Quality-of-Content (QoC)-Driven Rate Allocation for Video Analysis in Mobile Surveillance Networks

Xiang Chen*, Jenq-Neng Hwang*, Kuan-Hui Lee*, Ricardo L. de Queiroz †

*Department of Electrical Engineering, University of Washington,
Seattle, WA 98195, USA. Email: {xchen28, hwang, ykhlee}@uw.edu
† Department of Computer Science, Universidade de Brasilia,
Brasilia, Brazil. Email: queiroz@ieee.org

***Abstract*—Nowadays, more and more videos are transmitted for video analytics purposes rather than human perceptions. In mobile surveillance networks, a cloud server collects videos delivered from multiple moving cameras and detects suspicious people in all the camera views. However, all the videos recorded by moving cameras such as phone or dash cameras are uploaded through bandwidth-limited wireless networks. Therefore, videos are required to be encoded with high compression ratio to satisfy the total data rate constraint, which may affect the video analyses (e.g., human detection/tracking and action recognition, etc.) performance due to the degraded video decoding qualities at the server side. In this paper, we propose an effective content-driven video source coding rate allocation scheme, which can improve the human detection success rate in mobile surveillance networks under a total data rate constraint. The proposed scheme allocates appropriate amount of data rate to each moving camera based on the corresponding content information (i.e., human detection results). A model of human detection accuracy based on object area and video quality is provided. The rate allocation problem is formulated as a convex optimization problem and can be solved by standard solvers. Simulations with real video sequences demonstrate the effectiveness of our proposed scheme.**

***Keywords*—rate allocation; video analysis; human detection; visual surveillance; convex optimization**

I. Introduction

The rapidly increasing demand of video streaming applications has boosted the development of wireless video transmission technologies [1], [2]. As predicted in [3], 72 percent of all consumer mobile Internet traffic will be mobile video in 2019, up from 55 percent in 2014. Furthermore, mobile data traffic will exponentially increase between 2014 and 2019, representing a 57 percent of compound annual growth rate (CAGR), which is about three times faster than fixed IP traffic. Due to the bandwidth-limited nature of wireless channels, it is crucial to design efficient wireless video transmission schemes for the bandwidth-consuming real-time video streaming services [4].

In traditional wireless video transmission research, the optimization criteria are either quality-of-service (QoS) based design [1], or quality-of-experience (QoE) based design [5]–[9]. For QoS-based design, network parameters such as packet loss, delay, jitter, etc. are jointly considered in order to improve the video streaming applications from a network perspective. For QoE-based design, the user perception and experience of decoded videos are combined with the QoS parameters so that video transmission parameters can be adjusted to improve users' satisfaction [10]. Both subjective and objective video quality measurements have been developed to quantify the QoE-based system design [11].

Although most of video transmission services are designed for human perceptions, more and more video streaming data are collected for video analytics purposes. In [12], authors developed a vehicle tracking system with static surveillance cameras. In [13], a live fish tracking system is developed based on low-contrast and low-frame-rate stereo videos. Based on human detectors, pedestrian tracking systems in single moving camera are developed [14], [15]. Moreover, a system of on-road pedestrian tracking across multiple driving recorders for mobile surveillance network is proposed in [16]. Most existing human-perception-based (QoE-based) wireless video transmission designs may not be optimal for video analytics purposes. Therefore, it is necessary to develop more efficient video transmission schemes for surveillance and computer vision applications.

As intelligent surveillance systems become more and more important for crime investigation and tragedy prevention, mobile surveillance networks with multiple moving cameras, which have more flexible camera views comparing to traditional surveillance systems with static cameras, have thus been introduced [16]. As indicated in [16], videos are recorded by driving recorders (dash cameras) and uploaded to remote cloud servers for further automatic analyses. Due to the mobility nature of moving cameras, wireless wide area networks (WWAN) have to be used for video transmissions, where efficient rate allocation is necessary because of the limited wireless resources.

Among different applications in intelligent mobile surveillance networks, such as human tracking, action recognition, behavior understanding, etc., human detection is the first step and its result will critically affect the performance of other

This study is conducted under the 103-EC-17-A-03-S1-214 project from the Ministry of Economic Affairs (MOEA) of Taiwan and Advanced Wireless Broadband System and Inter-networking Application Technology Development Project of the Institute for Information Industry which is subsidized by the Ministry of Economy Affairs of Taiwan.

978-1-4673-7478-1/15/$31.00 ©2015 IEEE

human-related video analysis applications [16]. In [17], [18], image/video features instead of the full video sequences are uploaded to the cloud servers for video analyses. Although transmitting features can save lots of wireless resources, they are not suitable for surveillance purposes since the full video sequences are requried to be archived in the server for future investigations. In [19], authors proposed a saliency-based rate control for human detection with a single camera. Based on a properly designed saliency map, this scheme adaptively adjusts the quantization parameters (QPs) to preserve regions with small contrast from excessive smoothing so that the human detection accuracy can be improved. In this paper, we propose a quality-of-content (QoC)-driven video source coding rate allocation scheme for human detection in the mobile surveillance networks with multiple moving cameras. Instead of considering human perception in traditional video streaming design, the proposed scheme maximizes the overall human detection accuracy at the remote server when multiple moving cameras upload videos via WWAN with a total data rate constraint. We analytically evaluate the factors that affect the probability of successful human detections and propose a video source coding rate allocation algorithm based on the human detection results in the past group of pictures (GoP). To the best of our knowledge, there is no existing QoC-driven work conducted in video encoding rate allocation for human detections in mobile surveillance system when multiple moving cameras compete for the limited wireless resources.

The rest of this paper is organized as follows. In Section II, we will describe the scenario and system structure of mobile surveillance network. In Section III, evaluation of the factors that affect the successful human detections is provided. Section IV gives the proposed video source coding rate adaptation algorithm. Simulation results are shown in Section V, followed by the conclusion remarks in Section VI.

II. Scenarios and System Structure

As shown in Fig. 1, a mobile surveillance network consists of multiple moving cameras (mobile nodes) such as dash cameras and smartphone cameras, which are randomly distributed and moving around in the areas with different pedestrian densities. Each camera can encode and upload videos via a WWAN to a remote cloud server in real time for further video analyses, such as human detection. The system structure is shown in Fig. 2, where captured camera views are encoded with the high efficiency video coding (HEVC) [20] with different encoding data rates. To reduce the cost and computational complexity on each mobile node, human detection is performed in the cloud server. After human detection is performed, an upload scheduling and resource allocation module collects the human detection results (contents) and assigns different source encoding data rate target to each camera. The overall encoding data rate is constrained such that the data transmission can be better supported by the network. Therefore, the human detection accuracy is only related to the source coding data rate allocation. In this work, we assume all the video analyses are conducted in the cloud server. Not only can it archive videos in the cloud server for further investigations, it can save computational cost and power at mobile nodes as well, especially for smartphone cameras.

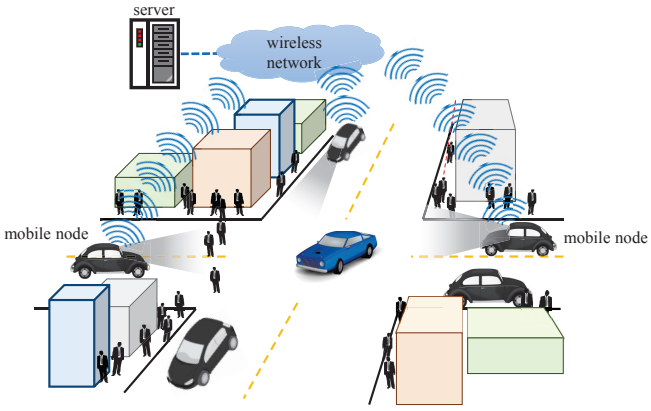

Fig. 1: Scenario of mobile surveillance network.

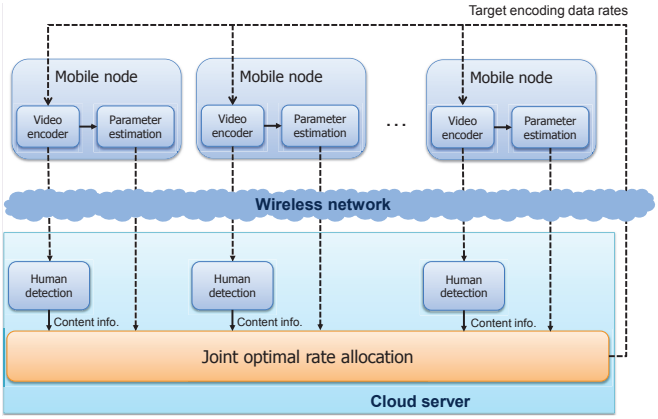

Fig. 2: Proposed system structure.

III. Effect of Video Quality on Human Detector

Many human detectors have been proposed in the literatures. In [21], a human detector, which can effectively represent the shape of human, has been proposed based on the histogram of oriented gradient (HOG) features. The implicit shape model (ISM) proposed in [22] applies a voting scheme based on multi-scale interest points to create plenty of detection hypotheses, and a codebook is used to preserve the trained features. The deformable part model (DPM) [23], an extension of the idea in [21], uses a root model and several part models to describe different partitions of an object. Based on a predefined geometry, the part models are spatially connected with the root model so that the object can be precisely depicted. Among different human detectors, the DPM is a well-accepted robust and computational efficient scheme. Therefore, we adopt the DPM as the human detection scheme in this paper. But similar concept can be applied to other detection schemes.

The DPM object detector is based on HOG features, which can be affected by the artifacts created from video encoder

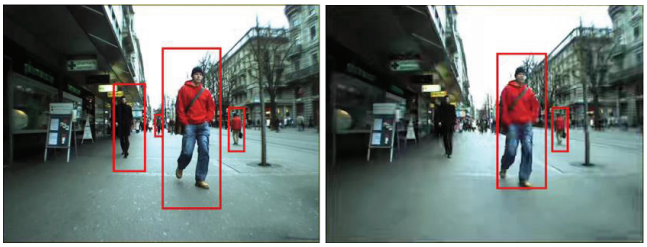

Fig. 3: Human detection result of DPM. Video clip: BAHNHOF in the ETHZ dataset [25]. Left: QP=15; Right: QP=39

with different compression ratios [19]. Therefore, the received video quality will affect the detection performance in the cloud server. Figure 3 shows a comparison of the DPM detection results with two different video encoding qualities in terms of different QPs. When the video quality is poor, smaller objects in the view have lower probability to be successfully detected compared to the larger objects in the view since a larger QP may smooth out the detailed shape information of smaller objects. Figure 4 illustrates the human detection accuracy with different object areas (in terms of number of pixels) and QPs of HEVC encoder [24]. Six video clips in ETHZ dataset [25] are tested and each video is encoded with 11 different QPs from 15 to 45. The detection results are compared with the ground-truth coordinate labels of each object in the dataset. If the overlapped area of the detection result and the ground-truth is larger than 50 percent of the ground-truth area, the detected object is regarded as a successful detection [23]. The detection accuracy of a specific object area a is calculated by counting all the true-positive detected objects whose areas are larger than this specific value a and divided by the total number of objects whose area is larger than this value a. According to Fig. 4, the detection accuracy increases with better video frame quality (smaller QP) and larger object area.

Suppose A is a random variable representing the object area, and Q is a random variable representing QP. Due to the independence of A and Q, the detection accuracy in Fig. 4 can be expressed as:

$$P_{A,Q}(a,q) = f(A \geq a)\, g(q), \tag{1}$$

where $f(\cdot)$ is the probability of true-positive detection result when the objects area is larger than a. $g(\cdot)$ is the probability of true-positive detection result as a function of video encoding QP q. In total 6 videos with VGA (640×480) resolution in ETHZ dataset [25] and 2 videos with 720p (1280×720) resolution recorded in the University of Washington (UW) are tested. We also investigate the human detection accuracy model by two-dimensional curve-fitting in Fig. 5, Eq. (1) can be approximated via regression as:

$$\begin{aligned} P_{A,Q}(a,q) = &\left(1 - 0.2865 \exp\left(-3.3934 \times 10^{-4} \cdot a\right)\right) \\ &\cdot \left(-0.0016 \cdot 2^{q/6} + 0.6762\right). \end{aligned} \tag{2}$$

The encoding data rate model function $r(q)$ can also be represented as a function of q. In this paper, we adopt a simple

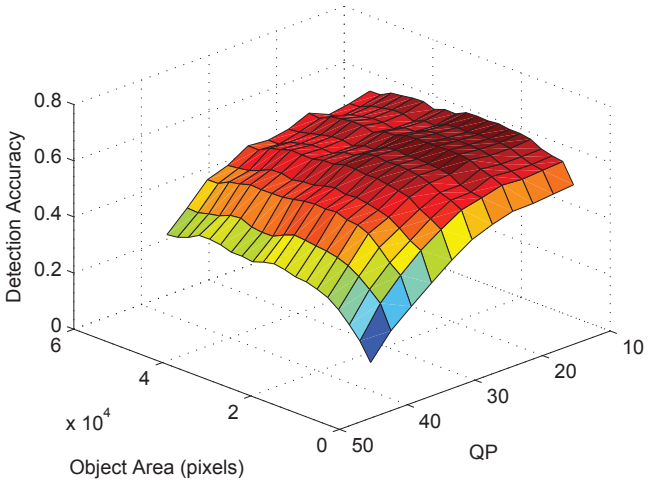

Fig. 4: Human detection accuracy with different object areas and QPs.

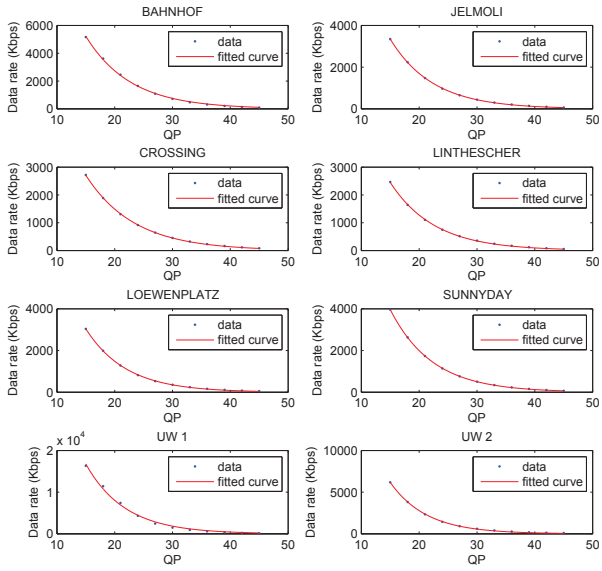

Fig. 5: Curve-fitting results of the source encoding data rate model in Eq. (2) with different videos of VGA and 720p resolutions.

exponential model to fit the source coding rate with respect to QP, i.e.,

$$r(q) = c_1 \exp(c_2 \cdot q), \tag{3}$$

where $c_1 \geq 0$ and $c_2 \leq 0$ are two parameters to be determined. Figure 5 shows the relationship between different QP and source coding rate using HEVC encoder.

IV. Proposed Video Encoding Rate Allocation Scheme

Since wireless video streaming is bandwidth consuming, and the overall wireless resource is limited in WWAN, it is crucial to design an efficient rate allocation scheme so that the true-positive detection result is maximized under a certain total data rate constraint. Therefore, the objective of our proposed system is to optimally allocate the video encoding data rate of each mobile node under a total data rate constraint so that the overall true-positive detection probability is maximized, i.e.,

$$\begin{aligned} \max_{\mathbf{r}} \quad & \prod_{m=1}^{M}\prod_{n=1}^{N_m} P\left(a_{m,n}, q_m\left(r_m\right)\right) \\ \text{subject to} \quad & \sum_{m=1}^{M} r_m \le R^{(\mathrm{T})}; r_m \ge R^{(\min)}, \forall m, \end{aligned} \tag{4}$$

where M is the total number of mobile nodes. $r = [r_1, r_2, \cdots, r_M]$ is the rate allocation vector and the element r_m represents the corresponding source coding rate of the mobile node m. N_m is the number of objects (people in human detection scenario) in the view of mobile node m. $R^{(\mathrm{T})}$ is the total available data rate of the system. $R^{(\min)}$ is the minimum data rate requirement so that the minimum detection capability can be maintained for each mobile node. By taking the logarithm of the objective function, the optimization problem in Eq. (4) can be reformulated as:

$$\begin{aligned} \max_{\mathbf{r}} \quad & \sum_{m=1}^{M} N_m \log\left(g\left(q_m\left(r_m\right)\right)\right) + \sum_{m=1}^{M}\sum_{n=1}^{N_m} \log\left(f\left(a_{m,n}\right)\right) \\ \text{subject to} \quad & \sum_{m=1}^{M} r_m \le R^{(\mathrm{T})}; r_m \ge R^{(\min)}, \forall m. \end{aligned} \tag{5}$$

In Eq. 5, the second term of the objective function can be considered as constant since the optimization variable only appears in the first term. Therefore, we remove the second term so that the final problem formulation is:

$$\begin{aligned} \max_{\mathbf{r}} \quad & \sum_{m=1}^{M} N_m \log\left(-0.0016 \cdot 2^{\frac{1}{6 \cdot c_2^{(m)}} \log\left(\frac{r_m}{c_1^{(m)}}\right)} + 0.6762\right) \\ \text{subject to} \quad & \sum_{m=1}^{M} r_m \le R^{(\mathrm{T})}; r_m \ge R^{(\min)}, \forall m. \end{aligned} \tag{6}$$

Note that in our problem formulation, the optimal solution of the source coding rate allocation is affected by human density indicated by N_m. The objective function in Eq. (6) can be proven as a convex function [26] (see Appendix A). Since the constraint is linear, the optimization problem in Eq. (6) becomes a convex optimization problem, which can be effectively solved by existing tools such as CVX [27]. In our implementation, the resource allocation is updated in every GoP time period and the human density N_m is determined by human detection results in the last GoP time period.

TABLE I: Video Resolutions and Human Densities

Video	Resolution	Human Density
UW 1	1280×720	Low
UW 2	1280×720	Medium
LINTHESCHER	640×480	High

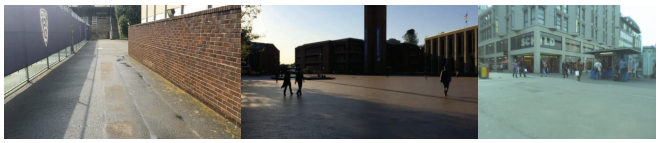

Fig. 6: The sample frames of the three videos. Left: "UW 1"; Middle: "UW 2"; Right: "LINTHESCHER".

V. Simulation Results

The proposed algorithm is tested in this section. Three video clips are used to compete for the limited wireless resources: one video "LINTHESCHER" from the ETHZ data set [25] and two videos recorded in UW campus. The resolutions and human densities of the three videos are listed in Table I. HEVC (X265 implementation) [24] is used as the video encoder. The frame rate of each video is set as 25 fps. GoP sizes are set as 16 for all the videos. The encoding pattern in each GoP block is one I-frame followed by 15 P-frames. 25 GoPs (400 frames) are tested for each video. The sample video frames of the three videos are shown in Fig. 6

We compare our proposed QoC-driven rate allocation scheme with two other schemes. One is the equal rate allocation scheme, which evenly allocates the total data rate to each mobile node. The other scheme is a distortion-driven rate allocation scheme, which tries to minimize the decoding mean-squared-error (MSE) of the system. We adopt a rate-distortion model as [28]:

$$d_m\left(r\right) = c_3^{(m)} r^{c_4^{(m)}}, \tag{7}$$

where d_m is the distortion in terms of MSE for the mobile node m, while $c_3^{(m)}$ and $c_4^{(m)}$ are two constants to be determined. The MSE-driven rate allocation problem can be expressed as:

$$\begin{aligned} \min_{\mathbf{r}} \quad & \sum_{m=1}^{M} d_m\left(r_m\right) \\ \text{subject to} \quad & \sum_{m=1}^{M} r_m \le R^{(\mathrm{T})}; r_m \ge R^{(\min)}, \forall m. \end{aligned} \tag{8}$$

In the simulations, the minimum data rate requirement $R^{(\min)}$ for our proposed QoC-driven scheme and the MSE-driven rate allocation scheme are both set as 200 Kbps.

Figure 7 shows the source coding rate allocation of these 3 videos when the total data rate constraint is 4.8 Mbps. With the MSE-driven rate allocation scheme, the data rate is allocated based on the distortion of each video, which is not directly related to human detection results. However, with the proposed QoC-driven rate allocation scheme, more data rate is allocated to the mobile nodes with higher human densities. Therefore,

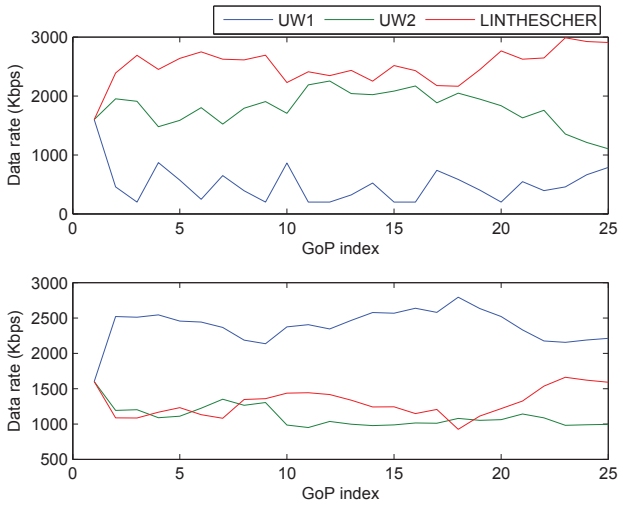

Fig. 7: Data rate allocation of the 3 videos with the proposed QoC-driven data rate allocation scheme (top) and the MSE-driven data rate allcoation scheme (bottom). Total data rate constraint: 4.8 Mbps.

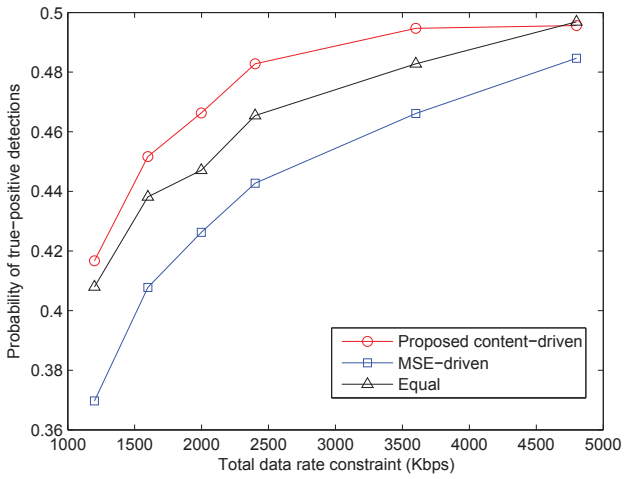

Fig. 8: Probability of true-positive human detections under different total data rate constraints.

the data rate of the video clip LINTHESCHER is higher than that of UW 2 and the data rate assigned to UW 1 is the lowest.

The probabilities of total true-positive detections with different total data rate constraints are plotted in Fig. 8. With more available data rate, the video encoding qualities become better, resulting in improving the true-positive detection rates at the cloud server. Moreover, with the same total data rate constraint, the proposed QoC-driven data rate allocation scheme has better human detection performance comparing with the equal data rate allocation scheme and the MSE-driven data rate allocation scheme. It is noticeable that the MSE-driven data rate allocation scheme has worse human detection performance than the equal data rate allocation scheme. This indicates that transmitting videos based on distortions (decoding qualities) may not be a suitable choice

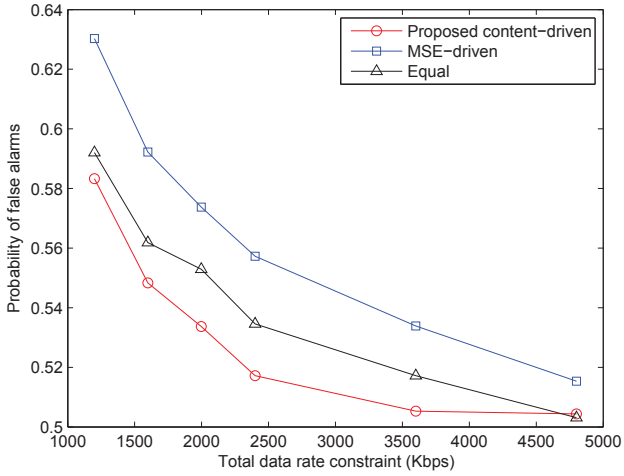

Fig. 9: False-alarm rate under different total data rate constraints.

if the delivered videos are used for video analysis other than human perception. Also, the performance gain of the proposed QoC-driven scheme becomes less when the total available data rate becomes higher. This is because of less video quality degradation with higher encoding data rate.

The human detector may generates some false-alarm detections (i.e., no human exists in the region of bounding box given by human detectors), which will cause problems for subsequent video analysis techniques based on human detections, such as human tracking, behavior understanding, etc.. Therefore, false-alarm is another performance indicator for human detections. Figure 9 shows the probability of false-alarms under different total data rate constraints. Obviously, the false-alarm rate becomes smaller when more data rate is available and high-quality videos are decoded at the cloud server. With the same total data rate constraint, the proposed QoC-driven data rate allocation scheme has the lowest false-alarm rate.

The videos of human detection results are available at http://allison.ee.washington.edu/xchen/MMSP_QoC/

VI. Conclusions

In this paper, we proposed a QoC-driven rate allocation scheme for video analytics purposes in mobile surveillance network with multiple moving cameras. Unlike the traditional wireless video transmission design for human perception, our proposed scheme tries to maximize the human detection rate. The DPM object detector is used for human detection and its accuracy model with respect to object area and video quality is given. Our proposed rate allocation scheme can be formulated as a convex optimization problem, which can be efficiently solved by existing solvers. Simulation results show the effectiveness of our proposed scheme and its favorable performance comparing with equal rate allocation and MSE-driven rate allocation schemes.

Plenty of future works can be conducted in both computer

vision and video transmission areas. In computer vision area, effects of video compression and transmission errors on existing video analysis and computer graphics technologies such as object detection and tracking, pose and event recognitions, 3-D scene reconstructions etc. can be investigated. While in video transmission area, it is necessary to develop noval video coding and transmissions schemes, which can preserve the required features (e.g., [29]) for existing computer vision technologies. As more and more videos are transmitted for video analysis purposes, we believe that combining wireless video transmission and computer vision techniques contains rich research topics and is crucial for next generation mobile networks based on the Internet of things (IoT) and the big data.

APPENDIX A
CONVEXITY OF THE OBJECTIVE FUNCTION IN EQ. (6)

Let $f_1(x)$ be defined as:

$$f_1(x) = \frac{1}{6 \cdot c_2} \log\left(\frac{x}{c_1}\right), \tag{9}$$

which is convex with respect to x if c_2 is non-positive, and $f_2(x)$ is defined as:

$$f_2(x) = -0.0016 \cdot 2^x + 0.6762, \tag{10}$$

which is concave and non-increasing with respect to x. By the composition rule [26], $f_3(x) = f_2(f_1(x))$ is concave. Similarly, since $f_4(x) = \log(x)$ is concave and non-decreasing, $f_5(x) = f_4(f_3(x))$ is also concave by the composition rule. Also, N_m is the detection result of mobile node m, which is non-negative. Therefore, the objective function of Eq. (6) is a non-negative sum of concave functions $f_5(r_m)$, which is also concave [26].

REFERENCES

[1] J.-N. Hwang, *Multimedia Networking: From Theory to Practice*. Cambridge University Press, 2009.

[2] X. Chen, J.-N. Hwang, P.-H. Wu, H.-J. Su, and C.-N. Lee, "Adaptive mode and modulation coding switching scheme in MIMO multicasting system," in *Proc. of IEEE Intl. Symp. on Circuits and Systems*, Beijing, China, May 19-23 2013.

[3] "Cisco Visual Networking Index: Forecast and Methodology, 2014-2019," 2015.

[4] X. Chen, J.-N. Hwang, J. A. Ritcey, and C.-N. Lee, "Quality-driven joint rate and power adaptation for scalable video transmissions over MIMO systems," *submitted to IEEE Trans. on Circuits and Systems for Video Technologies*, 2015.

[5] P.-H. Wu, C.-W. Huang, J.-N. Hwang, J. young Pyun, and J. Zhang, "Video-quality-driven resource allocation for real-time surveillance video uplinking over OFDMA-based wireless networks," *IEEE Trans. on Vehicular Tech.*, pp. 3233–3246, 2014.

[6] X. Chen, J.-N. Hwang, C.-Y. Wang, and C.-N. Lee, "A near optimal QoE-driven power allocation scheme for SVC-based video transmissions over MIMO systems," in *Proc. of IEEE Intl. Conf. on Communications*, Sydney, NSW, June 10-14 2014.

[7] X. Chen, J.-N. Hwang, C.-N. Lee, and S.-I. Chen, "A near optimal QoE-driven power allocation scheme for scalable video transmissions over MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 1, pp. 76–88, 2015.

[8] X. Chen, J.-N. Hwang, C.-J. Wu, S.-R. Yang, and C.-N. Lee, "A QoE-based APP layer scheduling scheme for scalable video transmissions over Multi-RAT systems," in *Proc. of IEEE Intl. Conf. on Communications*, London, UK, 2015.

[9] X. Chen, H. Du, J.-N. Hwang, J. A. Ritcey, and C.-N. Lee, "A QoE-driven FEC rate adaptation scheme for scalable video transmissions over MIMO systems," in *Proc. of IEEE Intl. Conf. on Communications*, London, UK, 2015.

[10] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *Network, IEEE*, vol. 24, no. 2, pp. 36–41, 2010.

[11] A. K. Moorthy, K. Seshadrinathan, R. Soundararajan, and A. C. Bovik, "Wireless video quality assessment: A study of subjective scores and objective algorithms," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 20, no. 4, pp. 587–599, 2010.

[12] K.-H. Lee, J.-N. Hwang, and S.-I. Chen, "Model-based vehicle localization based on three-dimensional constrained multiple-kernel tracking," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 25, no. 1, pp. 38–50, 2015.

[13] M.-C. Chuang, J.-N. Hwang, K. Willianms, and R. Towler, "Tracking live fish from low-contrast and low-frame-rate stereo videos," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 25, no. 1, pp. 167–179, 2015.

[14] K.-H. Lee, J.-N. Hwang, G. Okopal, and J. Pitton, "Driving recorder based on-road pedestrian tracking using visual SLAM and constrained multiple-kernel," in *Proc. IEEE International Conf. Intelligent Transportation System (ITSC)*, 2014, pp. 2629–2635.

[15] L. Hou, W. Wan, K.-H. Lee, J.-N. Hwang, G. Okopal, and J. Pitton, "Deformable multiple-kernel based human tracking using a moving camera," in *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2015.

[16] K.-H. Lee and J.-N. Hwang, "On-road pedestrian tracking across multiple driving recorders," *IEEE Trans. on Multimedia*, vol. 17, no. 9, pp. 1429–1438, 2015.

[17] B. Girod, V. Chandrasekhar, D. M. Chen, N.-M. Cheung, R. Grzeszczuk, Y. Reznik, G. Takacs, S. S. Tsai, and R. Vedantham, "Mobile visual search," *IEEE Signal Processing Magazine*, vol. 28, no. 4, pp. 61–76, 2011.

[18] A. Redondi, M. Cesana, and M. Tagliasacchi, "Rate-accuracy optimization in visual wireless sensor networks," in *Proc. of IEEE International Conference on Image Processing*, 2012, pp. 1105–1108.

[19] S. Milani, R. Bernardini, and R. Rinaldo, "A saliency-based rate control for people detection in video," in *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013, pp. 2016–2020.

[20] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[21] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2005, pp. 886–893.

[22] B. Leibe, A. Leonardis, and B. Schiele, "Robust object detection with interleaved categorization and segmentation," *International journal of computer vision*, vol. 77, no. 1-3, pp. 259–289, 2008.

[23] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1627–1645, 2010.

[24] The X265 website. [Online]. Available at http://bitbucket.org/multicoreware/x265/wiki/home.

[25] A. Ess, B. Leibe, K. Schindler, and L. V. Gool, "A mobile vision system for robust multi-person tracking," in *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2008, pp. 1–8.

[26] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[27] M. Grant and S. Boyd. CVX: MATLAB software for disciplined convex programming. [Online]. Available at http://stanford.edu/~boyd/cvx.

[28] Y.-H. Huang, T.-S. Ou, P.-Y. Su, and H. H. Chen, "Perceptual rate-distortion optimization using structural similarity index as quality metric," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 20, no. 11, pp. 1614–1624, 2010.

[29] J. Chao, R. Huitl, E. Steinbach, and D. Schroeder, "A novel rate control framework for sift/surf feature preservation in h. 264/avc video compression," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 25, no. 6, pp. 958–972, 2014.