A Quality-of-Content (QoC)-based Joint Source and Channel Coding for Human Detections in A Mobile Surveillance Cloud

Xiang Chen, Jenq-Neng Hwang, De Meng, Kuan-Hui Lee, Ricardo L. de Queiroz, Fu-Ming Yeh

***Abstract*—More than 70 percent of all the consumer mobile Internet traffic will be mobile video transmissions by 2019. The development of wireless video transmission technologies have been boosted by the rapidly increasing demand of video streaming applications. Although more and more videos are delivered for video analysis (e.g., object detection/tracking, action recognition, etc.), most of existing wireless video transmission schemes are developed to optimize human perception quality, and are sub-optimal for video analysis. In mobile surveillance networks, a cloud server collects videos from multiple moving cameras and detects suspicious persons in all the camera views. Camera mobility in smartphones or dash cameras, implies that video is to be uploaded through bandwidth-limited and error-prone wireless networks, which may cause quality degradation of the decoded videos and jeopardize the performance of video analyses. In this paper, we propose an effective rate allocation scheme for multiple moving cameras in order to improve human detection (content) performance. Therefore, the optimization criterion of the proposed rate allocation scheme is driven by quality-of-content (QoC). Both video source coding and application layer forward error correction (APP-FEC) coding rates are jointly optimized. Moreover, the proposed rate allocation problem is formulated as a convex optimization problem and can be efficiently solved by standard solvers. Plenty of simulations using high efficiency video coding (HEVC) standard compression of video sequences and the deformable part model (DPM) object detector are carried and the results demonstrate the effectiveness and favorable performance of our proposed QoC-driven scheme under different pedestrian densities and wireless conditions.**

***Index Terms*—Quality-of-content (QoC), rate allocation, application-layer forward error correction (APP-FEC), video analysis, human detection, visual surveillance, convex optimization**

I. Introduction

It is predicted that by 2019, 72 percent of all the consumer mobile Internet traffic will be video, up from 55 percent in 2014 [1]. Moreover, the compound annual growth rate (CAGR) of mobile data traffic is predicted to be 57 percent, which is about three times faster than the growth of fixed IP traffic. However, the bandwidth-limited and error-prone nature of wireless communication environements creates challenges for the bandwidth-consuming and delay-sensitive wireless video streaming applications [2]–[4]. High packet loss/error rate, large delay and jitter experienced in wireless networks can cause tremendous quality degradation of received videos.

The exponentially increasing demand of video streaming services has boosted the developement of wireless video transmission technologies [5]. In most of the existing studies, the optimization targets for wireless video delivery are either quality-of-service (QoS) [5]–[7] or quality-of-experience (QoE) [8]–[14]. For QoS-based design, the transmission scheme is optimized for network parameters such as packet loss rate, delay, jitter, etc [15]–[19]. For QoE-based design, QoS measures are mapped to users' perception and experience of decoded videos so that video transmission parameters can be adjusted to improve users' satisfaction [20]. Both subjective and objective video quality measures can be applied to quantify the QoE-based system design [21].

The Internet of Things (IoT) became a new paradigm in modern wireless telecommunications [22]–[24]. Video-based IoT, which integrates image processing, computer vision and network frameworks, is a new challenging scientific research area at the intersection of video and network technologies [25]. Such research areas include surveillance system, automatic behavior analysis, event detection, etc. [25]. In fact, more and more videos are transmitted for video analytics purposes rather than human consumption. In [26], [27], vehicle tracking and identification systems with static surveillance cameras are developed based on robust vehicle model construction. In [28], an unsupervised learning of camera link models for tracking humans across nonoverlapping static cameras is presented. Based on human detectors, pedestrian tracking systems in single moving cameras are proposed [29], [30]. A system of on-road pedestrian tracking across multiple moving cameras is studied in [31]. An occupancy detection and tracking system is developed for automatic monitoring and commissioning of a building [32], where image-based depth sensor and a programmable pan-tilt-zoom (PTZ) cameras are used. An innovative multiple-kernel adaptive segmentation and tracking system is presented in [33], which dynamically controls the decision thresholds of segmentation around the adaptive kernel regions based on the preliminary tracking results to improve

This study is partially supported by 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.

Xiang Chen is with the Tupl Inc., Bellevue, WA 98005, USA (e-mail: xchen28@uw.edu).

Jenq-Neng Hwang is with the Department of Electrical Engineering, University of Washington, Seattle, WA, 98195 USA (e-mail: hwang@uw.edu).

De Meng is with the Department of Electrical Engineering, University of Washington, Seattle, WA, 98195 USA (e-mail: demeng@uw.edu).

Kuan-Hui Lee is with the Department of Electrical Engineering, University of Washington, Seattle, WA, 98195 USA (e-mail: ykhlee@uw.edu).

Ricardo L. de Queiroz is with the Department of Computer Science, Universidade de Brasilia, Brasilia, Brazil (e-mail: queiroz@ieee.org).

Fu-Ming Yeh is with the Broadband Wireless Department, Gemtek Technology Co., Ltd., Taiwan(e-mail: fred_yeh@gemteks.com).

the robustness of tracking. Due to the large computational cost of video analysis applications and the inter-cooperating properties among multiple cameras, video sequences are required to be uploaded to a cloud server through wired and/or wireless networks. However, transmitting video optimized for QoS or QoE based on existing designs are no longer optimal if the videos is delivered for video analytics. Therefore, it is necessary to develope more efficient video transmission schemes specifically for video analysis and computer vision applications.

Mobile surveillance network with multiple moving cameras, which have more flexible camera views comparing to traditional surveillance systems with static closed circuit television (CCTV), is introduced for better crime investigations and tragedy prevention [31]. In mobile surveillance systems, videos are recorded by driving recorders (dash cameras) or smartphone cameras, and uploaded to remote cloud servers for further automated video analysis. Due to the mobility nature of moving cameras, wireless wide area networks (WWAN) have to be used for video transmissions. Therefore, efficient video compression, error protection and resource allocation are needed because of the bandwidth-limited and the error-prone nature of wireless networks.

In mobile surveillance networks, in order to recognize and track suspicious people, several video analytics technologies need to be applied such as human detection, human tracking, action recognition, behavior understanding, etc.. Among these technologies, human detection is the first step and the performances of other human-related video analytics applications are critically affected by the human detection results [31]. Although plenty of studies have been conducted to improve the object detector in the computer vision field [34]–[36], few studies can be found in wireless video transmission when the performance degradations caused by video compression and the wireless transmission errors are considered. In [37], a saliency-based rate control system for human detection with a single camera is proposed. To improve the human detection accuracy, this scheme adaptively adjusts the quantization parameters (QPs) to preserve regions with small contrast from excessive smoothing based on a properly designed saliency map. Video encoders can be modified to transmit videos while maintaining useful features for video analytics [38]. In [39], [40], image/video features instead of the full video sequences are uploaded to the cloud servers for video analyses. Although transmitting features can save 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 [41], a video source encoding rate allocation scheme is proposed when multiple moving cameras compete for limited wireless resources. This scheme adequately allocate the wireless resource to each moving camera based on its previous human detection results. Since the optimization objective is to improve human detection performance, this scheme is called a quality-of-content (QoC)-based design.

In this paper, we extend the work in [41] by considering packet losses in wireless networks. Among several techniques of improving the reliability of data transmissions, we adopt the application layer forward error correction (APP-FEC) code since it can provide certain correction capability without retransmissions, which is suitable for delay-sensitive real-time video streaming applications [2], [42], [43]. With APP-FEC, well-designed redundant packets are transmitted along with data packets. By receiving enough subset of data and redundant packets, all of the original data packets can be recovered [42]. The correction capability increases with the number of redundant packets in a coding block. In mobile surveillance networks, each mobile node (moving camera) is allocated a certain amount of wireless resources for video encoding and APP-FEC encoding. Nevertheless, the number of transmitted packets is limited by the total available data rate. With higher/lower source encoding rate, the video analysis performance becomes better/worse with the cost of weeker/stronger APP-FEC protection. Therefore, a QoC-driven joint source coding and APP-FEC rate allocation scheme is proposed in this paper. Instead of considering human perception in traditional video streaming design, the proposed scheme optimizes the human detection performance at the cloud server when multiple mobile nodes compete for the limited wireless resource and upload vidoes via WWAN with a total data rate constraint. The detected human density is used as the content information for the proposed rate allocation algorithm. Ideally, the proposed algorithm allocates more data rates to the mobile nodes located in the places with higher human densities. The proposed scheme can be formulated as a convex optimization problem [44] and can be solved by convex optimization toolbox such as CVX [45]. To the best of our knowledge, there is no QoC-driven work conducted when both video compressions and wireless transmission errors are considered.

The rest of this paper is organized as follows. In Section II, scenario and system structure of the mobile surveillance network is described. In Section III, the effects of source coding rate on the human detection performance are studied. Detailed description about the APP-FEC is given in Section IV. The proposed joint source and channel coding rate allocation scheme is given in Section V. Simulation results are shown in Section VI, followed by the conclusion remarks in Section VII.

Notations: The list of symbols used in this paper is summarized in Table I.

II. Scenarios and System Structure

In a mobile surveillance network, multiple mobile nodes are randomly distributed, moving around in the area with different pedestrian densities as illustrated in Fig. 1. Each mobile node can capture, encode and upload videos via a WWAN to a cloud server for video analyses such as human detection. The proposed system structure is shown in Fig. 2. The captured camera view of each node is encoded with the high efficiency video coding (HEVC) [46] encoder with different encoding data rates. The parameter estimation module determines the necessary parameters required by the cloud server. The encoded packets of video data and corresponding parameters of a single group of pictures (GoP) are then fed into the APP-FEC encoder, which generates certain amount of redundant packets

TABLE I: List of symbols

Symbol	Definition
q	Quantization parameter (QP)
$P(\cdot)$	Human detection accuracy model
n	Application layer forward error correction (APP-FEC) block size
k	Number of source packets in an APP-FEC block
t	Correction capability of an APP-FEC block
ϵ	Reception overhead efficiency of APP-FEC
p	Probability of packet losses
$f(\cdot)$	APP-FEC block correction rate
$\Phi(\cdot)$	Cumulative distribution function of the standard Gaussian distribution
M	Total number of mobile nodes in the system
$T^{(\mathrm{GoP})}$	Group of pictures (GoP) time period
S	Packet size
N_m	Number of detected people in the view of the mobile node m
$R^{(\mathrm{T})}$	Total data rate target of the system
$R^{(\min)}$	Minimum data rate requirement of each mobile node
$\mathbb{N}$	Set of natural numbers
$c_m^{(1)}, c_m^{(2)}$	Parameters in QP-rate model of the mobile node m
$r_m^{(\mathrm{S})}$	Source coding rate target of the mobile node m
$r_m^{(\mathrm{B})}$	Transmission data rate target of the mobile node m
$c_m^{(3)}, c_m^{(4)}$	Parameters in a rate-distortion model of the MSE-driven rate allocation scheme in the control group.
$d_m(\cdot)$	Distortion measured by the MSE in a rate-distortion model of the MSE-driven rate allocation scheme in the control group.
$P^{(\mathrm{req})}$	FEC block correction rate requirement of the MSE-driven rate allocation scheme in the control group.

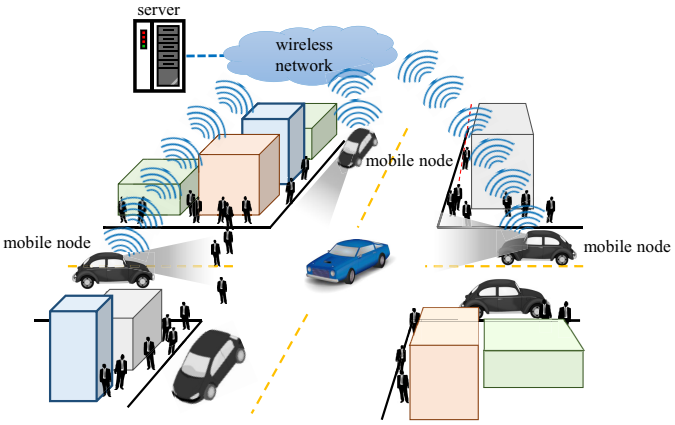


Fig. 1: Scenario of mobile surveillance network.

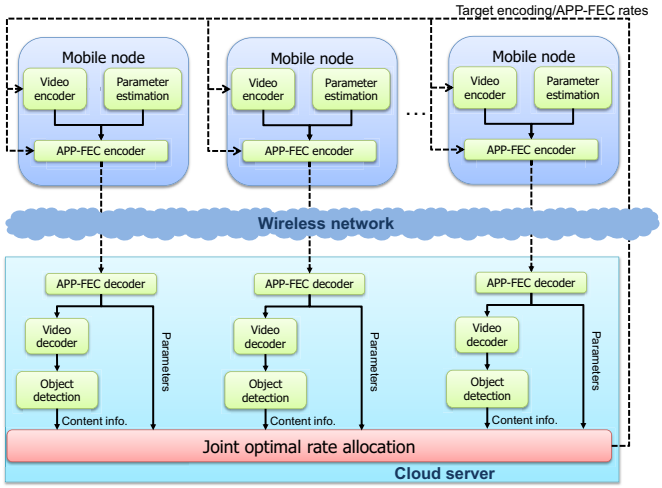


Fig. 2: Proposed system structure.

and forms an APP-FEC block [47], [48]. The concept of GoP-level APP-FEC encoding is shown in Fig. 3. The cloud server receives the transmitted APP-FEC block with some packet loss due to either wireless transmission errors or network congestion. In this paper, the packet loss rate of each mobile node is assumed to be perfectly known by a standard real-time protocol (RTP) and real-time control protocol (RTCP) [49], [50]. The APP-FEC decoder decodes the APP-FEC block and fed the video packets to the video decoder. The undecodable video frames are dropped. The video decoder can conceal the lost video frames by copying the last successfully decoded video frame. After decoding the video, an object detection module performs the human detections and sends the detection results (content information) to the rate allocation module. Since the pedestrian density is different in the view of each mobile node, the human detection result (content information) is therefore different. Based on the content information and the necessary parameters delivered by the mobile nodes, the rate allocation module jointly optimizes the source coding rate and the APP-FEC coding rate for each mobile node under a predetermined total data rate constraint, which is assumed to be affordable by the wireless network. The rate allocation result is then fed back to the mobile nodes for the video encoding and transmission of the next GoP. Note that the rate allocation results are targets for the mobile nodes and the actural source coding and APP-FEC rates can be slightly different to the targets.

III. Human Detection and the Effect of Video Quality on Human Detection Performance

Plenty of object detection schemes have been developed in existing literatures. A human detector based on the histogram of oriented gradient (HOG) feature, which can effectively represent the shape of human, is proposed in [34]. In [35], the implicit shape model (ISM) is proposed, which 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. In [36], authors extended the idea in [34] and proposed the deformable part model (DPM), which uses a root and several part models to describe different partitions of an object. The part models are spatially connected with the root model based on a predefined geometry so that the object can be accurately depicted. Among these object detection methods, the DPM is a well-accepted, robust and computationally efficient scheme. Therefore, the DPM is adopted as the human detection scheme in this paper.

Since the DPM object detector is based on the HOG feature, it can be affected by the artifacts created from

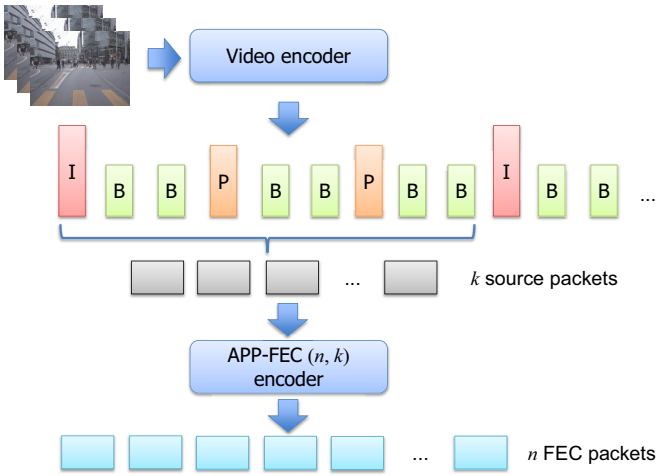


Fig. 3: GoP-level APP-FEC encoding [47].

(a) Video clip: LOEWENPLATZ in the ETHZ dataset [51]. Left: QP=15; Right: QP=39

(b) Video clip: BAHNHOF in the ETHZ dataset [51]. Left: QP=15; Right: QP=39

Fig. 4: Comparisons of Human Detection Performance of the DPM with Different Video Encoding Qualities.

video encoders at different compression ratios [37]. In mobile surveillance systems, human detection is performed in the cloud server based on compressed videos, which may affect the human detection performance. A comparison example of the DPM detection results are shown in Fig 4. The video clips "LOEWENPLATZ" and "BAHNHOF" in the ETHZ dataset [51] are encoded by two different QPs. When the QP is large, the artifact caused by higher compression ratio significantly distorts the original HOG information, which leads to noticeable human detection performance degradation.

Figure 5 shows the human detection accuracy with respect to different video encoding qualities in terms of the QPs. Six video clips in the ETHZ dataset [51] are encoded by the HEVC encoder [52] with 11 different QPs from 15 to 45. The detection results are compared to the ground-truth

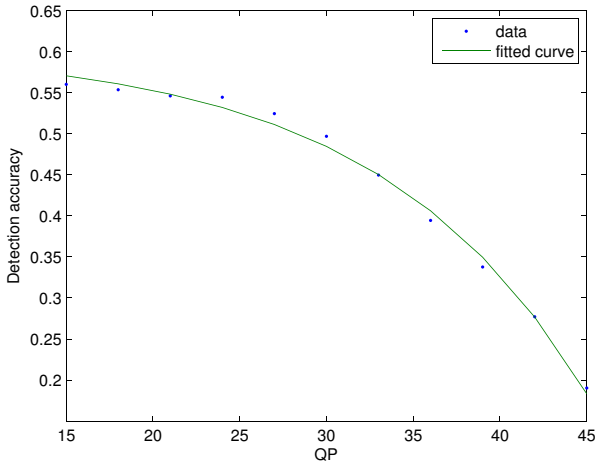


Fig. 5: Human detection accuracy with different QPs. The solid line is based on a human detection accuracy model given in Eq. (1). The RMSE of this curve-fitting is 1.104%

labels provided in the ETHZ dataset. An object is considered as correct detection if the overlapped area of the detection result and the corresponding ground-truth is larger than 50 percent of the ground-truth area [36]. Obviously, the higher the quality (lower QP), the higher the detection accuracy. But the rate of increasing becomes small when the video quality is high enough since the artifacts created by video compression are negligible. By adopting curve-fitting, the human detection accuracy model as a function of QP (i.e., q) can be approximated as:

$$P(q) = -0.0098 \cdot 2^{0.1206 \cdot q} + 0.6049. \quad (1)$$

The solid line in Fig. 5 is based on Eq. (1). The root mean square error (RMSE) of the curve-fitting in Fig. 5 is 1.104%.

IV. APP-FEC and Block Correction Rate

APP-FEC can provide reliable end-to-end streaming applications with packet level protection [43]. Suppose a source block with k packets is encoded into an APP-FEC block consisting of $n\,(n \geq k)$ packets by adding $n-k$ redundant packets. The APP-FEC rate of an (n,k) code is defined as the k/n. An ideal APP-FEC code, with correction capability $t = n-k$, can reconstruct the original k source packets from any k out of n received packets [43]. The Reed-Solomon (RS) code, which operates on non-binary symbols, is a well-known APP-FEC scheme with ideal correction capability. However, the block size of the RS code is constrained by the symbol size, which limits the flexibility of parameter selections in practice. For instance, in the most commonly used RS code with 8 bits per symbol, the total number of packets in an APP-FEC block is constrained by 255, i.e., $k \leq n \leq 255$. Furthermore, the non-binary operations of the RS code may cause high computational complexity in software implementation [43]. Such drawbacks make the RS code unattractive to high definition (HD) video streaming applications. With the development of APP-FEC, more practical schemes have been introduced. One

of the attractive APP-FEC solutions for HD video streaming services is the Raptor code [53], which has flexible parameters selection and linear decoding cost. Unlike the RS code, the correction capability of the Raptor code is nonideal, given by $t = n-(1+\epsilon)\,k$, where ϵ is the reception overhead efficiency. Nevertheless, the reception overhead efficiency of standardized Raptor code is very low and close to ideal correction capability [43]. In this paper, the APP-FEC scheme with ideal correction capability is assumed for simplicity.

An APP-FEC block can be successfully decoded if the number of packet loss/error is not more than the correction capability t. If the packet losses are independent and identically distributed (*i.i.d.*), the block correction rate (BCR) of an (n,k) ideal APP-FEC code is given by a cumulative distribution function (CDF) of a binomial distribution [54]:

$$f\left(t,n;p\right)=\sum_{j=0}^{t}\binom{n}{j}p^{j}\left(1-p\right)^{n-j}, \tag{2}$$

where p is the packet loss rate. When n is sufficiently large, np and $n\left(1-p\right)$ are much greater than 1, the CDF of a binomial distribution can be approximated by a CDF of Gaussian distribution with mean np and variance $np\left(1-p\right)$ [55]. Therefore, Eq. (2) can be approximated as:

$$f\left(t,n;p\right)\approx\Phi\left(\frac{t-np}{\sqrt{np\left(1-p\right)}}\right)-\Phi\left(\frac{-np}{\sqrt{np\left(1-p\right)}}\right), \tag{3}$$

where $\Phi\left(\cdot\right)$ is the CDF of the Gaussian distribution with 0 mean and unit variance. If $n\cdot p$ is large enough, The second term in Eq. (3) is approximately 0, and the BCR in Eq. (2) can be further approximated as:

$$\begin{aligned} f\left(k,n;p\right) &\approx\Phi\left(\frac{n-k-np}{\sqrt{np\left(1-p\right)}}\right)\\ &=\Phi\left(\frac{\sqrt{n}\left(1-p\right)-\frac{k}{\sqrt{n}}}{\sqrt{p\left(1-p\right)}}\right). \end{aligned} \tag{4}$$

When the APP-FEC block is generated every GoP time period (e.g., $0.5-1$ second), the APP-FEC block size n is normally large enough with the available data rate to transmit HD videos in modern wireless network (e.g., 10 Mbps) and a reasonable packet size (e.g., 600 bytes). With a typical packet loss rate in wireless networks without any other protection or retransmission schemes (e.g., $0.1\%-2\%$), the approximation in Eq. (4) is valid.

V. Proposed Joint Source Coding and APP-FEC Rate Allocation Scheme

Since the wireless resoures are limited, multiple moving nodes have to compete for the total available data rate. Moreover, for each moving node, part of the allocated data rate needs to be used for APP-FEC redundant packet in order to protect the source packets from possible losses. Therefore, it is necessary to design an efficient rate allocation scheme for both video and APP-FEC encodings. Unlike previous studies, which optimize the video decoding quality, the objective of our proposed scheme is to maximize the human detection performance at the cloud server. Suppose M is the number of mobile nodes in the system, $\mathbf{k}$ is an $M\times 1$ vector with each element k_m representing the number of source packets of the mobile node m in an APP-FEC block, $\mathbf{n}$ is an $M\times 1$ vector with each element n_m representing the APP-FEC block size of the mobile node m, and N_m is the total number of detected objects of the mobile node m in the last GoP time period, the proposed scheme maximizes the overall true-positive human detection probability under a total data rate constraint and minimum data rate requirements for each moving node, i.e.,

$$\begin{aligned} \max_{\mathbf{k},\mathbf{n}}\quad & \prod_{m=1}^{M}\left(P\left(q_m\left(\frac{k_m\cdot S}{T^{(\mathrm{GoP})}}\right)\right)f\left(k_m,n_m;p_m\right)\right)^{N_m}\\ \text{subject to}\quad & \sum_{m=1}^{M}\frac{n_m\cdot S}{T^{(\mathrm{GoP})}}\le R^{(\mathrm{T})},\\ & \frac{k_m\cdot S}{T^{(\mathrm{GoP})}}\ge R^{(\min)},\forall m,\\ & n_m\ge k_m,\forall m,\\ & \mathbf{k},\mathbf{n}\in\mathbb{N}^{M\times 1}, \end{aligned} \tag{5}$$

where $T^{(\mathrm{GoP})}$ is the GoP time period, $R^{(\mathrm{T})}$ is the total available data rate, S is the packet size. $R^{(\min)}$ is the minimum data rate requirement so that the minimum detection capability can be reached for each mobile node. The minimum allocated data rates for the mobile nodes are independent of the previous detection results so that the system is more robust to detection failures caused by wireless transmission errors. $P\left(\cdot\right)$ is the human detection accuracy in Eq. (1). $f\left(\cdot\right)$ is the APP-FEC BCR in Eq. (4). The QP of the mobile node m is denoted as $q_m\left(\cdot\right)$, which is a function of source encoding rate $r_m^{(\mathrm{S})}=k_m\cdot S/T^{(\mathrm{GoP})}$. In this paper, we adopt the following model to fit the QP with respect to the source coding rate, i.e.,

$$q_m\left(r^{(\mathrm{S})}\right)=\frac{1}{c_m^{(2)}}\log\left(\frac{r^{(\mathrm{S})}}{c_m^{(1)}}\right), \tag{6}$$

where $c_m^{(1)}\ge 0$ and $c_m^{(2)}\le 0$ are two model parameters to be estimated by the parameter estimation module in Fig. 2. Figure 6 illustrates the source coding rate vs. QP curves with the HEVC encoder. In total 6 videos with VGA (640×480) resolution in the ETHZ dataset [51] and 2 videos with 720p (1280×720) resolution recorded in the University of Washington (UW) are tested. And the QP vs. source coding rate model can be accurately fitted with the measurements by properly adjusting the two parameters.

The first constraint in Eq. (5) means that in one GoP time period, the sum of the transmission data rate of each mobile node is limited by the total system data rate. The second constraint indicates that the source coding data rate of each mobile node should not be smaller than a pre-defined minimum data rate requirement. Please note that even though human detection is considered as the only video analytics application in this paper, similar idea can be applied to other applications by replacing the human detection accuracy model $P\left(\cdot\right)$ in Eq. (1) with other video analytics performance models. The optimization problem in Eq. (5) is hard to be efficiently solved since $\mathbf{k}$ and $\mathbf{n}$ are positive integers. By eliminating the last constraint in Eq. (5) and taking logarithm of the objective function, the original problem in Eq. (5) can be reformulated

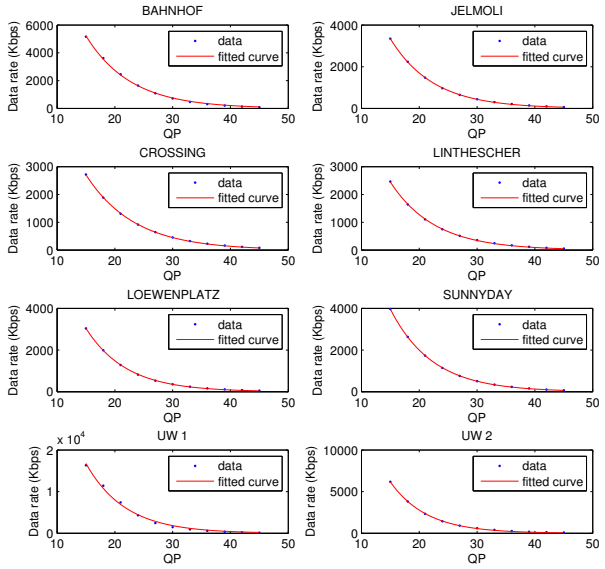


Fig. 6: Curve-fitting result of the model in Eq. (6) with different videos of VGA and 720p resolutions.

as:

$$\begin{aligned}\max_{\mathbf{k},\mathbf{n}} \sum_{m=1}^{M} N_m \left(\log\left(P\left(q_m\left(\frac{k_m S}{T^{(\mathrm{GoP})}}\right)\right)\right)+\log\left(f(k_m,n_m;p_m)\right)\right)\\ \text{subject to} \quad \sum_{m=1}^{M} n_m \le \frac{R^{(\mathrm{T})}T^{(\mathrm{GoP})}}{S},\\ k_m \ge \frac{R^{(\min)}T^{(\mathrm{GoP})}}{S}, \forall m,\\ n_m \ge k_m, \forall m.\end{aligned} \tag{7}$$

By applying the APP-FEC BCR approximation in Eq. (4) and substituting the optimization variables $\tilde{n}_m = \sqrt{n_m}$ and $\tilde{k}_m = \sqrt{k_m}$ for all m, The optimization problem becomes:

$$\begin{aligned}\max_{\tilde{\mathbf{k}},\tilde{\mathbf{n}}} \sum_{m=1}^{M} N_m \cdot \log\left(P\left(q_m\left(\frac{\tilde{k}_m^2 \cdot S}{T^{(\mathrm{GoP})}}\right)\right)\right)\\ + \sum_{m=1}^{M} N_m \log\left(\Phi\left(\frac{\tilde{n}_m(1-p_m) - \frac{\tilde{k}_m^2}{\tilde{n}_m}}{\sqrt{p_m(1-p_m)}}\right)\right)\\ \text{subject to} \quad \sum_{m=1}^{M} \tilde{n}_m^2 \le \frac{R^{(\mathrm{T})}T^{(\mathrm{GoP})}}{S},\\ \tilde{k}_m \ge \sqrt{\frac{R^{(\min)}T^{(\mathrm{GoP})}}{S}}, \forall m,\\ \tilde{n}_m \ge \tilde{k}_m, \forall m,\end{aligned} \tag{8}$$

which can be shown as a convex optimization problem [44] (see Appendix A) and can be solved by convex optimization tools such as CVX [45]. In the objective function of Eq. (8), the first line is for source coding rate allocation and the second line is for APP-FEC rate allocation. The rate allocation of each mobile node is also proportional to the number of detected people N_m, which reflects the pedestrian density in the camera view of the m^{th} mobile node. Note that the optimized solutions $\mathbf{k}^*$ and $\mathbf{n}^*$ do not need to be converted to integers since the source coding rate $r_m^{(\mathrm{S})} = k_m^* \cdot S/T^{(\mathrm{GoP})}$ and the transmission data rate $r_m^{(\mathrm{B})} = n_m^* \cdot S/T^{(\mathrm{GoP})}$ of the mobile node m are the corresponding encoding targets for the next round, which may slightly differ from the encoding rates due to the rate-control mechanism of the encoder. However, the numbers of transmitted source and the total packets (including both source coding and redundancy packets) are chosen as the largest integers smaller than the coresponding source coding and transmission rates respectively.

TABLE II: Video Resolutions and Human Densities

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

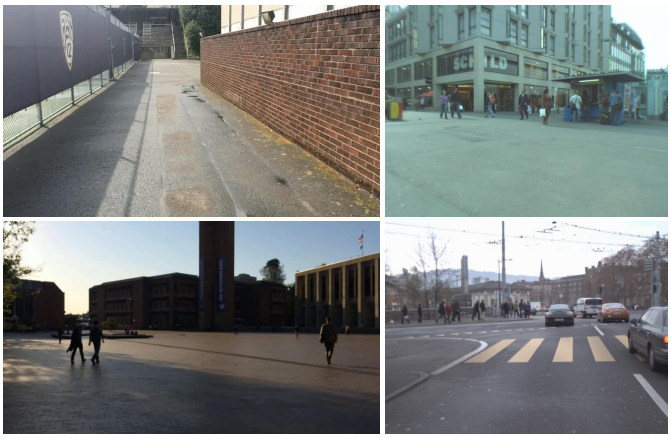


Fig. 7: The sample frames of the four videos. Top left: "UW 1"; Top right: "LINTHESCHER"; Bottom left: "UW 2"; Bottom right: "LOEWENPLATZ".

VI. Simulation Results

In this section, the proposed QoC-driven rate allocation scheme is evaluated by plenty of simulations. In total four video clips are used to compete for the limited wireless resources: two videos "LINTHESCHER" and "LOEWENPLATZ" from the ETHZ data set [51] and two videos recorded in the UW campus. Table II summarizes the resolutions and pedestrian densities of the four videos. The HEVC (X265 implementation) [52] is used as the video encoder. The frame rate and GoP are set as 25 fps and 16 respectively for all the videos. Hence the GoP period ($T^{(\mathrm{GoP})}$) is 0.64 s. The encoding pattern in each GoP block is one I-frame followed by 15 P-frames. 25 GoPs (400 frames) are simulated for each video. Figure 7 shows sample video frames of the four videos. The packet size (S) is set as 600 bytes.

Table III summarizes the proposed and other rate allocation schemes in the control group. We compare the proposed scheme with other three schemes. The first one is driven by video distortion measured by the mean-squared-error (MSE).

TABLE III: Simulated rate allocation schemes

Rate Allocation Scheme	Source Coding Rate	APP-FEC Rate
Proposed	Adaptive, QoC-driven	Adaptive, QoC-driven
MSE	Adaptive, MSE-driven	Adaptive, driven by pre-determined targets
Equal (0.5)	Fixed, 50% of total available throughput	Fixed, 50% of total available throughput
Equal (0.8)	Fixed, 80% of total available throughput	Fixed, 20% of total available throughput

We adopte a rate-distortion model in [56] as:

$$d_m(r) = c_m^{(3)} r^{c_m^{(4)}}, \tag{9}$$

where $d_m(r)$ is the distortion in terms of MSE for the mobile node m. $c_m^{(3)}$ and $c_m^{(4)}$ are two parameters to be determined by curve-fitting. The MSE-driven rate allocation problem is expressed as:

$$\begin{aligned} \min_{\tilde{\mathbf{k}},\tilde{\mathbf{n}}} \quad & \sum_{m=1}^{M} d_m\left(\frac{\tilde{k}_m^2 \cdot S}{T^{(\mathrm{GoP})}}\right) \\ \text{subject to} \quad & \sum_{m=1}^{M} \tilde{n}_m^2 \le \frac{R^{(\mathrm{T})} T^{(\mathrm{GoP})}}{S} \\ & \tilde{k}_m \ge \sqrt{\frac{R^{(\min)} T^{(\mathrm{GoP})}}{S}}, \forall m, \\ & \tilde{n}_m \ge \tilde{k}_m, \forall m \\ & \log \Phi\left(\frac{\tilde{n}_m(1-p_m) - \frac{\tilde{k}_m^2}{\tilde{n}_m}}{\sqrt{p_m(1-p_m)}}\right) \ge \log(P^{(\mathrm{req})}), \forall m, \end{aligned} \tag{10}$$

where $P^{(\mathrm{req})}$ is a pre-defined BCR requirement, which is set as 0.99 [47] in this paper. The MSE-driven rate allocation problem is also convex (similar proof as Eq. (8)) and can be solved by the CVX [45]. The minimum data rate requirement $R^{(\min)}$ in our proposed scheme and the MSE-driven rate allocation scheme are both set as 200 Kbps.

In the first simulation scenario, the packet loss rates of all the testing video sequences are set to 1%. When the packet loss rates are the same for all the videos, the only cause of data rate allocations is either content or distortion for the proposed content-driven scheme or the traditional distortion-driven scheme, respectively. Figure 8 shows the instantaneous source coding data rate (Kbps) of each video when the total transmission data rate constraint is 2800 Kbps. For the proposed content-driven scheme, more source coding data rates are allocated to the video sequences with higher pedestrian densities. Note that the rate allocation algorithm of the proposed content-driven scheme depends on the detected pedestrian densities of the previous GoP. For instance, the data rate allocated to the video sequence "LOEWENPLATZ" significantly increases from the 11$^{\text{th}}$ GoP to the 13$^{\text{th}}$ GoP. This is because of the increase of detected pedestrian densities as illustrated in Fig. 9. For the distortion-driven scheme, more data rate is allocated to the video sequence with more frame details (i.e., UW 1). The full video sequences with detection results are available at http://allison.ee.washington.edu/xchen/TCSVT-QoC/ .

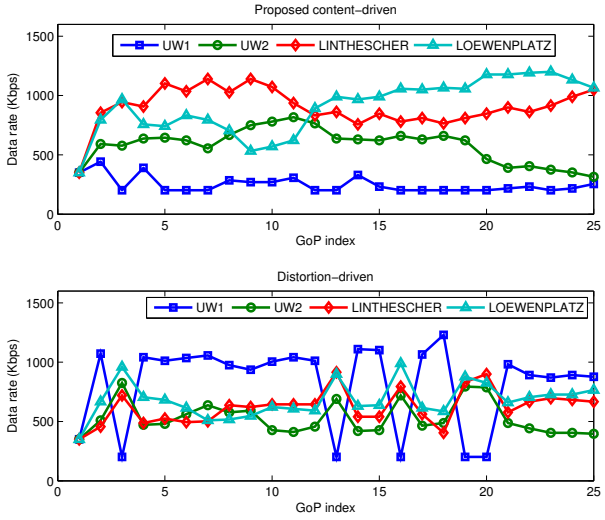


Fig. 8: Instantaneous source coding data rate (Kbps) of each video under total transmission data rate: 2800 Kbps. Packet loss rate of each video is 1%.

Fig. 9: The sample frames of the video sequence "LOEWENPLATZ" of different pedestrian densities. Left: a sample video frame of the 10$^{\text{th}}$ GoP; Right: a sample video frame of the 12$^{\text{th}}$ GoP.

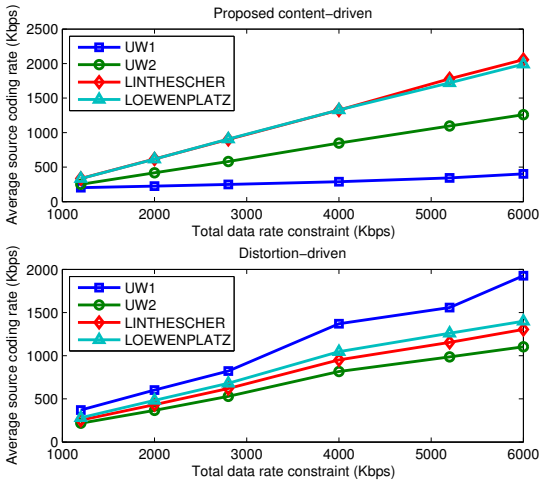


Fig. 10: Average source coding data rate (Kbps) of each video under different transmission data rate constraints. Packet loss rate of each video is 1%.

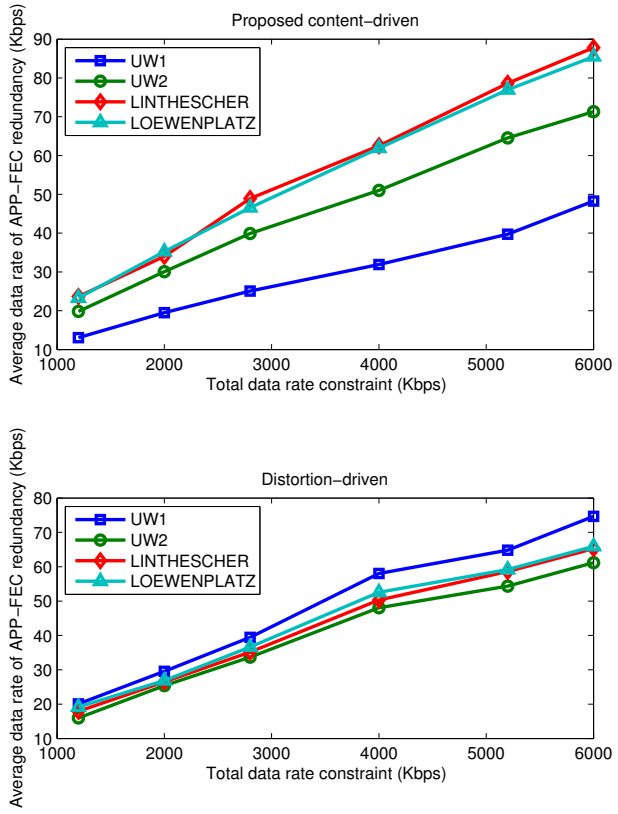


Fig. 11: Average data rate (Kbps) of APP-FEC redundancy under different transmission data rate constraints. Packet loss rate of each video is 1%.

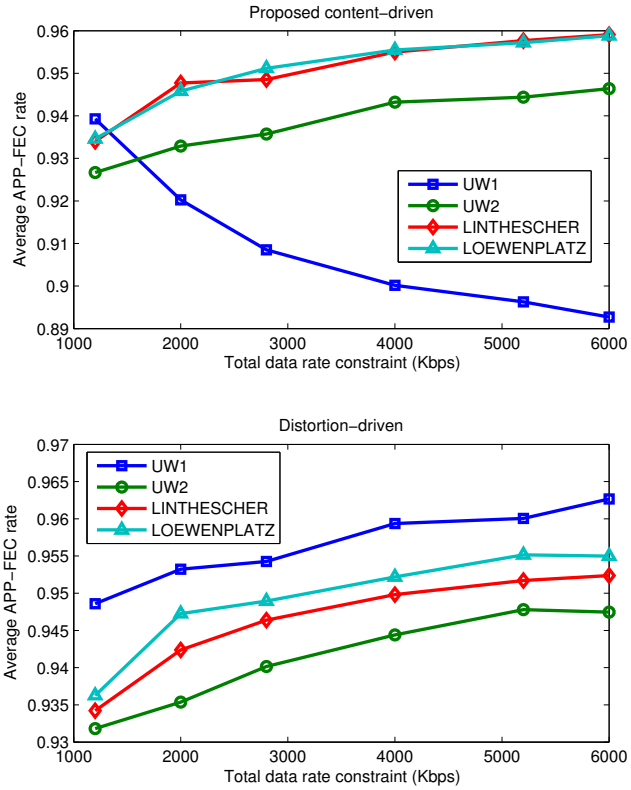


Fig. 12: Average APP-FEC rate of each video under different transmission data rate constraints. Packet loss rate of each video is 1%.

Figure 10 shows the average source coding data rate (Kbps) of each video sequence under different transmission data rate constraints. It is clear that larger source coding data rate is allocated to all the video sequences when the total data rate constraint increases. For the proposed content-driven scheme, the data increasing rates (slopes) of the videos with higher pedestrian densities (e.g., "LOEWENPLATZ") are higher than those with less pedestrian densities (e.g., "UW 1"). While for the distortion-driven scheme, the data increasing rates do not depend on the pedestrian densities. For instance, the data rate allocated to "UW 1" are higher than the other vidoes even though "UW 1" has the lowest human density. As shown in Fig. 11, more data rates are allocated to APP-FEC redundancies for better protections of the transmitted videos when the overall transmission data rate constraint increases. Similar to the source coding rate allocation, the proposed content-driven scheme allocates more FEC protections to the video sequences with more pedestrian densities, which is not the case for the distortion-driven scheme. The average APP-FEC rates of each video under different total transmission data rate constraints are plotted in Fig. 12. Most of the FEC rates slightly increase with available data rate for transmissions since the increase of source coding rates are higher than that of the APP-FEC redundancies. However, for the proposed content-driven

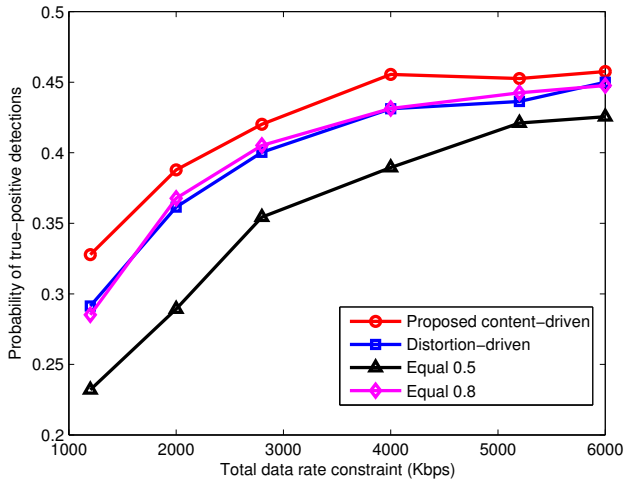


Fig. 13: Average true-positive detection probability under different transmission data rate constraints. Packet loss rate of each video is 1%.

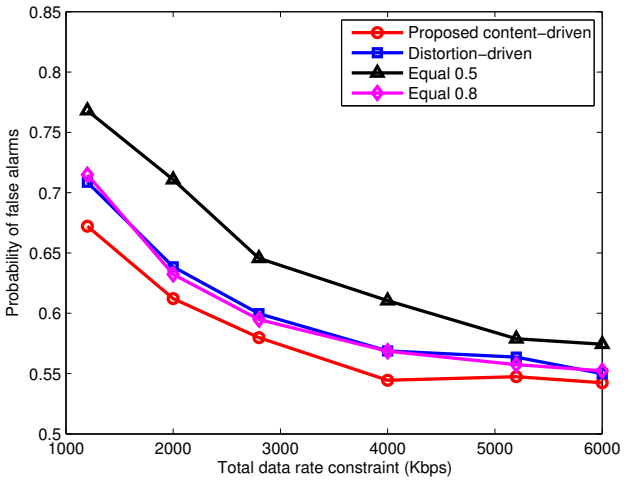


Fig. 14: Average false-alarm probability under different transmission data rate constraints. Packet loss rate of each video is 1%.

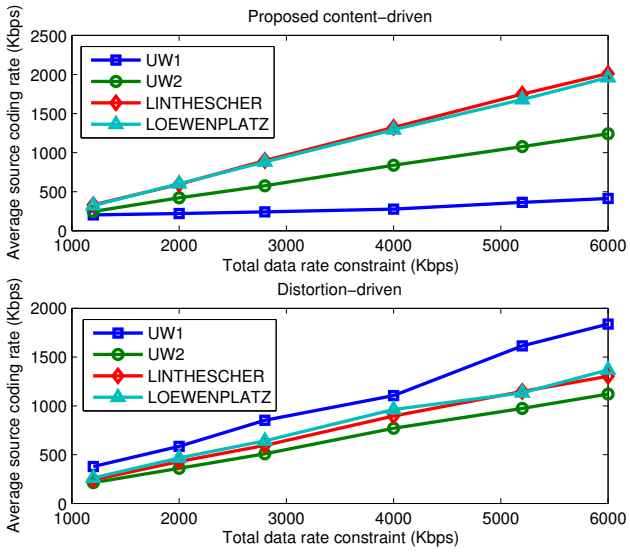


Fig. 15: Average source coding data rate (Kbps) of each video under different transmission data rate constraints. Packet loss rates of video "UW 1" and "LOEWENPLATZ" are 3%, and that of video "UW 2" and "LINTHESCHER" are 1%.

scheme, the APP-FEC rate of the video sequence "UW 1" has a decreasing trend, which is due to the fact that few source coding data rate is allocated even when more total data rate is available (shown in Fig. 10). This is because of the smaller pedestrian densities in the video scene.

The average true-positive detection probability and the average false-alarm probability for each rate allocation scheme are plotted in Fig. 13 and Fig. 14 respectively. The proposed content-driven scheme more effectively allocates both of the source coding rate and the APP-FEC rate and can achive higher true-positive detection probability and less false-alarms.

In the second simulation scenario, the packet loss rates of the video sequences "UW 1" and "LOEWENPLATZ" are set as 3% while that of the other two are set as 1%. Figure 15 shows the source coding rates of each video. Still, the proposed scheme allocates more data rates to the videos with higher pedestrian densities. The average data rates of APP-FEC redundancy are plotted in Fig. 16. For the proposed content-driven scheme, comparing to Fig. 11, more APP-FEC redundancy is allocated to the video sequence "LOEWENPLATZ", which has higher packet loss rate (3%) than the other video sequences (1%). However, the APP-FEC redundancy allocated to the video sequence "UW 1" is not significantly increased even though its packet loss rate is also 3%. This is because the pedestrian density of "UW 1" is low, and the proposed content-driven scheme allocates more data rates to the videos with more contents (i.e, pedestrian densities). Figure 17 shows the average APP-FEC rate of each video. Note that the APP-FEC rate of the video sequence "LOEWENPLATZ" is much lower than that in Fig. 12 because more redundancy is allocated. The average true-positive detection probability and the false-alarm probability are plotted in Fig. 18 and Fig. 19 respectively. It can be easily noted that the proposed content-driven scheme has better performance than the other schemes.

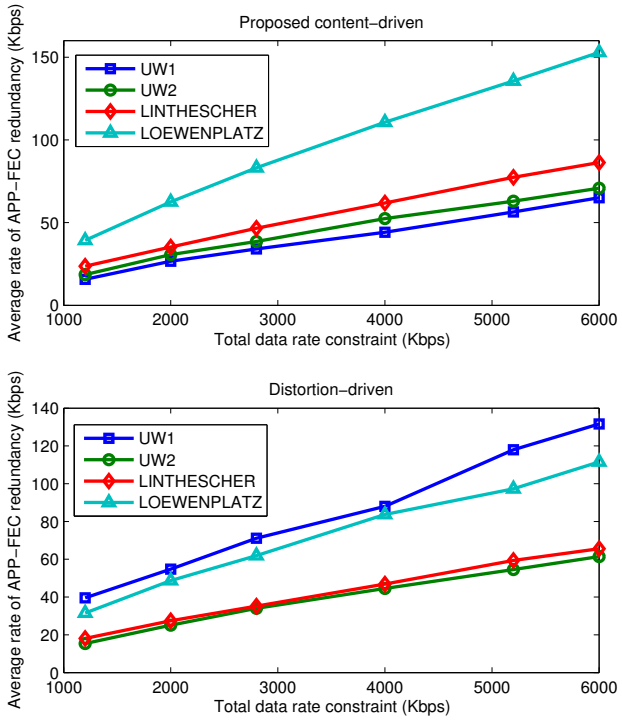


Fig. 16: Average data rate (Kbps) of APP-FEC redundancy under different transmission data rate constraints. Packet loss rates of video "UW 1" and "LOEWENPLATZ" are 3%, and that of video "UW 2" and "LINTHESCHER" are 1%.

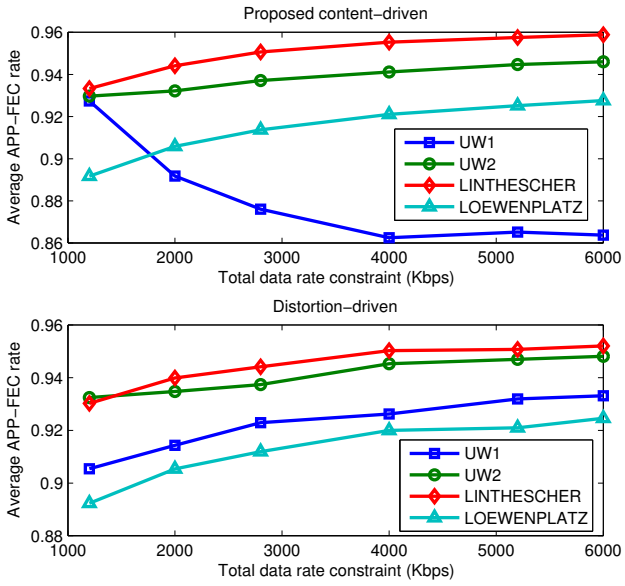


Fig. 17: Average APP-FEC rate of each video under different transmission data rate constraints. Packet loss rates of video "UW 1" and "LOEWENPLATZ" are 3%, and that of video "UW 2" and "LINTHESCHER" are 1%.

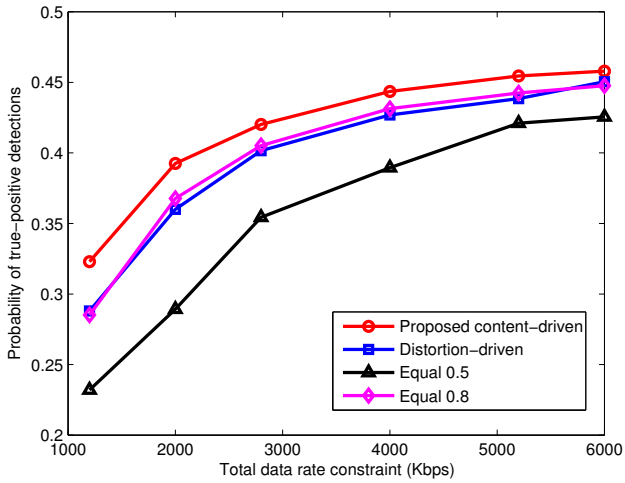


Fig. 18: Average true-positive detection probability under different transmission data rate constraints. Packet loss rates of video "UW 1" and "LOEWENPLATZ" are 3%, and that of video "UW 2" and "LINTHESCHER" are 1%.

VII. Conclusions and Future Work

In this paper, a QoC-driven joint source coding and APP-FEC rates allocation scheme for video analysis purposes in mobile surveillance network with multiple moving cameras is proposed. Different to the previous wireless video transmission studies, which focus on improving traditional QoS or QoE measures to meet the wireless network conditions or the users' perception satisfactions, the proposed scheme tries to optimize the wireless resource usage so that more accurate human detections can be performed at the cloud server based

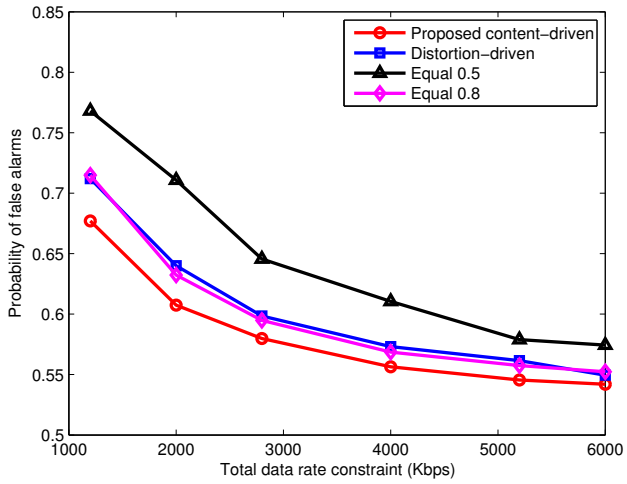


Fig. 19: Average false-alarm probability under different transmission data rate constraints. Packet loss rates of video "UW 1" and "LOEWENPLATZ" are 3%, and that of video "UW 2" and "LINTHESCHER" are 1%.

on the received videos. This study is conducted with HEVC video codec and the DPM object detector. We have evaluated the human detection model with different QPs of the HEVC encoder. The APP-FEC is used for more reliable wireless data deliveries. Also, our proposed joint source encoding and APP-FEC rate allocation problem can be formulated as a convex optimization problem, which can be efficiently solved by standard solvers. Plenty of simulations with different pedestrian densities and wireless conditions show the effectiveness of our proposed scheme and its favorable performance comparing to the equal rate allocation and MSE-driven rate allocation schemes.

The proposed scheme only considers human detection as the video analytics purpose, which is the first step for more sophisticated systems such as human tracking, behavior understanding etc. Therefore, plenty of furture studies can be conducted in computer vision, video compression and video transmission areas. In computer vision area, more robust video analytics and computer graphics technologies, including object detection/tracking, pose and event recognitions, 3-D scene reconstructions etc., are required when video compression and transmission errors exist. In video compression, traditional designs are developed to improve the rate-distortion properties so that more bit rate can be saved while keeping as much as the original video quality. However, more efficient video compression schemes could be developed if video quality is sacrificed instead of more useful features for video anayltics. In wireless video transmission, transmission protocols and strategies can be re-evaluated if the video sequences are transmitted for video analytics rather than human perceptions.

Appendix A
Convexity of the Objective Problem in Eq. (8)

For the first half of the objective function in Eq. (8), $q_m(.)$ is convex and non-increasing if $c_m^{(2)} \leq 0$.

Therefore, $q_m\left(\tilde{k}_m^2 \cdot S/T^{(\text{GoP})}\right)$ is convex by the composition rule [44]. Since $P(\cdot)$ is concave and decreasing, $P\left(q_m\left(\tilde{k}_m^2 \cdot S/T^{(\text{GoP})}\right)\right)$ is concave by the composition rule. Also, $\log(\cdot)$ is concave and increasing, $\log\left(P\left(q_m\left(\tilde{k}_m^2 \cdot S/T^{(\text{GoP})}\right)\right)\right)$ is therefore concave. For the second half of the objective function in Eq. (8), $g\left(\tilde{k}_m, \tilde{n}_m\right) = \tilde{k}_m^2/\tilde{n}_m$ is a quadratic-over-linear function, which is convex if $\tilde{n}_m > 0$ [44].

$$h\left(\tilde{k}_m, \tilde{n}_m\right) = \frac{\tilde{n}_m\left(1-p_m\right) - \frac{\tilde{k}_m^2}{\tilde{n}_m}}{\sqrt{p_m\left(1-p_m\right)}} \tag{11}$$

is concave. Furthermore, $\Phi(\cdot)$ is log-concave [44] and non-decreasing. $\log\left(\Phi\left(h\left(\tilde{k}_m, \tilde{n}_m\right)\right)\right)$ is concave. Since N_m is non-negative, the objective function in Eq. (8) is non-negative sums of concave functions, which is also concave [44]. Therefore, the optimization problem in Eq. (8) is a convex optimization problem since the objective function is concave and the feasible set determined by all the constraints is convex.

REFERENCES

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

[2] X. Chen, J.-N. Hwang, J. A. Ritcey, C.-N. Lee, and F.-M. Yeh, "Quality-driven joint rate and power adaptation for scalable video transmissions over MIMO systems," *IEEE Trans. on Circuits and Systems for Video Technology*, 2016.

[3] X. Chen, J.-N. Hwang, C.-N. Lee, and C.-W. Hwang, "An efficient CQI feedback resource allocation scheme for wireless video multicast services," in *Proc. of IEEE Global Telecommunications Conf.*, Atlanta, GA, December 9-13 2013, pp. 1663–1668.

[4] 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, pp. 441–444.

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

[6] S. Ehsan and B. Hamdaoui, "A survey on energy-efficient routing techniques with QoS assurances for wireless multimedia sensor networks," *IEEE Communications Surveys & Tutorials*, vol. 14, no. 2, pp. 265–278, 2012.

[7] S. Koli, R. Purandare, S. Kshirsagar, and V. Gohokar, "QoS-optimized adaptive multi-layer (OQAM) architecture of wireless network for high quality digital video transmission," *Journal of Visual Communication and Image Representation*, vol. 26, pp. 210–221, 2015.

[8] 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.

[9] 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, pp. 1675–1680.

[10] 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, pp. 6953–6958.

[11] R. Schatz, T. Hoßfeld, L. Janowski, and S. Egger, "From packets to people: quality of experience as a new measurement challenge," in *Data traffic monitoring and analysis*. Springer, 2013, pp. 219–263.

[12] E. Yaacoub, F. Filali, and A. Abu-Dayya, "QoE enhancement of SVC video streaming over vehicular networks using cooperative LTE/802.11 p communications," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 1, pp. 37–49, 2015.

[13] X. Chen, "Quality-driven cross layer design of video transmissions over mimo systems," Ph.D. dissertation, University of Washington.

[14] J. Park, X. Chen, and J.-N. Hwang, "Optimum power allocation and rate adaptation for scalable video streaming over multi-user MIMO networks," in *Proc. of IEEE Global Telecommunications Conf.*, San Diego, CA, December 6-10 2015.

[15] C. Wei, H. Chen, M. Song, M.-T. Sun, and K. Lau, "A capture-to-display delay measurement system for visual communication applications," in *Proc. of Signal and Information Processing Association Annual Summit and Conference*, Kaohsiung, Taiwan, 2013, pp. 1–4.

[16] H. Chen, C. Wei, M. Song, M.-T. Sun, and K. Lau, "Capture-to-display delay measurement for visual communication applications," *APSIPA Trans. on Signal and Information Processing*, vol. 4, 2015.

[17] H. Chen, C. Zhao, M.-T. Sun, and A. Drake, "Adaptive intra-refresh for low-delay error-resilient video coding," *Journal of Visual Communication and Image Representation*, vol. 31, pp. 294–304, 2015.

[18] Y. Yang and S. Roy, "PMU deployment for optimal state estimation performance," in *Proc. of IEEE Global Communications Conf.*, Anaheim, CA, December 3-7 2012, pp. 1464–1468.

[19] ——, "PMU deployment for three-phase optimal state estimation performan," in *Proc. of IEEE Intl. Conf. on Smart Grid Communications*, Vancouver, BC, October 21-24 2013, pp. 342–347.

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

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

[22] L. Atzori, A. Iera, and G. Morabito, "The Internet of things: A survey," *Computer networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[23] Y. Yang and S. Roy, "Grouping based MAC protocols for EV charging data transmission in smart metering network," *IEEE Journal on Selected Areas in Communications*, vol. 49, no. 7, pp. 1328–1343, 2014.

[24] ——, "PCF scheme for periodic data transmission in smart metering network with cognitive radio," in *Proc. of IEEE Global Communications Conf.*, San Diego, CA, December 6-10 2015.

[25] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (IoT): A vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[26] 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.

[27] H.-Y. Wang and H.-C. Shih, "A robust vehicle model construction and identification system using local feature alignment," in *Proc. of IEEE International Symposium on Consumer Electronics (ISCE)*. IEEE, 2013, pp. 57–58.

[28] C.-T. Chu and J.-N. Hwang, "Fully unsupervised learning of camera link models for tracking humans across nonoverlapping cameras," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 24, no. 6, pp. 979–994, 2014.

[29] 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.

[30] 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*, 2015.

[31] K.-H. Lee and J.-N. Hwang, "On-road pedestrian tracking across multiple driving recorders," *submitted to IEEE Trans. on Multimedia*, 2015.

[32] H.-C. Shih, "A robust occupancy detection and tracking algorithm for the automatic monitoring and commissioning of a building," *Energy and Buildings*, vol. 77, pp. 270–280, 2014.

[33] Z. Tang, J.-N. Hwang, Y. Lin, and J. Chuang, "Multiple-kernel adaptive segmentation and tracking (mast) for robust object tracking," in *Proc. of IEEE Intl. Conf. on Acoustics, Speech and Signal Processing*, Shanghai, China, March 20-25 2016.

[34] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection." IEEE, 2005, pp. 886–893.

[35] 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.

[36] 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.

[37] 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*, 2013, pp. 2016–2020.

[38] 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.

[39] 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.

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

[41] X. Chen, J.-N. Hwang, K.-H. Lee, and R. L. de Queiroz, "Quality-of-content (QoC)-driven rate allocation for video analysis in mobile surveillance networks," in *Proc. of IEEE Intl. Workshop on Multimedia Signal Processing*, 2015.

[42] D. Jurca, P. Frossard, and A. Jovanovic, "Forward error correction for multipath media streaming," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 19, no. 9, pp. 1315–1326, 2009.

[43] M. Luby, T. Stockhammer, and M. Watson, "Application layer FEC in IPTV services," *IEEE Communications Magazine*, vol. 46, no. 5, pp. 94–101, 2008.

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

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

[46] 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.

[47] J. Wu, Y. Shang, J. Huang, X. Zhang, B. Cheng, and J. Chen, "Joint source-channel coding and optimization for mobile video streaming in heterogeneous wireless networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, pp. 1–16, 2013.

[48] E. Baccaglini, T. Tillo, and G. Olmo, "Slice sorting for unequal loss protection of video streams," *Signal Processing Letters, IEEE*, vol. 15, pp. 581–584, 2008.

[49] C. Perkins, *RTP Audio and Video for the Internet*. Addison Wesley, 2003.

[50] 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, pp. 6779–6784.

[51] A. Ess, B. Leibe, K. Schindler, and L. V. Gool, "A mobile vision system for robust multi-person tracking," 2008, pp. 1–8.

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

[53] A. Shokrollahi, "Raptor codes," *IEEE Trans. on Information Theory*, vol. 52, no. 6, pp. 2551–2567, 2006.

[54] C.-W. Huang, S.-M. Huang, P.-H. Wu, S.-J. Lin, and J.-N. Hwang, "OLM: Opportunistic layered multicasting for scalable IPTV over mobile WiMAX," *IEEE Trans. on Mobile Computing*, vol. 11, no. 3, pp. 453–463, 2012.

[55] H. Stark and J. Woods, *Probability, Statistics, and Random Processes for Engineers*. Prentice Hall, 2011.

[56] 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.

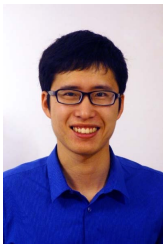
[57] Y. Yang, X. Wang, and X. Cai, "On the number of relays for orthogonalize-and-forward relaying," in *Proc. of IEEE Intl. Conf. on Wireless Communications and Signal Processing*, Nanjing, China, November 9-11 2011, pp. 1–5.

[58] H.-C. Shih and K.-C. Yu, "Spiral aggregation map (SPLAM): A new descriptor for robust template matching with fast algorithm," *Pattern Recognition*, vol. 48, no. 5, pp. 1707–1723, 2015.

[59] Y. Yang, "Contributions to smart metering protocol design and data analytics," Ph.D. dissertation, University of Washington.

[60] H. Chen and B. Zeng, "New transforms tightly bounded by DCT and KLT," *IEEE Signal Processing Letters*, vol. 19, no. 6, pp. 344–347, 2012.

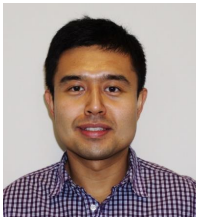
[61] H. Chen, Y. Chen, M.-T. Sun, A. Saxena, and M. Budagavi, "Improvements on intra block copy in natural content video coding," in *Proc. of IEEE Intl. Symp. on Circuits and Systems*, 2015.



Xiang Chen received the B.S. degree in electronic engineering from City University of Hong Kong, Hong Kong, in 2009 and the M.S. degree in electrical and computer engineering from University of Florida, Gainesville, in 2011, and the Ph.D. degree in electrical engineering from University of Washington in 2015. He is currently with Tupl Inc., a startup big data company for telecom operators. His research interests include multimedia networking, wireless communication and MIMO techniques.

Jenq-Neng Hwang (F'01) received the B.S. and M.S. degrees, both in electrical engineering from the National Taiwan University, Taipei, Taiwan, in 1981 and 1983 separately. He then received his Ph.D. degree from the University of Southern California. In the summer of 1989, Dr. Hwang joined the Department of Electrical Engineering of the University of Washington in Seattle, where he has been promoted to Full Professor since 1999. He is currently the Associate Chair for Research in the EE Department. He has written more than 300 journal, conference papers and book chapters in the areas of multimedia signal processing, and multimedia system integration and networking, including an authored textbook on "Multimedia Networking: from Theory to Practice," published by Cambridge University Press. Dr. Hwang has close working relationship with the industry on multimedia signal processing and multimedia networking.

Dr. Hwang received the 1995 IEEE Signal Processing Society's Best Journal Paper Award. He is a founding member of Multimedia Signal Processing Technical Committee of IEEE Signal Processing Society and was the Society's representative to IEEE Neural Network Council from 1996 to 2000. He is currently a member of Multimedia Technical Committee (MMTC) of IEEE Communication Society and also a member of Multimedia Signal Processing Technical Committee (MMSP TC) of IEEE Signal Processing Society. He served as associate editors for IEEE T-SP, T-NN and T-CSVT, T-IP and Signal Processing Magazine (SPM). He is currently on the editorial board of ETRI, IJDMB and JSPS journals. He was the Program Co-Chair of ICASSP 1998 and ISCAS 2009.



De Meng received the B.S. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2009 and the M.S. degree in applied mathematics in 2014 from the University of Washington, Seattle, WA, where he is currently pursuing the Ph.D. degree in electrical engineering. His research interests include convex optimization, distributed and online optimization, machine learning and data analysis.

Kuan-Hui Lee received the B.S. degree in the Department of Electrical Engineering from National Taiwan Ocean University in 2003, and the M.S. degree in the Institute of Computer and Communication Engineering from National Cheng Kung University in 2005. He has been in HTC Corporation for developing multi-media applications on smart phone from 2007 to 2009. In 2015, he received his Ph.D. degree in the Department of Electrical Engineering from University of Washington. His current research interests are in computer vision, image processing, and machine learning.

Dr. Ricardo L. de Queiroz received the Engineer degree from Universidade de Brasilia , Brazil, in 1987, the M.Sc. degree from Universidade Estadual de Campinas, Brazil, in 1990, and the Ph.D. degree from The University of Texas at Arlington , in 1994, all in Electrical Engineering.

In 1990-1991, he was with the DSP research group at Universidade de Brasilia, as a research associate. He joined Xerox Corp. in 1994, where he was a member of the research staff until 2002. In 2000-2001 he was also an Adjunct Faculty at the Rochester Institute of Technology. He joined the Electrical Engineering Department at Universidade de Brasilia in 2003. In 2010, he became a Full Professor (Professor Titular) at the Computer Science Department at Universidade de Brasilia. He was a Visiting Professor at the University of Washington, in Seattle, during 2015. Dr. de Queiroz has published over 160 articles in Journals and conferences and contributed chapters to books as well. He also holds 46 issued patents. He is a past elected member of the IEEE Signal Processing Society's Multimedia Signal Processing (MMSP) and the Image, Video and Multidimensional Signal Processing (IVMSP) Technical Committees. He is a an editor for IEEE Transactions on Image Processing and a past editor for the EURASIP Journal on Image and Video Processing, IEEE Signal Processing Letters, and IEEE Transactions on Circuits and Systems for Video Technology. He has been appointed an IEEE Signal Processing Society Distinguished Lecturer for the 2011-2012 term.

Dr. de Queiroz has been actively involved with the Rochester chapter of the IEEE Signal Processing Society, where he served as Chair and organized the Western New York Image Processing Workshop since its inception until 2001. He helped organizing IEEE SPS Chapters in Brazil. He was the General Chair of ISCAS'2011, and MMSP'2009, and is the General Chair of SBrT'2012. He was also part of the organizing committee of ICIP'2002, ICIP'2012, ICIP'2014 and ICIP'2016. His research interests include image and video compression, multirate signal processing, and color imaging. Dr. de Queiroz is a Senior Member of IEEE and a member of the Brazilian Telecommunications Society.

Fu-Ming Yeh received the Ph.D. in 1997 in electrical engineering from National Taiwan University. He is a CTO of Broadband Wireless Department at Gemtek Technology Co., Ltd. He was a deputy head at the Electronic System Research Division of Chung-Shan Research Institute of Science and Technology from 1997 to 2006. His research interests include LTE Small Cell system development, DSP system design, hardware verification, VLSI testing, and fault-tolerant computing.