2011

Video Surveillance Over Wireless Sensor and Actuator Networks Using Active Cameras

Dalei Wu
University of Nebraska-Lincoln, dalei-wu@utc.edu

Song Ci
University of Nebraska-Lincoln, sci2@unl.edu

Haiyan Luo
University of Nebraska-Lincoln, haiyan.luo@huskers.unl.edu

Yun Ye
University of Nebraska-Lincoln, yye@huskers.unl.edu

Haohong Wang
TCL Corporation, haohong@ieee.org

Video Surveillance Over Wireless Sensor and Actuator Networks Using Active Cameras

Dalei Wu, Song Ci, *Senior Member, IEEE*, Haiyan Luo, *Student Member, IEEE*, Yun Ye, and Haohong Wang, *Member, IEEE*

***Abstract*—Although there has been much work focused on the camera control issue on keeping tracking a target of interest, few has been done on jointly considering the video coding, video transmission, and camera control for effective and efficient video surveillance over wireless sensor and actuator networks (WSAN). In this work, we propose a framework for real-time video surveillance with pan-tilt cameras where the video coding and transmission as well as the automated camera control are jointly optimized by taking into account the surveillance video quality requirement and the resource constraint of WSANs. The main contributions of this work are: i) an automated camera control method is developed for moving target tracking based on the received surveillance video clip in consideration of the impact of video transmission delay on camera control decision making; ii) a content-aware video coding and transmission scheme is investigated to save network node resource and maximize the received video quality under the delay constraint of moving target monitoring. Both theoretical and experimental results demonstrate the superior performance of the proposed optimization framework over existing systems.**

***Index Terms*—Camera control, content-aware video coding and transmission, video tracking, wireless sensor networks.**

I. Introduction

Recently, as one type of the most popular wireless sensor and actuator networks (WSANs) [1], [2], wireless video sensor networks with active cameras have attracted a lot of research attentions and been adopted for various applications, such as intelligent transportation, environmental monitoring, homeland security, construction site monitoring, and public safety. Although significant amount of work has been done on wireless video surveillance in literature, major challenges still exist in transmitting videos over WSANs and automatically controlling cameras due to the fundamental limits of WSANs, such as, limitations on computation, memory, battery life, and network bandwidth at sensors, as well limitations on actuating speed, delay, and range at actuators.

Some work has been focused on automated camera control for video surveillance applications. In [3], an algorithm was proposed to provide automated control of a pan-tilt camera by using the captured visual information only to follow a person's face and keep the face image centered in the camera view. In [4], an image-based pan-tilt camera control method was proposed for automated surveillance systems with multiple cameras. The work in [5] focused on the control of a set of pan-tilt-zoom (PTZ) cameras for acquiring closeup views of subjects

Manuscript received April 03, 2010; revised March 18, 2011; accepted July 09, 2011. Date of publication August 08, 2011; date of current version October 05, 2011. This work was supported in part by the National Science Foundation (NSF) under Grant CCF-0830493. Recommended by Associate Editor I. Stojmenovic.

D. Wu, S. Ci, and Y. Ye are with the Department of Computer and Electronics Engineering, University of Nebraska-Lincoln, Omaha, NE 68182 USA (e-mail: dwu@unlnotes.unl.edu; sci@engr.unl.edu; yye@huskers.unl.edu).

H. Luo is with Cisco Systems, Austin, TX 79729 USA (e-mail: haiyan.luo@huskers.unl.edu).

H. Wang is with TCL Corporation, Santa Clara, CA 95054 USA (e-mail: haohong@ieee.org).

Color versions of one or more of the figures in this technical note are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2011.2164034

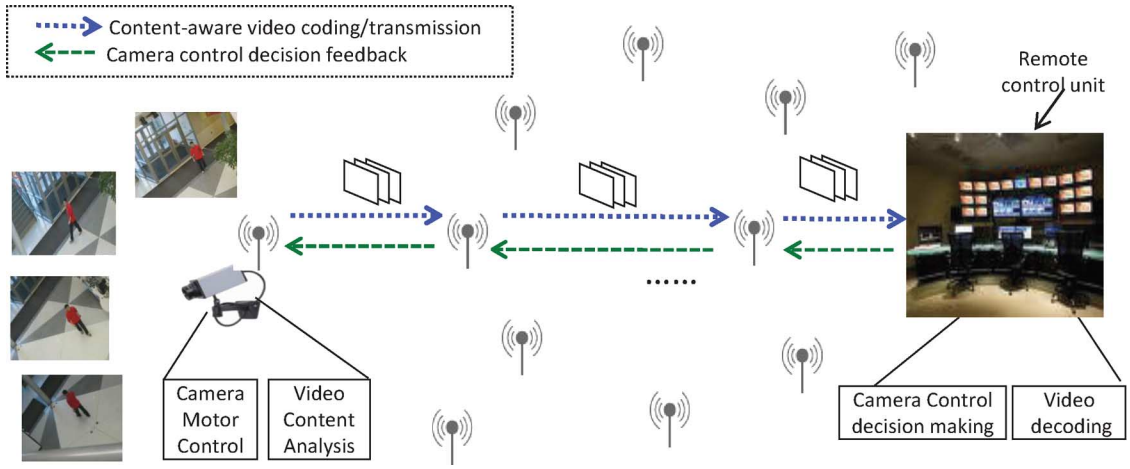

Fig. 1. System model of the proposed framework for real-time video surveillance over WSANs with active cameras.

based on the output of a set of fixed cameras. To extend the applications of these camera control methods into video surveillance over WSANs, both video transmission over WSANs and its impact on the accuracy of camera control need to be considered.

It is challenging to transmit videos over resource-constrained WSANs. Low network bandwidth results in large delay in video delivery, which may not satisfy the low latency requirements in real-time target detection and camera control decision making at the surveillance center. Also, lossy network channels lead to packet losses during video transmission. The resulting received video distortion may not satisfy the quality requirement of target observation at the surveillance center. Moreover, large video distortion may affect the accuracy of camera control since camera control decision is made based on the received videos. Finally, on the one hand, each node in WSNAs is powered by either batteries or solar-energy-harvesting devices, meaning that power is of utmost importance, and must be aggressively conserved. On the other hand, wireless video surveillance is extremely useful to provide continuous visual coverage in the surveillance areas. As a result, the continuous transmission of large amount of surveillance videos poses significant challenges to the power conservation of WSANs. Till now, although there has been work done on resource allocation for co-design of communication protocols and control strategies [6], few existing work focuses on the joint optimization of video coding, video transmission, and camera control for real-time video monitoring applications in WSANs.

In this technical note, we propose a framework for real-time video surveillance over WSANs in which pan/tilt cameras are used to track moving targets of interest and transmit the captured videos to a remote control unit (RCU). The main contributions of this technical note include: (i) an automated camera control method is developed to keep tracking the target by taking into account the impacts of the delay of video transmission and control decision feedback on the control-decision making process; (ii) a content-aware video coding and transmission scheme is investigated to save network node resource while maximizing the received video quality under the delay constraint of target monitoring. It is worth noting that although the camera control decision-making is performed by the central RCU, the proposed video processing method can also be used in smart cameras [7] or mobile agents [8] in distributed video surveillance systems [9], [10].

The remainder of this technical note is organized as follows. In Section II, we present the system model and formulated optimization problem. Section III presents an automated camera control method. The solution procedure is described in Section IV. Section V shows experimental results and Section VI concludes this technical note.

II. Problem Description

A. Proposed System Model

Fig. 1 shows the proposed system model for wireless video surveillance. The captured videos are coded at the camera, and the resulting video packets are transmitted over a wireless sensor network to a RCU where videos are reconstructed for target monitoring. To keep tracking the target, we propose that the camera control decision-making be performed at the RCU since RCU has more computational capability and more information about the target. Therefore, based on the video analysis at the RCU, the controller needs to calculate and determine the camera control parameters, i.e., pan and tilt velocities, and send them back to the camera motor.

Note that each captured video frame can be considered as being composed of a target part and a background part. To save node resource, in this work we aim to develop a content-aware approach to the video processing by which the target part is coded and transmitted with a higher priority than the background part. In other words, the limited network resources are concentrated on the coding and transmission of the target part while the background part is processed in a best-effort fashion. Note that in this work the motivation for background-foreground separation is mainly to roughly divide a captured video frame into different regions and find the region of interest (ROI, i.e., the target part) for resource allocation adaptation by content-aware video coding and transmission rather than to accurately perform target identification and extraction. Therefore, suitable background-foreground separation methods [11]–[14] can be selected to achieve a good tradeoff between the accuracy and complexity at the camera side based on the camera's computational capability.

B. Kinematic Model of Pan-Tilt Cameras

The control objective is to maintain the target being tracked in the center of the camera view. In the following, we adopt the kinematic model of pan-tilt cameras developed in [3] to explain the camera control objective. Let (x_c, y_c) be the target offset with respect to the center of the image, i.e., the coordinates of the centroid of the segmented target image inside the window where the target is located, as shown in Fig. 2. We next take the pan movement of the camera as an example to study the camera control problem.

Let f be the camera focal length. As shown in Fig. 3, the relationship between target coordinate x_c and pan angle θ can be expressed as

$$x_c = f\tan(\alpha - \theta) \tag{1}$$

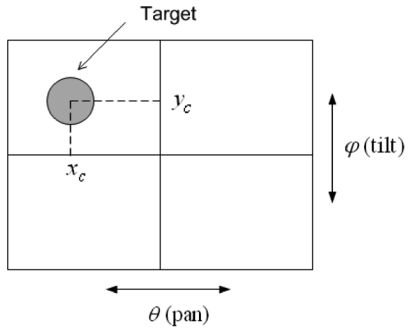

Fig. 2. Target offset in the image plane.

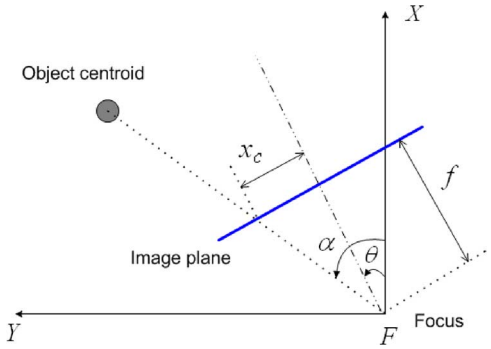

Fig. 3. Dependence of the target coordinate x_c on the pan angle θ.

where α is the angle of the target centroid in the frame FXY. Differentiating (1) with respect to time t, a kinematic model for the x_c offset is obtained as

$$\dot{x_c} = \frac{f}{\cos^2(\alpha-\theta)}(\omega_\alpha - \omega_\theta) \tag{2}$$

where the pan velocity ω_θ is considered as a control input. The angular velocity ω_α of the target depends on the instantaneous movement of the target with respect to the frame FXY. The control problem consists in finding an adaptive feedback control law for the system as shown in (2) with control input ω_θ such that $\lim_{t\to\infty} x_c(t) = 0$.

C. Proposed Content-Aware Video Coding and Transmission

Without loss of generality, it is assumed that the target and background parts in the current video frame are composed of I_o and I_b groups of blocks (GOB), respectively. We also assume that each GOB is encoded into a packet. Let $\{\pi_1, \pi_2, \cdots, \pi_{I_o}\}$ be the resulting I_o target packets and $\{\pi_{I_o+1}, \pi_{I_o+2}, \cdots, \pi_{I_o+I_b}\}$ the I_b background packets of the current video frame. For simplicity, define $I = I_o + I_b$. For each packet π_i, let s_i be the corresponding source coding parameter vector, and c_i the channel transmission parameter vector.

1) Delay Constraint: To perform accurate camera control and provide smooth video playback at the RCU, each video frame to be transmitted is associated with a frame decoding deadline $T^{\max}$. In real-time video communications, the average video frame decoding deadline $T^{\max}$ is linked with the video frame rate f_r as $T^{\max} \approx (1/f_r)$ [15]. The frame decoding deadline indicates that a delay deadline is imposed on the transmission of each packet composing the frame by $T^{\max}$ [15], i.e., $T_i(s_i, c_i) \le T^{\max}$ $(i = 1, 2, \cdots, I)$, where $T_i(s_i, c_i)$ is the end-to-end delay of packet π_i transmitted from the camera to the RCU, which is a function of s_i and c_i.

2) Expected Received Video Quality: We employ the ROPE algorithm [16] to evaluate the received video distortion. ROPE can optimally estimate the overall distortion of decoder frame reconstruction due to quantization, error propagation, and error concealment by recursively computing the total decoder distortion at pixel level precision. Moreover, in consideration of the importance difference for video surveillance between the target and background parts, we assume that each packet π_i is associated with a quality impact factor λ_i. λ_i can be determined by the quality requirement on the corresponding video parts. Given λ_i, the expected distortion of the whole video frame, denoted by $E[D]$, can be written as [15] $E[D] = \sum_{i=1}^{I} \lambda_i E[D_i](s_i, p_i, \zeta)$. Here $E[D_i]$ is the expected distortion of packet π_i, which is a function of the source coding parameter s_i, end-to-end packet loss rate p_i, and error concealment scheme ζ for the corresponding GOB [16]. The calculation of p_i will be presented in Section IV-B.

3) Problem Formulation of Content-Aware Video Coding and Transmission: The objective of video coding and transmission optimization is to determine the optimal values of source coding parameter and channel transmission parameter vectors $\{s_i, c_i\}$ for each packet π_i of the current video frame to maximize the expected received video quality under the delay constraint, i.e.

$$\{s_i, c_i\} = \arg\min \sum_{i=1}^{I} \lambda_i E[D_i]$$
$$s.t.: \quad T_i(s_i, c_i) \le T^{\max} \quad (i = 1, 2, \cdots, I). \tag{3}$$

Note that the optimization is performed one frame at a time. Nonetheless, this framework can potentially be improved by optimizing the video encoding and transmission over multiple buffered frames, which can integrate the packet dependencies caused by both error concealment and prediction in source coding into the optimization framework. Such a scheme, however, would lead to a considerably higher computational complexity. The detailed solution to the problem in (3) will be discussed in Section IV.

III. Camera Control Decision Making at the RCU

In this section, we modify and improve an existing pan-tilt camera control algorithm [3] by taking into account the impacts of both video transmission delay and control decision feedback delay on the camera control decision making process at the RCU. In the following, we take the pan angle control process as an example to explain the improved algorithm.

As discussed in Section II-B, the objective of pan angle control is to determine control input ω_θ such that $\lim_{t\to\infty} x_c(t) = 0$. Similar to the work in [3] we use the following control law:

$$\omega_\theta = \hat{\omega}_\alpha + \eta x_c \tag{4}$$

where η is a positive gain, and the estimate $\hat{\omega}_\alpha$ of ω_α is obtained from the dynamic part of the automated control, which is designed as a parameter update law. According to Proposition 1 in [3], to ensure that the resulting closed-loop adaptive camera control system is asymptotically stable, the update law is derived as

$$\dot{\hat{\omega}}_\alpha = \gamma \frac{f x_c}{\cos^2(\alpha-\theta)}. \tag{5}$$

Both target offset $\{x_c, y_c\}$ and pan/tilt angle $\{\theta, \varphi\}$ needed for control decision making are measured based on the received video frames. Let $\mathcal{F}_k$ be the video frame captured at discrete time instant t_k $(k = 0, 1, 2, \cdots)$. From $\mathcal{F}_k$ the offsets $\{x_{ct_k}, y_{ct_k}\}$ and the pan and tilt angles $\{\theta_{t_k}, \varphi_{t_k}\}$ can be measured at the RCU by using the Mean Shift algorithm as in [3]. Note that the motivation for conducting the Mean Shift algorithm at the RCU instead of at the camera side is based on the observation that in some scenarios the RCU may already have some a

TABLE I
MODIFIED CAMERA CONTROL ALGORITHM AT THE RCU

Initialization:	at $t = t_0$, $\hat{\omega}_{\alpha t_0}$ is set arbitrary.
Step 1:	Measure θ_{t_k} and x_{ct_k};
Step 2:	Calculate α_{t_k} following Eq. (9);
Step 3:	Calculate $\hat{\theta}_{t_k+T_k}$ and $\hat{\alpha}_{t_k+T_k}$ following Eqs. (7), (8);
Step 4:	Calculate $x'_{ct_k+T_k}$ following Eq. (10);
Step 5:	Calculate $\omega_{\theta t_k+T_k}$ following Eq. (6);
Step 6:	Calculate $\hat{\omega}_{\alpha t_{k+1}+T_{k+1}}$ following Eq. (11);
$t_k \to t_{k+1}$, repeat from Step 1.	

priori knowledge of the target, such as the color or size information of the target. It is expected that integrating these information into the Mean Shift algorithm could provide more accurate measurement of the offsets and quantities of the target.

Note that $\{x_{ct_k}, y_{ct_k}\}$ and $\{\theta_{t_k}, \varphi_{t_k}\}$ only represent the position information of the target and camera at time instant $t = t_k$ when video frame $\mathcal{F}_k$ is captured. In fact, a significant amount of delay is incurred during transmitting the frame $\mathcal{F}_k$ to the RCU as well as sending the control decision back to the camera. Therefore, in order to perform accurate camera control, the further change of the target position due to the possible movement of the target during this delay period also needs to be considered in camera control decision making.

At $t = t_0$, the first estimate for ω_α, $\hat{\omega}_{\alpha t_0}$, is set arbitrary. For the control law (4) at $t = t_k$, we modify the relation $\omega_{\theta t_k} = \hat{\omega}_{\alpha t_k} + \eta x_{ct_k}$ in [3] into

$$\omega_{\theta t_k+T_k} = \hat{\omega}_{\alpha t_k+T_k} + \eta x'_{ct_k+T_k} \tag{6}$$

where $x'_{ct_k+T_k}$ is the estimate of offset x_c when the camera control decision arrives at the camera, and T_k is the total delay of the video frame transmission and the control decision feedback.

Next we consider how to calculate $x'_{ct_k+T_k}$ based on measurement $\{x_{ct_k}, \theta_{t_k}, \alpha_{t_k}\}$ by taking into account T_k. Let T_{dk} be the end-to-end delay in transmitting video frame $\mathcal{F}_k$ from the camera to the RCU, which will be discussed in Section IV-B. Let T_{fk} be the delay in sending back the control decision based on $\mathcal{F}_k$ from the RCU to the camera. Without loss of generality, we assume that T_{fk} is a constant. Therefore, $T_k = T_{dk} + T_{fk}$. Based on $\{x_{ct_k}, \theta_{t_k}, \alpha_{t_k}\}$, we can compute both estimate $\hat{\theta}_{t_k+T_k}$ of θ and estimate $\hat{\alpha}_{t_k+T_k}$ of α at time instant $t_k + T_k$ as

$$\hat{\theta}_{t_k+T_k} = \theta_{t_k} + \omega_{\theta t_k} \cdot T_k; \tag{7}$$

$$\hat{\alpha}_{t_k+T_k} = \alpha_{t_k} + \hat{\omega}_{\alpha t_k} \cdot T_k \tag{8}$$

where α_{t_k} can be derived from (1) as

$$\alpha_{t_k} = \theta_{t_k} + \arctan \frac{x_{ct_k}}{f}. \tag{9}$$

Based on (1), estimate $x'_{ct_k+T_k}$ of offset x_c at time instant $t_k + T_k$ can be calculated as

$$x'_{ct_k+T_k} = f \cdot \tan\left(\hat{\alpha}_{t_k+T_k} - \hat{\theta}_{t_k+T_k}\right). \tag{10}$$

At $t = t_{k+1} + T_{k+1}$, to derive the control input $\omega_{\theta t_{k+1}+T_{k+1}}$, the estimate $\hat{\omega}_{\alpha t_{k+1}+T_{k+1}}$ of $\omega_{\alpha t_{k+1}+T_{k+1}}$ can be calculated based on (5) as follows:

$$\hat{\omega}_{\alpha t_{k+1}+T_{k+1}} = \hat{\omega}_{\alpha t_k+T_k} + \gamma \frac{f x'_{ct_k+T_k}}{\cos^2\left(\hat{\alpha}_{t_k+T_k} - \hat{\theta}_{t_k+T_k}\right)}. \tag{11}$$

The complete modified camera control algorithm at the RCU is summarized in Table I.

IV. OPTIMIZED CONTENT-AWARE VIDEO CODING AND TRANSMISSION

A. Content-Aware Video Coding

Let $s_i \in \mathbb{S}$ be the source coding parameters for the ith GOB, where $\mathbb{S}$ is the set of all admissible values of s_i and $|\mathbb{S}| = J$. For a given GOB, different coding parameters will lead to different amounts of compression-induced distortion. On the other hand, different coding parameters also result in different packet lengths and packet loss rates, which will lead to different amounts of transmission-induced distortion. A robust error concealment technique helps avoid significant visible error due to the packet loss in the reconstructed frames at the decoder. However, some error concealment strategies cause packet dependencies, which makes the source coding optimization further complicated. It is important to point out that, although some straightforward error concealment strategies do not cause packet dependencies, as a generic framework, the more complicated scenario is considered here as a superset for the simpler cases.

Without loss of generality, we assume that the current GOB depends on its previous z GOBs ($z \geq 0$). Therefore, the optimization goal in (3) becomes

$$\min_{s_1,\cdots,s_I} \sum_{i=1}^{I} E[D_i](s_{i-z}, s_{i-z+1}, \cdots, s_i), \quad i - z > 0 \tag{12}$$

where $E[D_i](s_{i-z}, s_{i-z+1}, \cdots, s_i)$ represents the expected distortion of the ith GOB, which depends on the $z + 1$ decision vectors $\{s_{i-z}, s_{i-z+1}, \cdots, s_i\}$ under the packet delay deadline. The problem (12) can be solved by a dynamic programming approach. Due to the limit of space, we will not elaborate the solution procedure in this technical note. Interested readers may refer to [15].

The computational complexity of the above algorithm is $O(I \times |\mathbb{S}|^z)$, which is much more efficient than the exponential computational complexity of an exhaustive search algorithm. Clearly for cases with smaller z, the complexity is quite practical to perform the optimization. On the other hand, for larger z, the complexity can be limited by reducing the cardinality of $\mathbb{S}$. The practical solution would be an engineering decision and tradeoff between the camera's computational capability and the optimality of the solution. For resource-constrained WSANs, simple error concealment strategies could be considered to further reduce the computational complexity.

B. Content-Aware Video Transmission

Let c_g denote the video packet class where $g = 0$ if packet π_i^j is a target packet, and $g = 1$ if it is a background packet. Without loss of generality, we assume that (u, v) is a link/hop of path $\mathcal{R}$, where nodes u and v are the hth and $(h + 1)$th nodes of path $\mathcal{R}$, respectively. In wireless environments, the transmission of a packet over each hop (u, v) of the network can be modeled as follows: once the packet arrives at node u, the packet header is extracted to determine the packet class. If it is a target packet, it will be put ahead of all background packets in the queue of node u, waiting for being transmitted/served over link (u, v). We assume that all packets of the same class in the queue of node u are scheduled following a first-come, first-served (FCFS) fashion. If the packet gets lost during transmission with the packet error probability $p_{g,(u,v)}$ due to signal fading over the link (u, v), it will be retransmitted until it is either successfully received or discarded because its delay deadline $T^{\max}$ was exceeded. As a result of the retransmission mechanism, the packet loss probability over link (u, v) is mainly exhibited as the probability of packet drop $p_{g,u}$ due to delay deadline expiration

TABLE II
PROCESSING LOAD DISTRIBUTION OF THE PROPOSED FRAMEWORK

Algorithm	Background-foreground separation	ROPE	Mean Shift
Processing load	$O(mN_f)$	$O(cN_f)$	$O(N_w^2)$

when queueing at node u. Based on priority queueing analysis, $p_{g,u}$ can be calculated as [14]

$$\begin{aligned} p_{g,u} &= \text{Prob}\left(E\left[W_{g,(u,v)}\right] + t_{g,u}^0 > T^{\max}\right) \\ &= \left(\sum_{g=0}^{1} \phi_{g,u} E[Z_{g,u}]\right) \\ &\quad \times \exp\left(-\frac{\left(T^{\max} - t_{g,u}^0\right)\sum_{g=0}^{1} \phi_{g,u} E[Z_{g,u}]}{E\left[W_{g,(u,v)}\right]}\right) \end{aligned} \tag{13}$$

where $t_{g,u}^0$ is the arrival time of the packets of class c_g at node u, $\phi_{g,u}$ the average arrival rate of the Poisson input traffic into the queue at node u, $E[Z_{g,u}]$ the average service time at node u, and $E[W_{g,(u,v)}]$ the average packet waiting time at the queue of node u.

From (13), the end-to-end packet loss probability of a packet of video class c_g transmitted over path $\mathcal{R}$ can be expressed as $p_g = 1 - \prod_{(u,v)\in\mathcal{R}}(1-p_{g,u})$. The end-to-end packet delay T_{dk} is the sum of the packet delay $t_{g,(u,v)}$ over each link (u,v), which can be obtained as [15] $T_{dk} = \sum_{(u,v)\in\mathcal{R}} t_{g,(u,v)} = \sum_{(u,v)\in\mathcal{R}} \{E[W_{g,(u,v)}] + E[Z_{g,u}]\}$.

V. EXPERIMENTAL RESULTS

Experiments of video surveillance were conducted in the lobby of a building to evaluate the performance of the proposed framework in terms of the camera control effectiveness and the surveillance video quality. All captured video frames except the first one were coded as inter frames. Each video frame was divided into a target part and a background part by the method of Mixture of Gaussians. The target and background parts were separately coded. The resulting video stream was transmitted to a server via a wireless sensor network. Based on the received video, the target offset was calculated at the server by using the Mean Shift algorithm for camera-control decision making. The processing load distribution of the proposed framework is presented in Table II, where N_f is the video frame size, m is the number of Gaussian distribution and practically set to be between 3 and 5 [12], c is usually between 10 and 20 [16], depending on both the frame types (intra-coded or inter-coded) and the adopted error concealment schemes, and N_w is the search window size.

In our first set of experiments, the target first made circular movement to the left at an angular velocity of $\omega_\alpha = 0.4$ rad/second for 20 seconds, and then made circular movement to the right at $\omega_\alpha = -0.3$ rad/second for another 20 seconds. Initially, the angle $\alpha = 0.5$ rad, offset $x_{ct_0} = 0.02$ m, and pan angle $\theta_{t_0} = 0$ rad. Fig. 4(a) shows that both the estimated angular velocity of the target and the angular velocity of pan tend asymptotically to the actual angular velocity of the target. Fig. 4(b) shows that the pan and target angles converge after some time duration under each movement scenario. Fig. 4(c) shows that the offset x_c tends to be zero after some time duration under each movement scenario, meaning that the captured target is at the center of the image plane. Therefore, Fig. 4(a)–(c) demonstrate the effectiveness of the proposed camera control method.

In the second set of experiments, the target was allowed to walk back and forth across the lobby. We compared the proposed camera control strategy with the control strategy without considering the impact of the delay of video transmission and control decision feedback. As shown

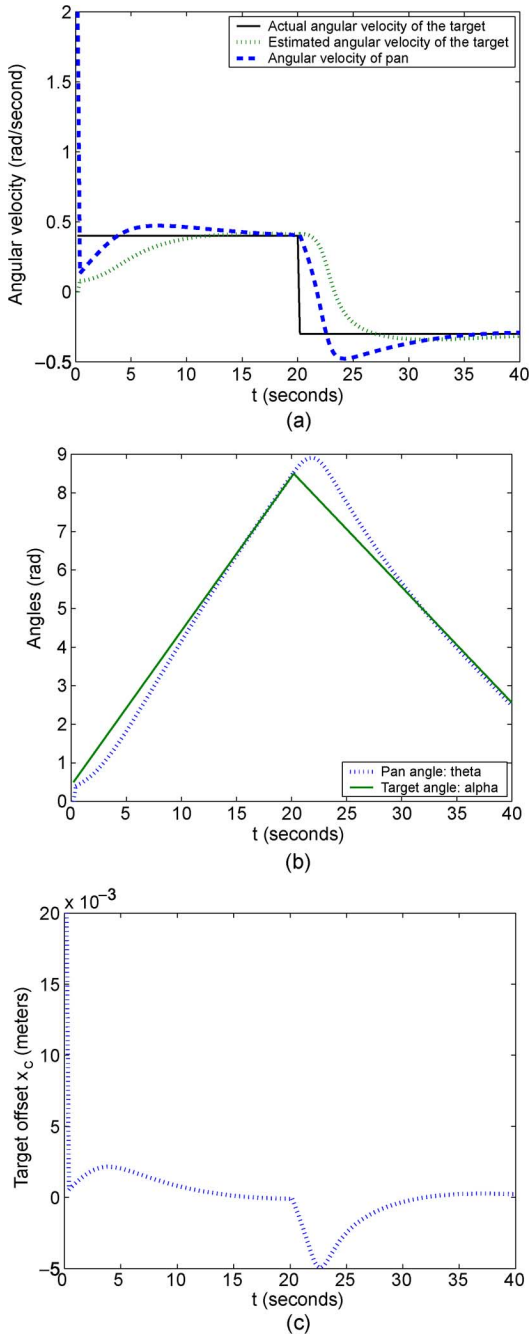

Fig. 4. Evaluation of the camera control effectiveness. (a) Pan velocity and estimated angular velocity of the target. (b) Pan angle and target angle. (c) Target offset.

in Fig. 5, the proposed control strategy is more effective, leading to smaller target offset x_c with respect to the center of the image. This is because the proposed strategy takes into account the target movement during the delay time of video transmission and control decision feedback.

We evaluated the video quality obtained by the proposed content-aware video coding and transmission. Under each movement scenario of the target, the surveillance videos were captured at two different video frame rates: 15 and 30 f/s. Fig. 6(a) and (b) show the received video quality of different video parts at the RCU, measured by peak signal-to-noise ratio (PSNR). We observe that the PSNRs of the target part are over 5 dB higher than those of the background part under both

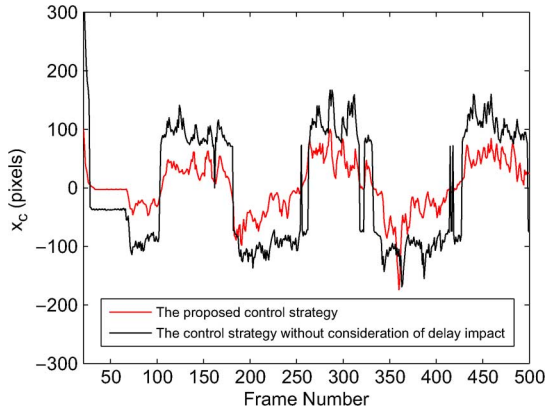

Fig. 5. Comparison of different camera control strategies.

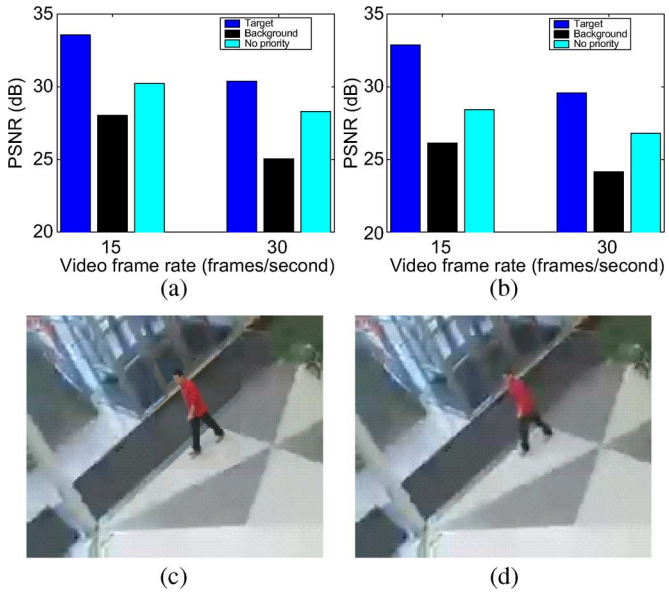

Fig. 6. Comparison of the received surveillance video quality. (a) Average PSNRs when $\alpha = 0.4\,\mathrm{rad}$. (b) Average PSNRs when $\alpha = -0.3\,\mathrm{rad}$. (c) Received video frame by using the proposed content-aware video coding and transmission. (d) Received video frame without using content-aware video coding and transmission.

video frame rates. Moreover, the PSNRs of the target are also much higher than those of the received videos without using content-aware coding and transmission.

Fig. 6(c) and (d) show the comparison of the received surveillance video quality with and without the proposed content-aware video coding and transmission scheme. As shown in Fig. 6(c), the target part has a much better visual quality than the background part. Moreover, the target part as shown in Fig. 6(c) also has a better visual quality than the received video without using the content-aware video coding and transmission scheme as shown in Fig. 6(d).

VI. Conclusion

In this technical note, we presented a co-design framework for video surveillance with active cameras by jointly considering the camera control strategy and video coding and transmission over wireless sensor and actuator networks. To keep tracking a target of interest, an automated camera control method was proposed based on the received surveillance videos. To save the network node resource, a content-aware video coding and transmission scheme was developed to maximize the received video quality under the delay constraint imposed by video surveillance applications. The impact of the delay of video transmission and control decision feedback on camera-control decision making has also been investigated to provide accurate camera control. Experimental results demonstrate the efficacy of the proposed co-design framework.

References

[1] I. F. Akyildiz and I. H. Kasimoglu, "Wireless sensor and actor networks: Research challenges," *Ad Hoc Netw.*, vol. 2, no. 4, pp. 351–367, May 2004.

[2] X. Cao, J. Chen, Y. Xiao, and Y. Sun, "Building environment control with wireless sensor and actuator networks: Centralized versus distributed," *IEEE Trans. Ind. Electron.*, vol. 27, no. 11, pp. 3596–3605, Nov. 2010.

[3] P. Petrov, O. Boumbarov, and K. Muratovski, "Face detection and tracking with an active camera," in *Proc. 4th Int. IEEE Conf. Intell. Syst.*, Varna, Bulgaria, Sep. 2008, pp. 14–39.

[4] S. Lim, A. Elgammal, and L. S. Davis, "Image-based pan-tilt camera control in a multi-camera surveillance environment," in *Proc. Int. Conf. Mach. Intell.*, 2003, pp. 645–648.

[5] N. Krahnstoever, T. Yu, S. Lim, and K. Patwardhan, "Collaborative control of active cameras in large-scale surveillance," in *Multi-Camera Networks—Principles and Applications*, H. Aghajan and A. Cavallaro, Eds. Amsterdam, The Netherlands: Elsevier, 2009.

[6] A. Deshpande, C. Guestrin, and S. Madden, "Resource-aware wireless sensor-actuator networks," *IEEE Data Eng.*, vol. 28, no. 1, pp. 40–47, 2005.

[7] Z. Zivkovic and R. Kleihorst, "Smart cameras for wireless camera networks: Architecture overview," in *Multi-Camera Networks—Principles and Applications*, H. Aghajan and A. Cavallaro, Eds. Amsterdam, The Netherlands: Elsevier, 2009.

[8] B. Rinner and M. Quaritsch, "Embedded middleware for smart camera networks and sensor fusion," in *Multi-Camera Networks—Principles and Applications*, H. Aghajan and A. Cavallaro, Eds. Amsterdam, The Netherlands: Elsevier, 2009.

[9] M. Dalai and R. Leonardi, "Video compression for camera networks: A distributed approach," in *Multi-Camera Networks—Principles and Applications*, H. Aghajan and A. Cavallaro, Eds. Amsterdam, The Netherlands: Elsevier, 2009.

[10] P. Dragotti, "Distributed compression in multi-camera systems," in *Multi-Camera Networks—Principles and Applications*, H. Aghajan and A. Cavallaro, Eds. Amsterdam, The Netherlands: Elsevier, 2009.

[11] M. Piccardi, "Background subtraction techniques: A review," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, 2004, pp. 3099–3104.

[12] C. Stauffer and W. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, 2000.

[13] Y. Sheikh, O. Javed, and T. Kanade, "Background Subtraction for Freely Moving Cameras," in *Proc. IEEE Int. Conf. Comp. Vision*, 2009, pp. 1219–1225.

[14] D. Wu, H. Luo, S. Ci, H. Wang, and A. Katsaggelos, "Quality-driven optimization for content-aware real-time video streaming in wireless mesh networks," in *Proc. IEEE GLOBECOM*, Dec. 2008, pp. 1–5.

[15] D. Wu, S. Ci, H. Luo, H. Wang, and A. Katsaggelos, "Application-centric routing for video streaming over multi-hop wireless networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1721–1734, Dec. 2010.

[16] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 6, pp. 966–976, Jun. 2000.