Spread Spectrum Visual Sensor Network Resource Management Using an End-to-End Cross-Layer Design

Elizabeth S. Bentley, Lisimachos P. Kondi, *Member, IEEE*, John D. Matyjas, Michael J. Medley, *Senior Member, IEEE*, and Bruce W. Suter

***Abstract*—In this paper, we propose an approach to manage network resources for a direct sequence code division multiple access (DS-CDMA) visual sensor network where nodes monitor scenes with varying levels of motion. It uses cross-layer optimization across the physical layer, the link layer, and the application layer. Our technique simultaneously assigns a source coding rate, a channel coding rate, and a power level to all nodes in the network based on one of two criteria that maximize the quality of video of the entire network as a whole, subject to a constraint on the total chip rate. One criterion results in the minimal average end-to-end distortion amongst all nodes, while the other criterion minimizes the maximum distortion of the network. Our experimental results demonstrate the effectiveness of the cross-layer optimization.**

***Index Terms*—Code division multiple access (CDMA), convolutional codes, cross-layer, H.264, joint source-channel coding, multimedia communications, power control, resource allocation, spread spectrum, visual sensor network.**

I. Introduction

In this paper, we consider a direct sequence code division multiple access (DS-CDMA) visual sensor network where we assume that the nodes in the network are equipped with a video camera deployed to survey a large area. Many sensor networks concern themselves with increasing the energy efficiency and maximizing the lifetime of the network as in [1]–[3]. Some visual sensor networks focus on image transmission as in [4], where the trade-off between image quality and energy consumption of different routing paths is considered. Visual sensor networks that transmit video are much more challenging than typical sensor networks due to the high bit rates and delay constraints. In [5], the demanding nature of visual sensor networks is acknowledged and a new wireless sensor node protocol stack is proposed. However, the improvements that can be gained with cross-layer interactions are not considered.

Manuscript received July 10, 2009; revised January 31, 2010 and July 16, 2010; accepted September 17, 2010. Date of publication October 11, 2010; date of current version January 19, 2011. This work was supported by a Marie Curie International Reintegration Grant within the 7th European Community Framework Programme. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Qian Zhang.

E. S. Bentley, J. D. Matyjas, M. J. Medley, and B. W. Suter are with the Air Force Research Laboratory, Rome, NY 13441 USA.

L. P. Kondi is with the Department of Computer Science, University of Ioannina, 45110 Ioannina, Greece.

Digital Object Identifier 10.1109/TMM.2010.2086441

In our set-up, some nodes will be imaging a relatively stationary field while other nodes will be imaging scenes with a high level of motion to create a more realistic scenario where scenes with varying levels of motion exist. Video sequences with less motion can be source encoded at a lower bit rate while still yielding good picture quality. The centralized control unit at the network layer should be able to request that the video at specific nodes be transmitted at a lower bit rate, if it is deemed as being capable of still producing adequate video quality. These nodes that compress their video at a lower bit rate are left with more bits for channel coding and can afford to transmit at a lower power so that they will cause less interference to the other nodes. For this reason, DS-CDMA is an appropriate choice for use in our visual sensor network set-up.

In this work, we present a multi-node cross-layer optimization technique that operates across the physical, data link, and application layers of the system. Our algorithm accounts for network performances all the way from the physical layer up to the application layer. At the application layer, the source coding rate for video compression is determined. At the data link layer, the channel coding rate is selected, and at the physical layer, the transmission power is determined. Our algorithm simultaneously allocates a source coding rate, a channel coding rate, and a power level to all nodes in a DS-CDMA visual sensor network. We propose to jointly optimize all nodes using one of two criteria. Our first criterion results in the minimal average end-to-end distortion over all the nodes in the network while our second criterion minimizes the maximum distortion amongst all nodes. This optimization algorithm uses universal rate distortion characteristics (URDC) to reduce the computational complexity. Zero-mean Gaussian interference is assumed to obtain the probability of error in a channel using Viterbi's upper bound on the probability of error. Some preliminary results appear in [6], and an earlier version of this article was published in [7].

The rest of the paper is organized as follows. In Section II, we describe the transmission parameters in a visual sensor network: the DS-CDMA physical layer in Section II-A, the source coding in Section II-B, and the channel coding in Section II-C. In Section III, the cross-layer optimization algorithm is explained. In Section IV, experimental results are presented, and in Section V, conclusions are drawn.

II. Visual Sensor Networks

Sensor networks previously focused on networks that transmit scalar information such as temperature, pressure, acoustic data, etc. Visual sensor networks are much more challenging due to the high bit rates and delay constraints required for video transmission. These networks are comprised of typically low-weight distributed sensor nodes that can communicate directly (not via intermediate nodes) with a centralized control unit at the network layer. The centralized control unit performs channel and source decoding to obtain the received video from each node. The control unit transmits information to the nodes in order to request changes in transmission parameters, such as source coding rate, channel coding rate, and transmission power. Applications of visual sensor networks include surveillance, automatic tracking and signaling of intruders within a physical area, command and control of unmanned vehicles, and environmental monitoring.

A. DS-CDMA

This work considers a wireless visual sensor network that utilizes DS-CDMA. In DS-CDMA, all users (nodes) transmit on the same frequency. In order to transmit a single bit, a node actually transmits L "chips". Thus, each node k is associated with a spreading code (signature sequence) $\mathbf{s}_k$, which is a vector of length L. Thus, in order to transmit the ith bit of a bit stream, node k actually transmits $b_k(i)\mathbf{s}_k$, which is a vector of L chips and $b_k(i)$ is either 1 or -1, depending on the value of the bit that is being transmitted.

Assuming there are K nodes in a synchronous single-path binary phase shift keying (BPSK) channel, the received signal can be expressed as $\mathbf{r}(i) = A_1 b_1(i)\mathbf{s}_1 + \sum_{k=2}^{K} A_k b_k(i)\mathbf{s}_k + \mathbf{n}_k$, where A_k, $b_k(i)$, $\mathbf{s}_k$, $\mathbf{n}_k$ are the amplitude, symbol stream, signature sequence, and noise of node k, respectively. $\mathbf{r}(i)$, $\mathbf{s}_k(i)$, and $\mathbf{n}_k$ are vectors of length L. DS-CDMA systems are usually interference-limited systems. Thus, it is reasonable to ignore thermal noise and background noise and assume that the interference can be approximated by a zero-mean White Gaussian random process [8]. Since user k has an associated power level in Watts, $S_k = E_k R_k$, the energy-per-bit to multiple-access-interference (MAI) ratio becomes

$$\frac{E_k}{I_0} = \frac{\frac{S_k}{R_k}}{\sum\limits_{j \neq k}^{K} \frac{S_j}{W_t}}; k = 1, 2, 3, \ldots, K \tag{1}$$

where E_k is the energy-per-bit, $I_0/2$ is the two-sided noise power spectral density due to MAI in Watts/Hertz, S_k is the power level of the node-of-interest in Watts, R_k is the transmitted bit rate in bits per second, S_j is the power level of interfering node j in Watts, and W_t is the total bandwidth in Hertz [8]. The term "power level" refers to the power that is received by the centralized control unit. For a given power level, nodes can determine the required transmit power using a propagation model. R_k is taken to be the total bit rate used for source and channel coding. Assuming K users, R_k can be expressed as

$$R_k = \frac{R_{s,k}}{R_{c,k}}; k = 1, 2, 3, \ldots, K \tag{2}$$

where $R_{s,k}$ is the source coding rate for node k and $R_{c,k}$ is the channel coding rate for node k. Since $R_{s,k}$ has units of bits per second and $R_{c,k}$ is a dimensionless number, R_k will be measured in bits per second [9].

Let us also define the vectors $\underline{R_s} = [R_{s,1}, R_{s,2}, \ldots, R_{s,K}]^T$, $\underline{R_c} = [R_{c,1}, R_{c,2}, \ldots, R_{c,K}]^T$, and $\underline{S} = [S_1, S_2, \ldots, S_K]^T$.

B. Source Coding

In our visual sensor network, we assume that the nodes are equipped with video cameras that monitor various fields. We assume that each node has the computational power necessary for video compression. The video captured by the cameras is compressed using the H.264/AVC video coding standard. H.264/AVC has two conceptual layers, the video coding layer (VCL) and the network abstraction layer (NAL). The VCL forms the main part of the H.264/AVC and performs the required tasks for video compression to efficiently represent the content of the video data. The NAL achieves the network-friendly objective of H.264/AVC. It defines the interface between the VCL and the broad variety of systems and transport media [10]. All data are encapsulated in NAL units which contain an integer number of bytes.

C. Channel Coding

In this work, we use rate compatible punctured convolutional (RCPC) codes for channel coding [11]. Using RCPC codes allows us to utilize Viterbi's upper bounds on the bit error probability, P_b, given by

$$P_b \leq \frac{1}{P} \sum_{d=d_{free}}^{\infty} c_d P_d \tag{3}$$

where P is the period of the code, d_{free} is the free distance of the code, c_d is the information error weight, and P_d is the probability that the wrong path at distance d is selected. An AWGN channel with binary phase-shift keying (BPSK) modulation has a P_d given by

$$P_d = Q\left(\sqrt{\frac{2dR_c E_b}{I_0}}\right) \tag{4}$$

where R_c is the channel coding rate and E_b/I_0 is the energy-per-bit to multiple-access-interference ratio, measured in Watts/Hertz.

D. Expected Video Distortion Calculation

Clearly, the expected video distortion for a node should depend on the corresponding bit error rate. In this work, we utilize universal rate-distortion characteristics (URDCs) [9]. These characteristics show the expected distortion as a function of the bit error probabilities, P_b after channel decoding. However, since video encoded with the H.264 codec is designed to handle packet errors as opposed to bit errors, we need to calculate the resulting packet loss rate. We calculate a real-time transport protocol (RTP) packet loss rate (PLR) from a certain bit error rate (BER), drop packets from the H.264 bitstream according to the RTP PLR, and pass the corrupted H.264 bitstream to the H.264 decoder to calculate the distortion of the uncompressed video.

The link layer packet size is LL_{size} (measured in bits). Thus, the link layer PLR is $PLR_{LL} = 1 - (1 - BER)^{LL_{size}}$, where PLR_{LL} is the packet loss rate for a link layer packet of size,

LL_{size}. Similarly, we calculate the RTP packet loss rate with $PLR_{RTP} = 1-(1-PLR_{LL})^{RTP_{size}}$, where PLR_{RTP} is the packet loss rate for an RTP packet of size, RTP_{size} (measured in the number of link layer packets). The RTP provides a packet format for real-time data transmissions [12]. We assume that we know when a packet has an error, and we manually drop packets with any errors from the H.264 encoded video stream, in accordance with the PLR_{RTP} calculated from the BER. We then calculate the distortion of this "corrupted" video stream. This creates the relation between the BER and the distortion of a packet-based video stream with packet errors. As mentioned previously, the bit error rate we are interested in for the URDCs is the bit error rate after channel decoding. Thus, in our case, $P_b = BER$.

Since channel errors are random, the video distortion $D_{s+c,k}$ of node k, which is due to both the lossy compression and channel errors, is a random variable. Thus, it does not suffice to calculate the video distortion for just one realization of the channel. Instead, we will consider the expected value of the distortion, $E[D_{s+c,k}]$.

As in [13] and [14], we assume the following model for the URDC for each user k:

$$E[D_{s+c,k}] = a\left[\log_{10}\left(\frac{1}{P_b}\right)\right]^b \tag{5}$$

where a and b are such that the square of the approximation error is minimized. Thus, instead of calculating the URDCs based on experimental results for every possible P_b, we instead experimentally calculate the expected distortion for a few packet loss rates associated with specific bit error rates, P_b's. We then use the model, given in (5), to approximate the distortion for other bit error rates. The parameters a and b depend on the video sequence and the source coding rate.

The expected distortion $E[D_{s+c,k}]$ for node k is a function of the source and channel coding rates $R_{s,k}$ and $R_{c,k}$, for node k, and the power levels of all nodes S_k, $k = 1,\ldots,K$. This can be seen as follows. From (1), assuming that all users transmit at the same total bit rate (and thus chip rate), the E_k/I_0 for node k is a function of the power levels of all nodes. Parameters c_d and d_{free} depend on the channel coding rate. Thus, from (3) and (4), it follows that the bit error probability P_b for node k is a function of E_k/I_0 and the channel coding rate $R_{c,k}$. Parameters a and b depend on the source coding rate and the encoded video sequence. Therefore, from (5), it follows that $E[D_{s+c,k}]$ is a function of the bit error probability P_b, the source coding rate $R_{s,k}$, and the encoded video sequence. Thus, to summarize, we can write the expected distortion as $E[D_{s+c,k}](R_{s,k}, R_{c,k}, \underline{S})$.

III. Optimal Resource Allocation

A. Problem Formulation

A centralized control unit at the network layer determines how network resources should be allocated amongst the nodes. It can request changes in transmission parameters, such as the source coding rates, channel coding rates, and power levels. There are two criteria we will utilize to optimally allocate the network resources to each node in the network. The constraint for both criteria is that the chip rate be the same for all nodes. Assuming that the spreading code length is the same for all nodes, a constraint on the chip rate corresponds to a constraint on the transmission bit rate R_k. Thus, we can equivalently impose a constraint on the bit rate instead of the chip rate. The first criterion we will employ can be formally stated as follows: Given a total target bit rate, R_{budget}, determine the vectors of optimal source coding rates $\underline{R_s}^*$, channel coding rates $\underline{R_c}^*$, and power levels $\underline{S}^*$ such that the overall end-to-end distortion $D_{ave}(\underline{R_s}, \underline{R_c}, \underline{S})$ over all nodes is minimized:

$$\begin{aligned}\left\{\underline{R_s}^*, \underline{R_c}^*, \underline{S}^*\right\} &= \arg\min_{\underline{R_s},\underline{R_c},\underline{S}} D_{ave}(\underline{R_s}, \underline{R_c}, \underline{S})\\ &\text{subject to } R_1 = R_2 = \cdots = R_K = R_{budget}\end{aligned} \tag{6}$$

with $R_k = (R_{s,k}/R_{c,k})$, and $D_{ave}(\underline{R_s}, \underline{R_c}, \underline{S}) = (1/K)\sum_{k=1}^{K} E[D_{s+c,k}](R_{s,k}, R_{c,k}, \underline{S})$.

The second criterion we will use to allocate resources to the nodes in the network minimizes the maximum distortion:

$$\begin{aligned}&\left\{\underline{R_s}^*, \underline{R_c}^*, \underline{S}^*\right\} = \arg\min_{\underline{R_s},\underline{R_c},\underline{S}}\{\max_k E[D_{s+c,k}](R_{s,k}, R_{c,k}, \underline{S})\}\\ &\text{subject to } R_1 = R_2 = \cdots = R_K = R_{budget}\end{aligned} \tag{7}$$

with $R_k = (R_{s,k}/R_{c,k})$. This formulation assumes that the videos from all sensors are equally important, but allows sensors that image low-motion scenes to use a lower source coding rate. This criterion guarantees fairness among all sensors, since we are minimizing the worst distortion among all sensors. The problem is a discrete optimization problem, that is, $R_{s,k}$, $R_{c,k}$, and S_k can only take values from discrete sets $\mathbf{R}_s$, $\mathbf{R}_c$, and $\mathbf{S}$, respectively, i.e., $R_{s,k} \in \mathbf{R}_s$, $R_{c,k} \in \mathbf{R}_c$, $S_k \in \mathbf{S}$ [9].

We assume that the K nodes are grouped into N *motion classes* according to the amount of motion in the scenes they are imaging. For example, if $N = 2$, we can have two classes of nodes, low-motion nodes and high-motion nodes. We assume that each class has its own set of URDC curves (5). Thus, instead of determining the source coding rate, channel coding rate, and power for each node, we just determine these parameters for each class.

B. Optimization Algorithm and Computational Complexity

We next discuss our proposed discrete optimization algorithm and its computational complexity. Given that for all admissible $(R_{s,k}, R_{c,k})$ pairs, we should have $(R_{s,k}/R_{c,k}) = R_{budget}$, the cardinalities $|\mathbf{R_s}|$ and $|\mathbf{R_c}|$ should be equal, i.e., $|\mathbf{R_s}| = |\mathbf{R_c}| = C$. The number of admissible $(R_{s,k}, R_{c,k})$ pairs should also be equal to C. For each class of nodes, the source coding rate-channel coding rate pair and the power level should be selected. The number of possible choices for each class of nodes is $C \cdot |\mathbf{S}|$, where $|\mathbf{S}|$ is the cardinality of set $\mathbf{S}$. If there are N motion classes, the total number of admissible combinations of source coding rate, channel coding rate, and power level for each class of nodes is $(C \cdot |\mathbf{S}|)^N$.

The problems of (6) and (7) could be solved using exhaustive search, by trying out all $(C \cdot |\mathbf{S}|)^N$ combinations and selecting the one that minimizes the corresponding expression. However, the computational complexity can be reduced by noting that $E[D_{s+c,k}](R_{s,k}, R_{c,k}, \underline{S})$ for node k is not affected by the choices of source coding rates and channel coding rates of the other users. It is only affected by the power selections of the

TABLE I
MAD WITH VARIOUS DISTRIBUTIONS OF NODE TYPES: $\text{Target bit rate} = 144\,000 \text{ bits/s}$

Low	(R_{s1}, R_{c1}, S_1)	$D_{s+c,1}$	High	(R_{s2}, R_{c2}, S_2)	$D_{s+c,2}$	D_{ave}	$PSNR_{ave}$
90	$(72k, 1/2, 10)$	7.8	10	$(96k, 2/3, 15)$	19.6	9.0	$38.6dB$
70	$(72k, 1/2, 10)$	9.0	30	$(96k, 2/3, 15)$	21.0	12.6	$37.1dB$
50	$(48k, 1/3, 5)$	11.9	50	$(96k, 2/3, 10)$	20.6	16.3	$36.0dB$
30	$(48k, 1/3, 5)$	13.8	70	$(96k, 2/3, 10)$	23.3	20.4	$35.0dB$
10	$(48k, 1/3, 5)$	16.3	90	$(96k, 2/3, 10)$	26.9	25.8	$34.0dB$

TABLE II
MAD WITH VARIOUS DISTRIBUTIONS OF NODE TYPES: $\text{Target bit rate} = 96\,000 \text{ bits/s}$

Low	(R_{s1}, R_{c1}, S_1)	$D_{s+c,1}$	High	(R_{s2}, R_{c2}, S_2)	$D_{s+c,2}$	D_{ave}	$PSNR_{ave}$
90	$(48k, 1/2, 5)$	7.9	10	$(64k, 2/3, 15)$	23.5	9.5	$38.4dB$
70	$(48k, 1/2, 5)$	8.5	30	$(64k, 2/3, 10)$	31.7	15.4	$36.2dB$
50	$(48k, 1/2, 5)$	9.7	50	$(64k, 2/3, 10)$	35.1	22.4	$34.6dB$
30	$(48k, 1/2, 5)$	11.3	70	$(64k, 2/3, 10)$	38.8	30.5	$33.3dB$
10	$(48k, 1/2, 5)$	13.3	90	$(64k, 2/3, 10)$	43.0	40.0	$32.1dB$

other users. There are $|\mathbf{S}|^N$ possible power allocations among the classes of nodes. For each power allocation, each class of nodes should select the best $(R_{s,k}, R_{c,k})$ pair (the one that minimizes the expected distortion). Since there are C such pairs, $C-1$ comparisons will be needed. In order to do that for all classes of nodes, the total number of comparisons will be $N(C-1)|\mathbf{S}|^N$. Thus, for each of the $|\mathbf{S}|^N$ power allocations $\underline{S}$, we have found the source-channel coding rate combinations that would minimize the expected distortion for each class of nodes.

Thus, to solve the problem of (6), we need to find the minimum of $|\mathbf{S}|^N$ numbers. For that, we need $|\mathbf{S}|^N - 1$ comparisons. To summarize, we need a total of $N(C-1)|\mathbf{S}|^N + |\mathbf{S}|^N - 1$ comparisons to solve the problem of (6).

In order to solve the problem of (7), for each of the $|\mathbf{S}|^N$ power combinations, we need to compare the distortions for each class of nodes and find the maximum distortion among the node classes. For this, we will need $N-1$ comparisons. After we do that, we need to find the minimum of these values among the $|\mathbf{S}|^N$ combinations. So, we need a total of $N(C-1)|\mathbf{S}|^N + (N-1) + |\mathbf{S}|^N - 1$ comparisons in order to solve the problem of (7).

IV. EXPERIMENTAL RESULTS

We next provide experimental results using software simulations. We perform the optimization procedure discussed in Section III using the proposed model for URDCs. The data points used to obtain the parameters a and b in (5) are obtained by corrupting the video stream with packet errors based on a calculated P_b, decoding the corrupted video bit stream with the H.264/AVC codec, calculating the distortion, repeating this experiment 300 times, and then taking the average distortion. We assume that there are two possible motion levels viewed by the sensor nodes, low motion and high motion, Thus, there are two node classes $(N = 2)$. The "Akiyo" sequence is used to represent a low-motion node, and the "Foreman" sequence is used to represent a high-motion node. It is necessary to have two sets of URDC curves, one for each level of motion. The characteristics were obtained for both video sequences at a frame rate of 15 f/s.

We use BPSK modulation and RCPC codes with mother code rate 1/4 from [11] for channel coding. We set the link layer packet size, LL_{size}, to 400 bits. We examine target bit rate constraints at 144 000 bits/s and 96 000 bits/s. The total bandwidth, W_t, was set to 20 MHz. The set of admissible source coding rates and corresponding channel coding rates for the different target bit rates are

$$R = 144\,000 \ \frac{\text{bits}}{\text{s}} \rightarrow \mathbf{R}_s, \quad \mathbf{R}_c \in \left\{ \left(48 \text{ kbps}, \frac{1}{3}\right), \left(72 \text{ kbps}, \frac{1}{2}\right), \left(96 \text{ kbps}, \frac{2}{3}\right) \right\} \tag{8}$$

$$R = 96\,000 \ \frac{\text{bits}}{\text{s}} \rightarrow \mathbf{R}_s, \quad \mathbf{R}_c \in \left\{ \left(32 \text{ kbps}, \frac{1}{3}\right), \left(48 \text{ kbps}, \frac{1}{2}\right), \left(64 \text{ kbps}, \frac{2}{3}\right) \right\}. \tag{9}$$

The power levels in Watts were chosen from $\mathbf{S} \in \{5, 10, 15\}$ Watts. Thus, $C = 3$ and $|\mathbf{S}| = 3$.

In Tables I–VIII, we show how the network resources should be assigned for various distributions of the two types of nodes for different target bit rates. The low-motion nodes' source coding rate in bits per second, channel coding rate, and power level in Watts are represented by R_{s1}, R_{c1}, and S_1, respectively, and the high-motion nodes' parameters are represented by R_{s2}, R_{c2}, and S_2. The number of low-motion nodes is given under column, "Low", and the number of high-motion nodes is given under column, "High". "MAD" corresponds to the method of Minimizing the Average end-to-end Distortion over all users, and "MMD" corresponds to the technique of Minimizing the Maximum Distortion. In Tables I–VI, the distribution of the two types of nodes is varied while the total number of

TABLE III
MMD WITH VARIOUS DISTRIBUTIONS OF NODE TYPES: Target bit rate = 144 000 bits/s

Low	(R_{s1}, R_{c1}, S_1)	$D_{s+c,1}$	High	(R_{s2}, R_{c2}, S_2)	$D_{s+c,2}$	D_{ave}	$PSNR_{ave}$
90	$(48k, 1/3, 5)$	9.6	10	$(96k, 2/3, 15)$	14.5	10.1	$38.1dB$
70	$(48k, 1/3, 5)$	12.8	30	$(96k, 2/3, 15)$	16.5	13.9	$36.7dB$
50	$(48k, 1/3, 5)$	17.9	50	$(96k, 2/3, 15)$	18.9	18.4	$35.5dB$
30	$(48k, 1/3, 5)$	13.8	70	$(96k, 2/3, 10)$	23.3	20.4	$35.0dB$
10	$(48k, 1/3, 5)$	16.3	90	$(96k, 2/3, 10)$	26.9	25.8	$34.0dB$

TABLE IV
MMD WITH VARIOUS DISTRIBUTIONS OF NODE TYPES: Target bit rate = 96 000 bits/s

Low	(R_{s1}, R_{c1}, S_1)	$D_{s+c,1}$	High	(R_{s2}, R_{c2}, S_2)	$D_{s+c,2}$	D_{ave}	$PSNR_{ave}$
90	$(48k, 1/2, 5)$	7.9	10	$(64k, 2/3, 15)$	23.5	9.5	$38.4dB$
70	$(48k, 1/2, 5)$	10.4	30	$(64k, 2/3, 15)$	27.9	15.7	$36.2dB$
50	$(48k, 1/2, 5)$	14.6	50	$(64k, 2/3, 15)$	32.2	23.4	$34.4dB$
30	$(32k, 1/3, 5)$	21.1	70	$(64k, 2/3, 15)$	36.9	32.1	$33.1dB$
10	$(32k, 1/3, 5)$	25.7	90	$(64k, 2/3, 15)$	42.2	40.6	$32.0dB$

TABLE V
MAD WITH EQUAL DISTRIBUTIONS OF NODE TYPES: Target bit rate = 144 000 bits/s

Low	(R_{s1}, R_{c1}, S_1)	$D_{s+c,1}$	High	(R_{s2}, R_{c2}, S_2)	$D_{s+c,2}$	D_{ave}	$PSNR_{ave}$
10	$(96k, 2/3, 15)$	1.8	10	$(96k, 2/3, 15)$	12.1	6.9	$39.7dB$
30	$(96k, 2/3, 10)$	4.7	30	$(96k, 2/3, 15)$	16.0	10.4	$38.0dB$
50	$(48k, 1/3, 5)$	11.9	50	$(96k, 2/3, 10)$	20.6	16.3	$36.0dB$
70	$(48k, 1/3, 5)$	20.0	70	$(96k, 2/3, 10)$	32.8	26.4	$33.9dB$
90	$(48k, 1/3, 10)$	24.5	90	$(48k, 1/3, 15)$	68.3	46.4	$31.5dB$

TABLE VI
MAD WITH EQUAL DISTRIBUTIONS OF NODE TYPES: Target bit rate = 96 000 bits/s

Low	(R_{s1}, R_{c1}, S_1)	$D_{s+c,1}$	High	(R_{s2}, R_{c2}, S_2)	$D_{s+c,2}$	D_{ave}	$PSNR_{ave}$
10	$(64k, 2/3, 15)$	3.0	10	$(64k, 2/3, 15)$	22.4	12.7	$37.1dB$
30	$(48k, 1/2, 5)$	7.9	30	$(64k, 2/3, 15)$	23.5	15.7	$36.2dB$
50	$(48k, 1/2, 5)$	9.7	50	$(64k, 2/3, 10)$	35.1	22.4	$34.6dB$
70	$(48k, 1/2, 10)$	11.7	70	$(48k, 1/2, 15)$	49.7	30.7	$33.3dB$
90	$(32k, 1/3, 10)$	19.8	90	$(48k, 1/2, 15)$	58.9	39.3	$32.2dB$

TABLE VII
MMD WITH EQUAL DISTRIBUTIONS OF NODE TYPES: Target bit rate = 144 000 bits/s

Low	(R_{s1}, R_{c1}, S_1)	$D_{s+c,1}$	High	(R_{s2}, R_{c2}, S_2)	$D_{s+c,2}$	D_{ave}	$PSNR_{ave}$
10	$(72k, 1/2, 15)$	1.8	10	$(96k, 2/3, 15)$	12.1	6.9	$39.7dB$
30	$(48k, 1/3, 5)$	9.6	30	$(96k, 2/3, 15)$	14.5	12.1	$37.3dB$
50	$(48k, 1/3, 5)$	17.9	50	$(96k, 2/3, 15)$	18.9	18.4	$35.5dB$
70	$(48k, 1/3, 5)$	20.0	70	$(96k, 2/3, 10)$	32.8	26.4	$33.9dB$
90	$(48k, 1/3, 10)$	24.5	90	$(48k, 1/3, 15)$	68.3	46.4	$31.5dB$

nodes is kept constant. Tables I and II use the MAD criterion while Tables III and IV utilize the MMD criterion. We give the resulting average end-to-end peak signal-to-noise ratio (PSNR) in dB for the entire network as a measure of performance for the MAD experiments. We also use the minimum PSNR as a measure of performance for the MMD experiments. The PSNR is calculated from the expected distortion PSNR $= 10\log\left(255^2/E\left[D_{s+c}\right]\right)$ where PSNR is the peak signal-to-noise ratio and $E\left[D_{s+c}\right]$ is the expected distortion due to source coding and channel errors.

TABLE VIII
MMD WITH EQUAL DISTRIBUTIONS OF NODE TYPES: $\text{Target bit rate} = 96\,000\ \text{bits/s}$

Low	(R_{s1}, R_{c1}, S_1)	$D_{s+c,1}$	High	(R_{s2}, R_{c2}, S_2)	$D_{s+c,2}$	D_{ave}	$PSNR_{ave}$
10	$(48k, 1/2, 15)$	3.0	10	$(48k, 1/2, 15)$	22.4	12.7	$37.1dB$
30	$(48k, 1/2, 5)$	7.9	30	$(64k, 2/3, 15)$	23.5	15.7	$36.2dB$
50	$(48k, 1/2, 5)$	14.6	50	$(64k, 2/3, 15)$	32.2	23.4	$34.4dB$
70	$(32k, 1/3, 5)$	25.7	70	$(64k, 2/3, 15)$	42.2	34.0	$32.8dB$
90	$(32k, 1/3, 5)$	51.5	90	$(48k, 1/2, 15)$	50.5	51.0	$31.1dB$

From the discussion of Section III-B, we can see that we need 44 comparisons for MAD and 45 comparisons for MMD.

As expected, the PSNR decreases as you move down the MAD tables because the number of high-motion nodes are increasing. When using MMD, we observe how the value of the average PSNR can actually increase at some points as you move down the table. This occurs when the maximum distortion switches from being that of the high-motion node to that of the low-motion node and vice versa. We see that in most MAD cases, high-motion nodes are assigned a higher source coding rate than the low-motion nodes. This is because the drop in the end-to-end distortion when increasing the source coding rate for a high-motion video sequence is more significant than the effect of employing stronger channel coding. However, the distortions for the low-motion video sequence remain relatively low, even when the source coding rate is decreased, so it can afford to transmit at a lower source coding rate in some cases.

V. CONCLUSIONS

In this paper, we presented a cross-layer optimization algorithm that works across the physical layer, the data link layer, and the application layer in a wireless visual sensor network. This algorithm accounts for network performances all the way from the physical layer up to the application layer. At the application layer, we determined the source coding rate, R_s, for video compression. At the data link layer, we assigned the channel coding rate, R_c. At the physical layer, we selected the power level, S. The algorithm shows how to distribute these parameters among all the nodes transmitting in the network. To create a realistic DS-CDMA visual sensor network, different levels of motion were assumed to be imaged by the nodes. By utilizing the parametric model for the URDCs, we estimated each node's expected distortion in a computationally efficient manner. We presented the combinations of $\{R_s, R_c, S\}$ for each node that result in the minimal average end-to-end distortion over all nodes in the system and the combinations that minimize the maximum distortion. We also showed how to determine the minimum total bandwidth needed to obtain a specific level of quality for the desired number of nodes and which target chip rate achieves the highest average PSNR for a given total bandwidth. Our experimental results demonstrated the effectiveness of the proposed cross-layer optimization scheme.

REFERENCES

[1] H. Kwon, T. H. Kim, S. Choi, and B. G. Lee, "A cross-layer strategy for energy-efficient reliable delivery in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3689–3699, Dec. 2006.

[2] R. Madan, S. Cui, S. Lall, and A. Goldsmith, "Cross-layer design for lifetime maximization in interference-limited wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 11, pp. 3142–3152, Nov. 2006.

[3] G. Messier, J. Hartwell, and R. Davies, "A sensor network cross-layer power control algorithm that incorporates multiple-access interference," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 2877–2883, Aug. 2008.

[4] K.-Y. Chow, K.-S. Lui, and E. Y. Lam, "Balancing image quality and energy consumption in visual sensor networks," in *Proc. 2006 1st Int. Symp. Wireless Pervasive Computing*, Jan. 2006.

[5] T. Zahariadis and S. Voliotis, "Open issues in wireless visual sensor networking," in *Proc. 14th Int. Workshop Systems, Signals, and Image Processing, 2007 and 6th EURASIP Conf. Focused on Speech and Image Processing, Multimedia Communications and Services*, Jun. 2007, pp. 335–338.

[6] E. S. Pynadath and L. P. Kondi, "Cross-layer optimization with power control in DS-CDMA visual sensor networks," in *Proc. IEEE Int. Conf. Image Processing*, Atlanta, GA, 2006, pp. 25–28.

[7] E. S. Bentley, L. P. Kondi, J. D. Matyjas, and M. J. Medley, "Wireless visual sensor network resource allocation using cross-layer optimization," in *Proc. Visual Communications and Image Processing Conf.*, San Jose, CA, 2009.

[8] Y. S. Chan and J. W. Modestino, "A joint source coding-power control approach for video transmission over CDMA networks," *IEEE J. Select. Areas Commun.*, vol. 21, no. 10, pp. 1516–1525, Dec. 2003.

[9] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint source-channel coding for motion-compensated DCT-based SNR scalable video," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 1043–1052, Sep. 2002.

[10] T. Stockhammer, M. M. Hannuksela, and T. Wiegand, "H.264/AVC in wireless environments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 657–672, Jul. 2003.

[11] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. 36, no. 4, pp. 389–400, Apr. 1988.

[12] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson*, RTP: A Transport Protocol for Real-Time Applications*. Reston, VA: The Internet Society: Network Working Group Std., 2003.

[13] L. P. Kondi and A. K. Katsaggelos, "Joint source-channel coding for scalable video using models of rate-distortion functions," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Salt Lake City, UT, 2001, pp. 1377–1380.

[14] M. Bystrom and T. Stockhammer, "Modeling of operational distortion-rate characteristics for joint source-channel coding of video," in *Proc. IEEE Int. Conf. Image Processing*, Kobe, Japan, 1999, pp. 359–362.

Elizabeth S. Bentley received the B.S. degree in electrical engineering from Cornell University, Ithaca, NY, in 1999, the M.S. degree in electrical engineering from Lehigh University, Bethlehem, PA, in 2001, and the Ph.D. degree in electrical engineering from State University of New York at Buffalo in 2007.

From 2007 to 2008, she was a National Research Council Post-Doctoral Research Associate at the Air Force Research Laboratory in Rome, NY. Currently, she is employed by the Air Force Research Laboratory, performing research and development in the wireless communications branch. Her research interests are in cross-layer optimization, wireless multiple-access communications, video processing, and wireless video transmission.

Lisimachos P. Kondi (S'92–M'99) received the Diploma in electrical engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 1994, and the M.S. and Ph.D. degrees in electrical and computer engineering from Northwestern University, Evanston, IL, in 1996 and 1999, respectively.

During the 1999–2000 academic year, he was a postdoctoral research associate at Northwestern University. He is currently an Assistant Professor in the Department of Computer Science at the University of Ioannina, Ioannina, Greece. He was previously with the faculty of the University at Buffalo, The State University of New York, and has also held summer appointments at the Naval Research Laboratory, Washington, DC, and the Air Force Research Laboratory, Rome, NY. His research interests are in the general areas of signal and image processing and communications, including image and video compression and transmission over wireless channels and the Internet, super-resolution of video sequences, and shape coding.

Dr. Kondi is an Associate Editor for the *EURASIP Journal of Advances in Signal Processing* and an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS.

John D. Matyjas received the A.S. degree in pre-engineering from Niagara University in 1996 and the B.S., M.S., and Ph.D. degrees in electrical engineering from the State University of New York at Buffalo in 1998, 2000, and 2004, respectively.

He was a Teaching Assistant (1998–2002) and a Research Assistant (1998–2004) with the Communications and Signals Laboratory, Department of Electrical Engineering, State University of New York at Buffalo. Currently, he is employed since 2004 by the Air Force Research Laboratory in Rome, NY, performing R&D in the area of information connectivity. His research interests are in the areas of cognitive wireless multiple-access communications and networking, statistical signal processing and optimization, and neural networks. Additionally, he serves as an adjunct faculty in the Department of Electrical Engineering at the State University of New York Institute of Technology at Utica/Rome.

Dr. Matyjas is the recipient of the 2010 IEEE ICC "Best Paper Award," 2009 Mohawk Valley Engineering Executive Council "Engineer of the Year" Award, and the 2009 Fred I. Diamond Basic Research Award for "best technical paper." He is a member of the IEEE Communications, Information Theory, Computational Intelligence, and Signal Processing Societies; chair of the IEEE Mohawk Valley Chapter Signal Processing Society; and a member of the Tau Beta Pi and Eta Kappa Nu engineering honor societies.

Michael J. Medley (S'91–M'95–SM'02) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Rensselaer Polytechnic Institute, Troy, NY, in 1990, 1991, and 1995, respectively.

Since 1991, he has been a research engineer for the United States Air Force at the Air Force Research Laboratory, Rome, NY, where he has been involved in communications and signal processing research related to adaptive interference suppression, spread spectrum waveform design, covert messaging, and airborne networking and communications links. In 2002, he joined the State University of New York Institute of Technology in Utica, NY, where he currently serves as an Associate Professor and Coordinator of the Electrical Engineering program.

Bruce W. Suter received the B.S. and M.S. degrees in electrical engineering in 1972, and the Ph.D. degree in computer science in 1988 from the University of South Florida, Tampa.

In 1998, he joined the technical staff at the U.S. Air Force Research Laboratory, Rome, NY, where he was the founding Director of the Center for Integrated Transmission and Exploitation (CITE). He has held visiting appointments at Harvard University and the Massachusetts Institute of Technology. His current research interests are focused on wireless computing networks and their applications to signal and image processing. His previous positions include academia at the U.S. Air Force Institute of Technology and the University of Alabama at Birmingham, together with industrial positions at Honeywell Inc. and Litton Industries. He is author of over 100 technical publications and the author of a widely accepted monograph *Multirate and Wavelet Signal Processing* (New York: Academic, 1998).

Dr. Suter is a member of Tau Beta Pi and Eta Kappa Nu. He is a former associate editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He has received a number of awards for his engineering and research contributions. These include the Air Force Research Laboratory (AFRL) Fellow, an AFRL-wide award for his accomplishments in the theory, application, and implementation of signal processing algorithms, the Arthur S. Flemming Award, a government-wide award for his pioneering Hankel transform research, the General Ronald W. Yates Award for Excellence in Technology Transfer for his patented Fourier transform processor, and the Fred I. Diamond Award for best laboratory research publication.