution, they are important for PTZ camera automatic control if the equation was used for camera's FoV optimization.

As far as the author concerned, it is the first time that one model includes the two-conflicted dimensions at the same time. There is no doubt that this attribute of visual surveillance risk entropy certainly brings big convenience to the simulation of coverage optimization.

IV. AN EXAMPLE OF VISUAL SURVEILLANCE RISK ENTROPY APPLICATION

In the above sections, we proposed a risk entropy model for police video surveillance applications. As an illustration of measuring the decision-making uncertainty caused by insufficient visual information support, the proposed entropy model is employed to optimize the FoV of a surveillance camera installed at one corner of the street. Nevertheless, it should acknowledge that the current experiment of FoV optimization is simplified to test the validity of the proposed model. A real optimization of camera control has to deal with many technical details of camera control and the requirements representation as well, which are a bit far beyond the scope of this article.

A. DESCRIPTION OF APPLICATION SCENARIO

As mentioned in the first section of this article, once a site is selected to equip a surveillance camera, coverage detail of the camera's FoV is further determine the efficiency of applying in public security. Hence, it needs to analyze targets' characteristics and task requirements before setting the camera scope of view. As a directional sensor, the coverage of the camera is a compelling character in deployment optimization [19], [20]. Despite the depth of field (DoF), which is the region between near field and far field of acceptable sharpness, it is proposed to calculate the camera's coverage [20]. Still, most researches of camera network take all of the regions between the camera and the acceptable far-field sharpness into account, and system optimization is pursuing a maximum FoV coverage for a site or seamless coverage along curtain path or primer [20], [21]. Because only simple visual information requirements are involved, traditional FoV optimization research for WSN is not suitable for case investigation and similar application scenarios of public security. At the same time, there is a common phenomenon over video surveillance system construction that FoV is set randomly or based on personal subjective aesthetic sense, even though the surveillance site and camera are chosen discreetly. Therefore, the examples we presented in the following has some practical application significance.

In the following experiment, an Ultra-high-definition surveillance camera was installed at the corner of a street.



FIGURE 3. An illustration of the surveillance scenario in this experiment. The camera is installed at one corner of the street, and pedestrians are moving targets concerned. Vehicles are not included in the experiment to simplify the experiment because there are distinctions of projection between the front view and side view.

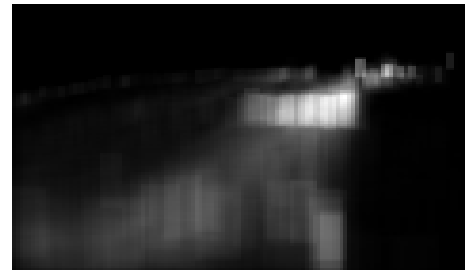


FIGURE 4. Spatial distribution of moving targets, which are labeled out manually. A higher value of the pixels (whiter) represents a higher probability.

The camera type is Hikvision DS-2CD4085F with a 4k image resolution (4096×2160 pixels). A screenshot is presented in Figure 3. Because the camera covers the whole road intersection, and 4K resolution can record more details of targets. There be a space for optimization of FoV if it is replaced by a low-resolution camera, which is cheaper but more widespread.

As shown in Figure 3, surveillance targets are pedestrians, and there is no fixed target considered. The length of the video is 10 minutes, and its content can represent the state of the street intersection. To get the spatial distribution of pedestrians, all moving objects that image size is larger than 15×15 pixels are manually labeled, and there are nearly 300 thousand pedestrians' images region were labeled. The spatial distribution characteristics of targets are illustrated in Figure 4. Vehicles are not included in this experiment because the distinctions of projection between front view and side view are too larger area to calculate spatial distribution accurately from the result of object detection. Beside manually labeled data of targets, a computer vision algorithm [27] is also used to label video data for subsequent comparison. Cross or intersection of roads is always drawing security guard's attention, so in the following experiment, the region was supposed as the control area of surveillance task.

During the simulation experiment, a smaller portion of the original picture is selected and looked as new coverage of camera with lower resolution, which equals that camera's

FoV zooms onto the selected rectangle in a real operation. In principle, parameters of FoV (such as focal length, azimuth angle) can be further calculated based on the theory of geometrical optics.

B. RISK ENTROPY CALCULATION

1) CALCULATION OF S_1

According to Equation (6) and (7):

$$\begin{aligned} S_1 &= \omega_j \cdot S_{1,j} = -\omega_j \cdot \log \frac{\sum R_{w,j} p_j a_j}{\sum R_{w,j} p_j A_j} \\ &= -\omega_j \cdot \log \frac{a}{A} (j = 1) \end{aligned} \quad (18)$$

Because there is only one control area in the current experiment, the first portion of risk entropy is mainly contributed by the ratio of the actual and the desired area. In the above equation, A , the size of pedestrians' crossing, which is the control area, is calculated by pixels of the image region, and a is the corresponding portion of the area covered by the camera's FoV. As discussed in Section 3.2, ω_j is a weight factor decided by the importance or sensitiveness of the intersection of roads in this problem.

2) CALCULATION OF S_2

Most surveillance tasks of public security control are apt to contain targets as many as possible. As a preliminary handling or first-order approximation in the current calculation, a simple but typical surveillance scenario of which interaction between targets is ignored. The second portion of risk entropy, S_2 , calculated only by Equation (8) or (9). Under a constraint of time and space, probability of security events of all moving targets, i.e., P_M is regarded as a constant, so the risk entropy described by Equation (9) is:

$$S_2 = -R_M \log P - \text{constant} \cong -R_M \log \left(\frac{n}{N} \right) \quad (19)$$

$$P = \frac{n}{N} \quad (20)$$

In the current experiment, a smaller portion of the original picture is selected to simulate FoV of optimized camera installation. The probability P of moving targets captured by new FoV is calculated by the ratio of target number under different coverages according to target density distribution, which is shown in Figure 4.

3) CALCULATION OF S_{VQ}

Computer vision research and video surveillance application have proved that the size of the image target is an important index when considering image quality. NIST evaluation shows that the performance of computer face recognition is about to attain its best value when eyes' distance in facial image equals to 60 to 96 pixels [23]. In IEC standard 62676-4 [28], image quality requirements are classified into six categories according to the height of image targets from 5 to 400 pixels. These two image requirements are similar when considering the ratio of height and width of

the human body. In the current experiment, we take the height of the body region as quality requirements, and the satisfying degree function is the following:

$$\theta = 1 - e^{-\frac{w}{D}} \quad (21)$$

in the above equation, w denotes the height of image target in pixels, and D , also in pixels, is the surveillance requirement. According to Equation (13):

$$\begin{aligned} S_{VQ} &= -\omega_k \log(\theta) - \omega_k \log(p) \\ &= -\omega_k \log(\theta) - \text{constant} \end{aligned} \quad (22)$$

C. SIMULATION AND EXPERIMENT RESULTS

The total value of visual surveillance risk entropy is summarized as

$$\begin{aligned} S &= \text{constant} - \alpha \cdot \omega_j \log \left(\frac{a}{A} \right) - \beta \\ &\quad \cdot R_M \log \left(\frac{n}{N} \right) - \gamma \cdot \omega_k \log(\theta) \end{aligned} \quad (23)$$

During the optimizing, coverage of the camera with a lower image resolution of 1920×1080 is set to cover the selected part of the original FoV. Once the risk entropy S reaches its minimum value, optimization is achieved.

$$\min S = \arg \min \left\{ - \left[\alpha \cdot \omega_j \log \left(\frac{a}{A} \right) + \beta \cdot R_M \log \left(\frac{n}{N} \right) + \gamma \cdot \omega_k \log(\theta) \right] \right\} \quad (24)$$

This processing needs to traverse within parameter space, which is a typical optimization solution with a large number of calculations. To simplify the procedure, we take the method used in [29], and pseudo-code is shown in Table 1. The original picture was divided by small grids with the size of 16×9 pixels, a rectangular part of the image made up of adjacent grids is used as a new overlay for each iteration. Three main problems need to solve in optimizing: the start points, optimization direction of the next iteration, and terminating conditions.

1) THE START POINTS

Because original coverage is divided into small grids and each grid is looked at as new coverage of the camera, visual surveillance risk entropy of each grid now can be calculated out. Figure 5 illustrates S value over each grid after the first division is done under four experimental configurations. Then, the start point is chosen as the grids with the minimum value of S .

2) OPTIMIZATION DIRECTION OF THE NEXT ITERATION

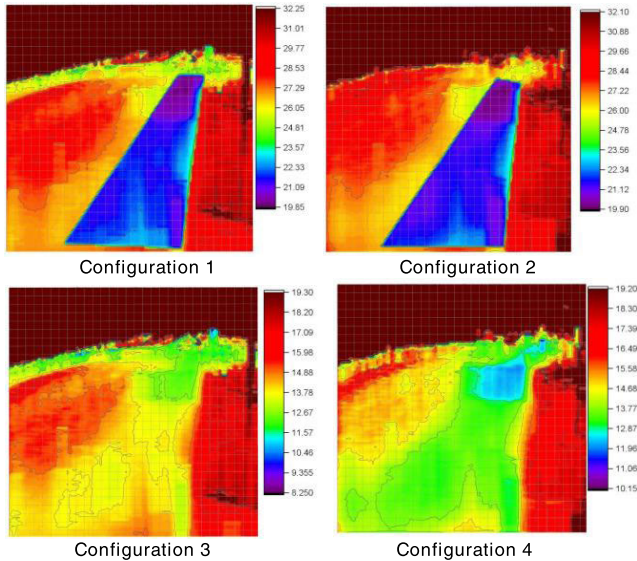
Once the start points or new coverage is selected, the optimizing direction for the next iteration in and adjacent grids in this direction are selected to form a newer coverage. Image coverage selected by the algorithm in each iteration is a rectangle and has eight potential directions in the image of the original FoV. The value S of each potential direction is

TABLE 1. The Algorithm of camera coverage optimizing.

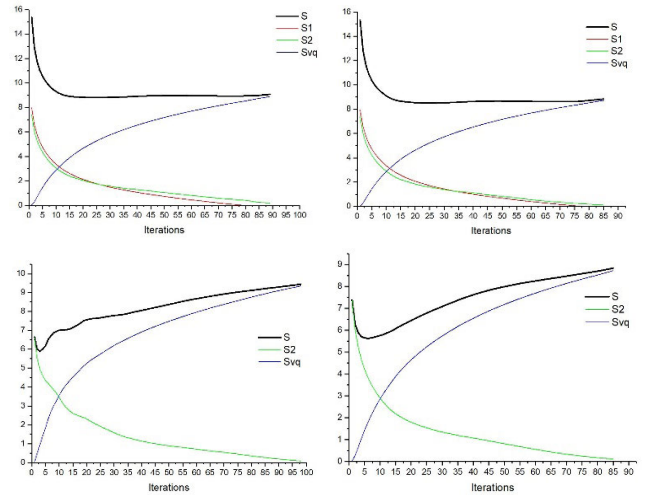
```

Divided image into m grids  $g(i)$  with  $16 \times 9$  pixels;
set(IterationMax)
set(StepLength = 1)
for  $i=1:m$ 
  Calculate  $S(i)$   $i \in g(i)$  // Risk entropy of each grid as coverage
  Iteration = 0
  Sd1 ← empty // main direction of NewView
  Sd2 ← empty // neighbor direction of main direction
  set(NewView =  $g(i)$ )  $i \in \min(s(i))$ 
  TerminateFlag ← False
  S(previousView) ← MaximumValue
  //  $S(*)$  is Visual Surveillance Risk Entropy (VSRE)
  while TerminateFlag = false do
    Compute  $S(\text{NewView})$  // VSRE of NewView as
    coverage
    if  $S(\text{NewView}) < S(\text{previousView})$ 
      S(previousView) =  $S(\text{NewView})$ 
    else
      end while
    end if
    for  $i = 1:8$  //eight neighbor directions of
    NewView
      Compute  $S\_temp(i) = \sum(S(i)) \forall g(i) \in \text{Direction}(i)$ 
      Sd1 =  $i \in \max(S\_temp(i))$ 
      If  $Sd(i+1) > Sd(i-1)$ 
        Sd2 =  $i+1$ 
      else  $Sd2 = i-1$ 
      end if
      set(NewView)
    //execute merge of NewView with its neighbor grids
    //in direction Sd1, Sd2 with StepLength
    Iteration = Iteration + 1
    if Iteration = IterationMax
      TerminateFlag ← True
    end if
  end while

```

**FIGURE 5.** Distribution of the value S over grids. Four sub-figures are corresponding to different configurations in Table 1 from the up-left to the down-right. The color bar on the left side of each subplot indication the value S .

calculated; the rectangle moves towards the direction with the smallest value. The length of the step in each movement is limited within the span of one grid.

**FIGURE 6.** Variation of value S and each portion, $S1$, $S2$, SVQ in iterations. The horizontal ordinate is the number of iteration duration simulation. Four sub-figures are corresponding to different configurations in Table 1 from the up-left to the down-right.

3) TERMINATING CONDITIONS

At each step of the iteration, S is calculated and compared with the value in the last iteration. Once S has a tendency to grow, the terminating conditions are satisfied because S has the meaning of entropy that the smallest value corresponding to minimize uncertainty when the maximum of a satisfying degree of visual information requirements has reached. Certainly, there is the ordinary termination condition that all grids have been selected, or some certain numbers of iterations been reached.

The total number of iterations is less than one hundred in our simulation, and iterations were executed until it went through all grids. In Figure 6, the value of S in each iteration, along with every part of it is shown.

D. RESULT ANALYSIS

1) SIMULATING PARAMETERS AND CONFIGURATION

The primary goal of the current experiment is to check the competence of the proposed entropy model in measuring the effectiveness of the grab of visual information. As we discussed in Section 3.2 and 3.4, it is not a trifle to represent information requirements of police decision-making by quantitative parameters, and such a systematical study is beyond the scope of current research. As listed in Table 2, a typical but straightforward parameter configuration of applying is adopted, taking into account the realness and feasibility.

There are six weights to assign. Each weight in Table 2 is design to adjust the corresponding portion of visual surveillance risk and keep each value, $S1$, $S2$, and SVQ varying in the same numerical range. Therefore, ω_k is assigned the value of 3 and all the other weights equal to one, in this setting. As a result, the value of each portion of the entropy varies in a similar interval, and the result also can be drawn from Figure 6.

The image quality requirement index used in the experiment is the height of image targets, which takes the value

TABLE 2. Simulation configurations and parameters.

Configuration	Moving targets	Pedestrian crossing	Targets data
1	✓	✓	Manual labeled
2	✓		CV algorithm
3	✓	✓	Manual labeled
4	✓		CV algorithm
weights	$\alpha = 0.125, \beta = 0.125, \gamma = 0.125$ $\omega_j = 0.125, R_M = 0.125, \omega_k = 0.375,$		
height of image targets	D=216 pixels		

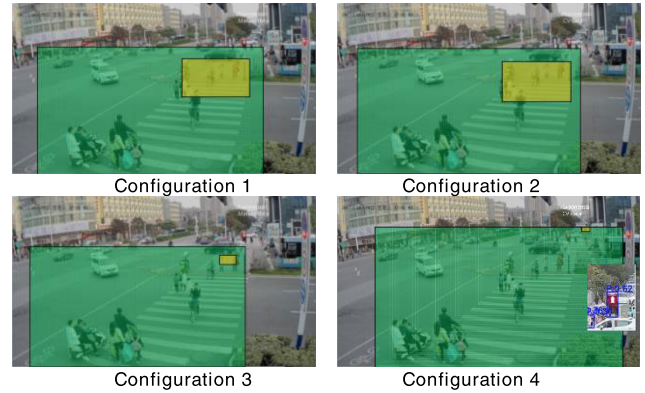
of 216 pixels that is the value recommended by IEC standard 62676-4 [28]. There are six levels of image quality in the standard: monitor, detect, observe, recognize, identify, and inspect recommended by IEC. The value of 216 pixels was classified into recognition levels, by which people can distinguish an individual whether they appeared in the same scenario before with a high degree of certainty.

Two types of surveillance application configurations, security control, and case investigation, were simulated. For the former one, Configuration 1 and 3 in Table 2, both moving targets and control area (pedestrian crossing) are taken into account, and algorithm drives camera to capture pedestrians' image and covers more areas of pedestrians crossing as many as possible at the same time. It is the situation that a police officer keeps a close monitor on the situation of public security for a given site. In such a scenario, a certain coverage ratio of the field can provide the holistically information of security situation and numbers of pedestrian images, which can help to identify the principal offender from the crowd. For the later one, Configuration 2 and 4 in Table 2, only moving targets were considered because more pedestrians' image with higher clarity is promised to achieve case investigation. Although the experiment simplifies the real surveillance scenario, the above two configurations can still reflect the characters of real applying with high confidence.

CV algorithm recognition result is also compared in the experiment because it is popular and more feasible to replace manual working data. With the same parameter configuration, Configuration 2 and 4 adopted a state of art pedestrian detecting algorithm [27] to get objects' spatial distribution, different with Configuration 1 and 3, which use the manually labeled data.

2) ANALYSIS OF OPTIMIZATION RESULTS

Simulation results with corresponding application configuration are illustrated in Figure 7. Each rectangle filled with light yellow is the optimized camera coverage in which the smallest entropy value is obtained, other rectangles filled with

**FIGURE 7.** Simulation result of optimization. The rectangle filled with light yellow is optimized camera coverage in which the smaller entropy value is obtained, the other rectangles filled with light green are the coverage under the maximum iteration number. Four sub-figures are corresponding to different configurations in Table 1 from the up-left to the down-right.

light green are the coverage when the maximum iteration number is reached.

It shows that the proposed risk entropy model, as we expected, can balance different types of information requirements of typical police surveillance tasks. Considering the camera's initial view showed by Figure 3, the entropy calculation of S_1 requires the camera's coverage on the pedestrian crossing; meanwhile, S_2 makes the coverage as large as possible to cover the whole initial view. However, S_{VQ} focuses on the width of the pedestrian's body area in the image, which means a narrow FoV.

As a result of gaming, all of the optimized coverage locates near the far end of the pedestrian crossing, where people gathered and waited for the traffic lights of two directions. If comparing configuration 1 (or 2) and Configuration 3 (or 4), it is found that a more considerable portion of the pedestrian crossing is covered in Configuration 1 and 2 because S_1 requires a possible higher coverage ratio at the pedestrian crossing. Different from S_1 and S_2 , S_{VQ} prevents camera coverage from becoming too large. As mentioned above, optimized coverage, or the rectangle filled with light yellow in the original image in Figure 7, is supposed to be covered by the camera with a resolution of 1920×1080 pixels at the same installing position. The image quality requirement, 216 pixels high image targets, are guaranteed to a large extent. Recognizing image-level given by IEC 62676-4 is a relatively strict requirement, so, the optimized coverage showed in Figure 7 is acceptable.

Different calculations of a satisfying degree affect the result of optimization. The satisfying degree in Equation (22) is an exponential type that simulates human cognition characteristics of intensive variation, such as the perception of sound intensity. Another straightforward choice is a linear function as follows:

$$\theta = \min\left(1, \frac{w}{D}\right) \quad (25)$$

Two functions of satisfying degrees are compared in Figure 8, and the corresponding results of Configuration 1

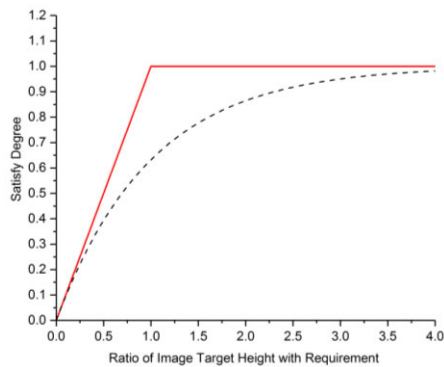


FIGURE 8. Comparison of satisfying degree function. The solid red line denotes the linear satisfaction of Equation 23, and the dark dash line denotes the exponential satisfying function of Equation (21).



FIGURE 9. Comparison of the optimization results of different satisfying degree functions with Configuration 1 in Table 1. The rectangle filled with yellow is the exponential function, and the others are linear functions.

are shown in Figure 9. It can be concluded that the exponential type is more reasonable for a real application.

Comparing simulation results of data labeled by manual and pedestrian detection algorithm, such as Configuration 1 and 2 shown in Figure 7, it is found that there is only a little difference between them. It suggests that the performance of the CV algorithm is fairly well. When considering pedestrian solely in Configuration 3 and 4 in the figure, the difference becomes more distinct. The reason may be that the human-shaped signal light was judged as a pedestrian by the algorithm. A sample of the wrong detection shows in the down-right of the sub-figure (d) in Figure 7, which indicates that the confidence degree of detection is up to 0.62 at that time. The comparison result suggests that the proposed entropy model has the potential to evaluate CV algorithms, and we further discuss in the next section.

The rectangles filled with light green in Figure 7 are the broadest coverage during the simulation procedure. The rule of thirds in photography suggests that the skyline or horizon should be imaged on the upper or lower third of the picture. Apparently, the original FoV of the camera follows the rule to a more significant extent. All optimized outputs, as well as each step of the procedure, avoided going through the areas located in the upper quarter to the third of the original view, the reason is that such area has few targets and the

useful information is scarce. It suggests strongly that the FoV of the surveillance camera should be configured based on the characteristics of the target and control area rather than photographic aesthetics.

It must be pointed out that there are some different considerations in real operation to which we try to make the configurations as close as possible. For example, the pedestrian crossing is treated as the only control area because the other places where the emergence of targets is sporadic, such as the sky and building area. In some rare surveillance scenarios, such areas are required to pay very close attention because the incident is hard to predict, and its consequences are severe. For example, unauthorized access by unmanned aerial vehicles (UAV) may be a hostile attack. In such a situation, each sensitive area should be treated as a control area and assign the appropriate value of $R_{w,j}$ and p_i of Equation (6).

Moreover, the current experiment of FoV optimization is simplified for testing validity of the proposed model, optimization for installation in a real situation of camera control has to deal with many technical details of camera control and the requirements representation as well, which are a bit far beyond the scope of this article.

V. DISCUSSION AND FURTHER WORK

In this article, we put forward a visual surveillance risk entropy model based on relevant analysis of visual information detail and public security risk, especially from the perspective of the police. The calculation of risk entropy is explained and illustrated by the optimization experiment of the configuration of the FoV of surveillance cameras as a preliminary application attempt. The current experiment has plainly shown that the proposed risk entropy model can cope with the matter of effectiveness in police video surveillance.

There are many efforts ought to be made to improve the current work further.

First, the mathematical character of the proposed risk entropy needs to be adequately explored. As mentioned at the beginning, the selection of surveillance sites from a complete view of the city is controversial without the optimization of cameras' FoV. It is worthwhile to explore how visual surveillance risk entropy can be used as a metric of security risk over the whole urban applications on the condition of individual camera's risk entropy. In particular, the calculation of two or more cameras involves the additivity of visual surveillance risk entropy, which can be used to measure the cooperative performance of police surveillance systems. It is noticeable that some pioneering work on measuring the spatial correlation of visual information has been done in the field of multimedia research. A joint information entropy model is proposed to evaluate the coding efficiency in compress videos captured by view-overlapped cameras in [30], and the correlation coefficient and its computing method using imaging geometry.... [31] are also employed to for a similar purpose. Their consideration of view-correlated video is similar to the cooperative operation of the police surveillance systems, which has great reference value for future work.

Second, the composition of metrics set, which is related to the mathematical behavior of visual surveillance risk entropy, is also essential to further research. In the current simulation, the size of the image target is selected as metrics of visual surveillance information quality. Researchers in image processing and computer vision have proposed many indices for different applications, yet actually, not all of them are suitable for police video surveillance. It deserves a further study on the selection of image quality metrics and focusing on the most suitable subset for risk entropy modeling and how it affects the character of risk entropy.

Third, the configuration of weight factors for a particular type of police application needs to be further probed. This issue is mentioned in section 3.4 and 4.4. If we further analyze the results of optimization shown in Figure 7 and Figure 9, it is found that the height of most image targets is larger than 216 pixels in new FoV coverage. Although these optimized FoV meets the image quality requirement well, image quality is somehow overemphasized. It can be reasonably considered that the configuration of weight is relevant to the type of target and the type of police tasks as well. A criterion of weight factor assignment should be dug up deeply to ensure the accuracy of the application.

Fourth, visual surveillance risk entropy has shown some potential abilities in the evaluation of computer vision algorithms. The mainstream of evaluation method [7], [32], [33] focuses on detection accuracy from the view of pixel or region level and does not take the amount of useful information into account. The comparison between manually labeled data and computer vision algorithm output shows the potential application of proposed risk entropy in the evaluation of automatic detection methods at the light of police application requirements.

REFERENCES

- [1] P. Cocca, F. Marciano, and M. Alberti, "Video surveillance systems to enhance occupational safety: A case study," *Saf. Sci.*, vol. 84, pp. 140–148, Apr. 2016, doi: [10.1016/j.ssci.2015.12.005](#).
- [2] M. L. Garcia, *Vulnerability Assessment of Physical Protection Systems*. Oxford, U.K.: Heinemann, 2006.
- [3] M. P. J. Ashby, "The value of CCTV surveillance cameras as an investigative tool: An empirical analysis," *Eur. J. Criminal Policy Res.*, vol. 23, no. 3, pp. 441–459, Apr. 2017, doi: [10.1007/s10610-017-9341-6](#).
- [4] B. C. Welsh, D. P. Farrington, and S. A. Taheri, "Effectiveness and social costs of public area surveillance for crime prevention," *Annu. Rev. Law Social Sci.*, vol. 11, no. 1, pp. 111–130, Nov. 2015, doi: [10.1146/annurev-lawsocsci-120814-121649](#).
- [5] *The Effectiveness of Public Space CCTV: A Review of Recent Published Evidence Regarding the Impact of CCTV on Crime*, Police Community Saf. Directorate Scottish Government, Edinburgh, U.K., 2009.
- [6] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 34, no. 3, pp. 334–352, Aug. 2004, doi: [10.1109/TSMCC.2004.829274](#).
- [7] P. L. Venetianer and H. Deng, "Performance evaluation of an intelligent video surveillance system—A case study," *Comput. Vis. Image Understand.*, vol. 114, no. 11, pp. 1292–1302, Nov. 2010, doi: [10.1016/j.cviu.2010.07.010](#).
- [8] V. Tsakanikas and T. Dagiuklas, "Video surveillance systems-current status and future trends," *Comput. Electr. Eng.*, vol. 70, pp. 736–753, Aug. 2018, doi: [10.1016/j.compeleceng.2017.11.011](#).
- [9] L. Patino, T. Nawaz, T. Cane, and J. Ferryman, "PETS 2017: Dataset and challenge," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 2126–2132, doi: [10.1109/CVPRW.2017.264](#).
- [10] G. Awad, A. Butt, J. Fiscus, D. Joy, A. Delgado, M. Michel, A. F. Smeaton, Y. Graham, W. Kraaij, G. Quénot, M. Eskevich, R. Ordelman, G. J. F. Jones, and B. Huet, (2018). *TRECVID 2017: Evaluating Ad-Hoc and Instance Video Search, Events Detection, Video Captioning, and Hyperlinking*. [Online]. Available: <http://www-nlpir.nist.gov/projects/tvpubs/tv17.papers/tv17overview.pdf>
- [11] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Sep. 2009, doi: [10.1007/s11263-009-0275-4](#).
- [12] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, Jun. 2009, pp. 248–255, doi: [10.1109/CVPR.2009.5206848](#).
- [13] *Face Recognition Vendor Test (FRVT)*. Accessed: Sep. 20, 2018. [Online]. Available: <https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt>
- [14] B. Wang, "Coverage problems in sensor networks," *ACM Comput. Surv.*, vol. 43, no. 4, pp. 1–53, Oct. 2011, doi: [10.1145/1978802.1978811](#).
- [15] E. Yildiz, K. Akkaya, E. Sisikoglu, and M. Y. Sir, "Optimal camera placement for providing angular coverage in wireless video sensor networks," *IEEE Trans. Comput.*, vol. 63, no. 7, pp. 1812–1825, Jul. 2014, doi: [10.1109/TC.2013.45](#).
- [16] T.-W. Sung and C.-S. Yang, "Voronoi-based coverage improvement approach for wireless directional sensor networks," *J. Netw. Comput. Appl.*, vol. 39, pp. 202–213, Mar. 2014, doi: [10.1016/j.jnca.2013.07.003](#).
- [17] D. Tao and T.-Y. Wu, "A survey on barrier coverage problem in directional sensor networks," *IEEE Sensors J.*, vol. 15, no. 2, pp. 876–885, Feb. 2015, doi: [10.1109/JSEN.2014.2310180](#).
- [18] M. Amac Guvensan and A. Gokhan Yavuz, "On coverage issues in directional sensor networks: A survey," *Ad Hoc Netw.*, vol. 9, no. 7, pp. 1238–1255, Sep. 2011, doi: [10.1016/j.adhoc.2011.02.003](#).
- [19] A. Mavrinac and X. Chen, "Modeling coverage in camera networks: A survey," *Int. J. Comput. Vis.*, vol. 101, no. 1, pp. 205–226, 2013, doi: [10.1007/s11263-012-0587-7](#).
- [20] Y. Fan, Z. Zhang, S. Zheng, and C. Sha, "Research on scientific point layout of urban image monitoring system," (in Chinese), *TV Technol.*, vol. 35, no. 1, pp. 107–110, 2011.
- [21] S. Chen, "The application and perfection of public security video surveillance system in investigation," (in Chinese), *Crime Res.*, vol. 3, pp. 32–36 Mar. 2009.
- [22] T. Cheng and D. Williams, "Space-time analysis of crime patterns in central London," *ISPRS-Int. Arch. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. 39-B2, pp. 47–52, Jul. 2012, doi: [10.5194/isprarchives-XXXIX-B2-47-2012](#).
- [23] J. Lai, T. Cheng, and G. Lansley, (2015). *Spatio-Temporal Patterns of Passengers' Interests at London Tube Stations*. [Online]. Available: <http://discovery.ucl.ac.uk/1468374/>
- [24] X. Guo and R. Hu, "The effectiveness evaluation for security system based on risk entropy model and Bayesian network theory," in *Proc. 44th Annu. IEEE Int. Carnahan Conf. Secur. Technol.*, Oct. 2010, pp. 57–65.
- [25] H. Lv and R. Hu, "Risk evaluation for security network based on protection model and risk entropy," in *Proc. Int. Conf. Teach. Comput. Sci.*, May 2014, pp. 146–149, doi: [10.2991/ictes-14.2014.39](#).
- [26] B. Wang, W. Wang, V. Srinivasan, and K. C. Chua, "Information coverage for wireless sensor networks," *IEEE Commun. Lett.*, vol. 9, no. 11, pp. 967–969, Nov. 2005, doi: [10.1109/LCOMM.2005.11002](#).
- [27] W. Liu, S. Liao, W. Hu, X. Liang, and X. Chen, "Learning efficient single-stage pedestrian detectors by asymptotic localization fitting," in *Proc. ECCV*, in Lecture Notes in Computer Science, Munich, Germany, vol. 11218. Springer, Sep. 2018, pp. 618–634, doi: [10.1007/978-3-030-01264-9_38](#).
- [28] *Video Surveillance Systems for Use in Security Applications—Part 4: Application Guidelines*, Standard IEC 62676-4:2014, International Electrotechnical Commission, 2014.
- [29] R. W. Lee and J. J. Kulesz, "A risk-based sensor placement methodology," *J. Hazardous Mater.*, vol. 158, nos. 2–3, pp. 417–429, Oct. 2008, doi: [10.1016/j.jhazmat.2008.01.111](#).
- [30] P. Wang, R. Dai, and I. F. Akyildiz, "A spatial correlation-based image compression framework for wireless multimedia sensor networks," *IEEE Trans. Multimedia*, vol. 13, no. 2, pp. 388–401, Apr. 2011, doi: [10.1109/TMM.2010.2100374](#).

- [31] S. Colonnese, F. Cuomo, and T. Melodia, "An empirical model of multi-view video coding efficiency for wireless multimedia sensor networks," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1800–1814, Dec. 2013, doi: [10.1109/TMM.2013.2271475](https://doi.org/10.1109/TMM.2013.2271475).
- [32] Y. Xu, J. Dong, B. Zhang, and D. Xu, "Background modeling methods in video analysis: A review and comparative evaluation," *CAAI Trans. Intell. Technol.*, vol. 1, no. 1, pp. 43–60, Jan. 2016, doi: [10.1016/j.trit.2016.03.005](https://doi.org/10.1016/j.trit.2016.03.005).
- [33] P. J. Grother, G. W. Quinn, and P. J. Phillips. (2011). Report on the Evaluation of 2D Still-Image Face Recognition Algorithms. National Institute of Standards and Technology. Accessed: Sep. 20, 2018. [Online]. Available: <http://face.nist.gov/mbe/>



HONGZHOU ZHANG received the M.Sc. and bachelor's degrees in theoretical physics from Northwest University, Xi'an, China, in 2001 and 1996, respectively, and the Ph.D. degree in image processing and pattern recognition from the Graduate School, CAS, in 2007. He studied at the Shanghai Institute of Applied Physics, CAS. From 2007 to 2012, he was a Research and Development Engineer of the Third Research Institute of the Ministry of Public Security of PRC. From 2012 to 2016, he was also a part-time Postdoctoral Researcher with the School of Automation, Huazhong University of Science and Technology (HUST). Since 2012, he has been an Associate Professor with the School of Information Technology and Cyber Security of People's Public Security University of China (PPSUC). His research interests include the effectiveness evaluation of the video surveillance system for public security applying and domain knowledge modeling and information analysis methods for crime investigation.



PEIYUE LI received the bachelor's degree in security engineering and the M.Sc. degree from the People's Public Security University of China (PPSUC). He is currently pursuing the Ph.D. degree in computer application technology with Beihang University. Since 2014, he has been a Lecturer with the Graduate School, PPSUC. His research interests include security system engineering, artificial intelligence, machine learning, and risk analysis.



ZHIGUO DU received the B.S. degree in telecommunication engineering from the People's Public Security University of China, in 1999, and the M.E. degree in information and communication system from the Beijing University of Posts and Telecommunications, in 2007. He is currently an Associate Professor with the School of Information Technology and Cyber Security of People's Public Security University of China. His research interests include vulnerability assessment of security systems and effectiveness evaluation of video surveillance systems.



WEI DOU received the B.Eng. degree from CAAC, in 2018. He is currently pursuing the M.Sc. degree in security and prevention with the School of Information Technology and Cyber Security of the People's Public Security University of China. His research interests include front-end layout evaluation of police video surveillance and pedestrian recognition.

...