Optimization of System Performance for DVC Applications with Energy Constraints over Ad Hoc Networks

Lifeng Sun[1], Ke Liang[2], Shiqiang Yang[1], and Yuzhuo Zhong[1]

[1] Department of Computer Science and Technology, Tsinghua University, 100084, Beijing, China
[2] School of Computing, National University of Singapore
sunlf@tsinghua.edu.cn, liangke@comp.nus.edu.sg

Abstract. We investigate optimization of system performance in the below scenario: capturing and transmitting videos by single or multiple video sensors using distributed video coding (DVC) over ad hoc networks. There is an intrinsic contradiction in this scenario that could affect the system performance: the contradiction between the decoding quality and network lifetime. In this paper, we propose a joint optimization between the decoding quality and network lifetime using a quantitative metric of system performance, which is defined as the amount of collected visual information during the operational time of the video sensor. Based on the proposed metric, an optimal encoding rate is determined, which results in an optimal system performance. The simulation results show that the optimal encoding rate can be determined to achieve the optimal system performance.

Keywords: Distributed video coding, ad hoc networks, energy, optimization.

1 Introduction

An ad hoc network is formed by wireless nodes without any established infrastructure. Each wireless node in an ad hoc network can act as a source, a receiver, or a relay for transmission. Ad hoc networks are unstable due to high packet loss probability, rate fluctuate, frequent route updates and node mobility. Recently, video streaming over ad-hoc networks becomes a young and hot topic in research and industry. In this application, video cameras are embedded into the sensors, which can capture, process, and transmit videos to receivers.

There are two challenges in transmitting video over ad hoc networks: First, since video cameras in ad hoc networks are usually powered by batteries and have limited computing capability and memory, it is necessary to design a new video codec to replace the conventional video encoder, which is complex and power-consuming. Second and more critically, since the resource is limited, a rate allocation strategy should be proposed to find the optimal encoding rate which results in the optimal performance of system, which means high decoding quality and long network lifetime.

T.-J. Cham et al. (Eds.): MMM 2007, LNCS 4352, Part II, pp. 23 – 31, 2007.

To cope with the first challenge, distributed video coding (DVC) [1,2,3,4] is proposed to lower down the computational burden of the encoder by shifting the motion compensation module from the encoder to the decoder. Further more, DVC has built-in error-resilient ability due to the channel coding techniques. Therefore, DVC encoder is lightweight and power-saving, and these features are propitious to wireless terminals.

It is more critical but more difficult to cope with the second challenge. In order to achieve higher decoding quality, higher encoding rate should be allocated to the video sensor. However, higher encoding rate brings higher power consumption which results in shorter network lifetime. The problem can be formulated as the mutual conditionality between the decoding quality and network lifetime. In other words, how to utilize the limited resource efficiently to achieve best system performance. In [5], the amount of collected visual information is defined as a quantitative metric of performance for single-source system. This metric is intuitionist to describe the properties of visual information function V(R): Non-decreasing, homogenous, concave and bounded, fast degradation around zero.

In this paper, we give a quantitative expression of visual information based on DVC rate distortion function of the video frame. Using this expression, we can find an optimal encoding rate for the video sensor of DVC application, and the optimal rate results in the optimal system performance.

The rest of the paper is organized as follows: In section 2, we will describe the fundamental models (include the ad-hoc network scenario model, the proposed rate-distortion model and the power model). In section 3, we will define the visual information and give the quantitative metric of system performance. In section 4, we will present the rate allocation scheme based on section 3 and analyze the simulation results. Finally, section 5 concludes the paper.

2 System Models

In this section, some system models are given to formulate the optimal problem. We first introduce the DVC rate distortion model based on Slepian-Wolf and Wyner-Ziv theorem [6,7]. And following, base on the DVC rate distortion model, the expression of visual information of a video frame is given.

2.1 DVC Rate Distortion Model

DVC originates from DSC (Distributed Source Coding), and the source is replaced by video sequences. Consider two successive video frames, X_{n-1} and X_n, in the video sequence, where the subscript n-1 and n denote the time indexes. $\hat{X}_{n-1}$ and $\hat{X}_n$ are reconstructed frames of X_{n-1} and X_n separately. In DVC scheme, X_{n-1} is encoded by a conventional intraframe encoder, unlike predictive coding scheme, X_n is encoded by a Wyner-Ziv encoder without the knowledge of X_{n-1}. At the decoder, with the help of $\hat{X}_{n-1}$, which is regard as *side information*, X_n is reconstructed as $\hat{X}_n$.

The information-theoretic rate-distortion bounds for Wyner-Ziv coding are first established in [8]:

$$R_{X|Y}(D) \le R_{WZ}(D) \le R_X(D) \tag{1}$$

Where X denotes the to-be-encoded source, Y denotes the side information. $R_{X|Y}(D)$ is the rate distortion function of inter-frame coding (predictive coding), and $R_X(D)$ denotes the rate distortion of intra-frame coding. $R_{X|Y}(D) = R_{WZ}(D)$ means that Wyner-Ziv coding is lossless in case of the source X and the side information Y are jointly Gaussian sequence, and MSE (Mean Squared Error) is used as the distortion measure. In [9], a generalization of the rate distortion function was given for Wyner-Ziv coding of noisy sources in the quadratic Gaussian case.

Notice that the distortion decays exponentially with rate, we give a more precise distortion function using exponential decay function, and the proposed rate distortion function can be represented as:

$$D = D_0 + \alpha \cdot e^{-R/\beta} \tag{2}$$

Where D_0, α, and β (decay constant) can be estimated experimental data of the video sequence via regressive technologies. The main difference between the proposed rate distortion function and that in [9] is the decay constant β, which is fixed in [9] and does not change with video sequence.

We use *foreman* sequence for experiment, as is shown in Fig.1. The proposed rate distortion has good fitting property with (SSE, EMSE) = (71.24, 2.179).

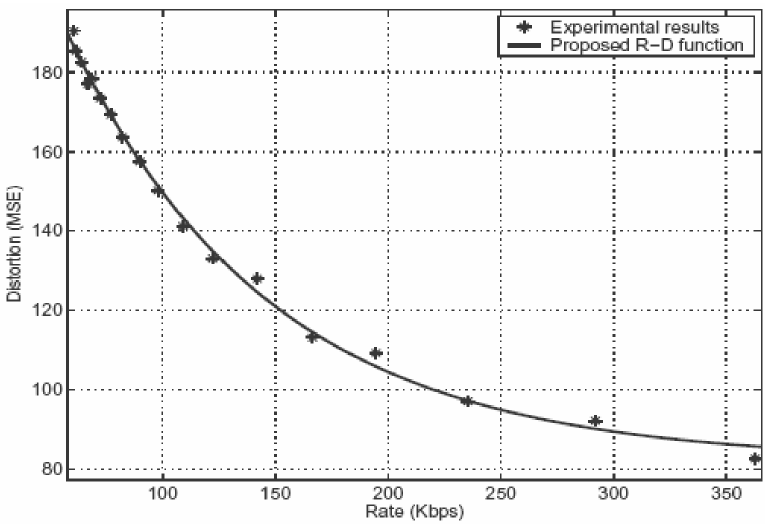

Fig. 1. Proposed R-D function

2.2 Network Scenario Model

We study the case where a number of nodes capturing and transmitting videos to a receiver in an ad hoc networks. As shown in Fig.2, $node_{1\text{-}6}$ is deployed in different place with different initial energy and transmits videos to the receiver $node_0$ using single hop or multiple hops.

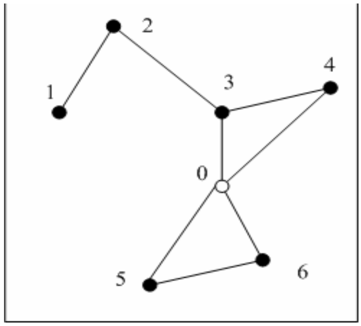

Fig. 2. Network scenario of ad hoc video sensor network

2.3 Power Consumption Model

We use the radio model presented in [10], which is an d^2 energy loss estimation for channel transmission. In order to transmit a k-bit message from node i to node j directly, the energy consumed of sending and receiving can be figured out by:

$$E_S(k,d) = k(\alpha_1 + \alpha_2 d^2)$$
$$E_R(k,d) = k\alpha_r \tag{3}$$

where E_S and E_R are the energy consumptions of sending and receiving; d is the transmission range, typically 250 meters; α_1 is for the energy dissipated to run the radio transmitter, α_r is for the energy dissipated to run the receiver circuitry, and α_2 is for the energy dissipated at the transmit amplifier, $\alpha_1 = \alpha_r = 50nJ/bit$, $\alpha_2 = 100pJ/bit/m^2$.

In a typical ad hoc network, power consumption of all the nodes can be represented as:

$$P_S = k_s R, \quad P_R = k_R R \tag{4}$$

Where P_S and P_R are the powers consumed for network sending and receiving; R is the transfer rate; $k_s = \alpha_1 + \alpha_2 d^2 = 6.3\mu J/bit$, $k_r = \alpha_r = 0.05\mu J/bit$.

For a whole ad hoc network system, we define the lifetime of the network system as the duration from the startup time of the network to the time when any single node runs out of energy. The power consumed by each source node is formed by the spending of CPU, camera, network device and network transmission. According to [11], the power consumed by a node is relatively constant over time. The only variability comes from the network transmission. So the lifetime can be figured out by:

$$T = \min_i \{E_i / (P_{Ti} + P_{Ci})\} \tag{5}$$

Where E_i is the total energy of node i, P_{Ci} is the constant part of power consumed by it, which includes power of CPU, camera and other devices. P_{Ti} represents the average power for transmission. The receiver is not taken into account because it is very different from other source nodes. During the final time, nodes energies fall to a very low level and BER arises, so the theoretical lifetime is a little longer than in reality.

3 Visual Information and System Performance Model

In order to measure the system performance, visual information is introduced. For an image, visual information is determined by the encoding rate and the variance of the image. For a video sequence, motion activity is regard as an more important factor. When a uniform video coding technology is used, visual information is determined by the activity of the to-be-encoded video frame. Generally speaking, more motion (comparing with the previous video frame) the to-be-encoded video frame has, more visual information it contains.

The amount of collected visual information during the operational lifetime has been defined as the metric of system performance in [5]. It should be noted that when the rate is low, small increment of rate brings big increment of decoding quality. However, when the rate is relatively high, large increment of rate brings small increment of decoding quality. Therefore, there exists a contribution function $\phi(R)$, which denotes the contribution of encoding rate R to the decoding quality.

For an image, visual information is defined as the amount of information needed to communicate for the image. Undoubtedly, visual information is in direct proportion to bit-rate of the image, in another word, the visual information function is non-decreasing to bit-rate. Intuitively, the visual information is convex at high rate.

However, we have no idea of the feature of visual information function at low rate. There are two possibilities at low rate: concave or convex. If it is convex, the visual information function is similar with the PSNR curve, thus the visual information can be regarded as the information of the image.

In our opinion, visual information of a video sequence is determined by not only the variance of video frames in the video sequence, but also the similarity of successive video frames. Therefore, we define the visual information from R-D function: D(R):

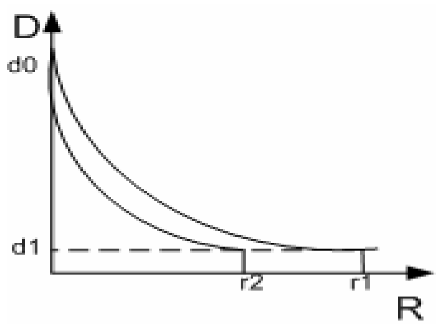

Fig. 3. R-D curve of video sequence

From Fig. 3, we can find some interesting facts in R-D curve:

1. d0 means the largest difference (MSE measured) between the two successive video frames. The curve will have a larger d0 if large motion is observed in the video

sequence. If the bit-rate is zero, it means the to-be-decoded frame is taken place by the previous frame.

2. Different R-D curve have different variance. Larger variance, the curve farther from the zero. And this is the feature of visual information in images.

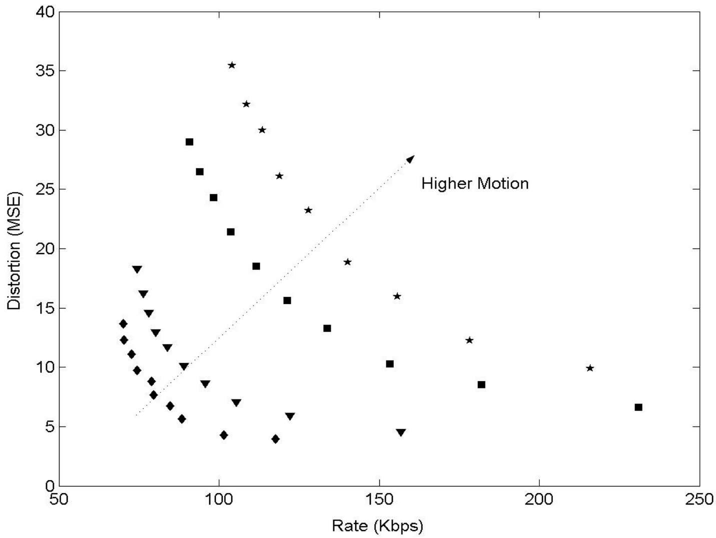

Fig. 4. Relationship of visual information with motion

For a video sequence, as shown in Fig. 4, larger motion, more visual information can be achieved. On the other hand, for an image, larger variance, more visual information can be achieved.

Therefore, we regard the area, which is formed by the two coordinate axes and the R-D function, as the visual information. For example in fig.3, the visual information is the area formed by d0, d1, and r2.

$$D = D_0 + \alpha \cdot e^{-R/\beta} \tag{6}$$

In this equation, D_0, α, β are parameters determined by the video sequence (quantizes, variance, and motion respectively).

We can compute the visual information using this R-D function:

$$V(R) = \int_{r=0}^{R} D(r)dr - D(R) \cdot R = \alpha \cdot \left[\beta - (R + \beta) \cdot e^{-R/\beta}\right] \tag{7}$$

As shown in the equation above, the upper bound of visual information is determined by the motion of the video sequence and the variance of the video frames.

We can also get the contribution of bit rate to the visual information:

$$\phi(R) = \frac{dV}{dR} = \frac{1}{\beta} \cdot R \cdot e^{-R/\beta} \tag{8}$$

4 Rate Allocation and Validation

We first use the concept of visual information in single-source video applications, in this case, we joint optimize the decoding quality and the network lifetime.

We use the ratio of visual information achieved to the energy consumed to indicate the efficiency of the system.

If the fame-rate is fixed, the energy consumed is in proportion to the bit rate according to equation 3, and the efficiency can be formulated as equation 8.

$$\frac{V}{E} = \frac{\alpha \cdot \left[\beta - (R+\beta)\cdot e^{-R/\beta}\right]}{K \cdot R} \tag{9}$$

Where K is a constant, which means that the transmission energy consumed is determined by the bit-rate. To get the optimized rate allocation, let $\frac{d\frac{V}{E}}{dR} = 0$, we get the equation:

$$e^{x} = x^{2} + x + 1, \qquad x = \frac{R}{\beta}$$
$$R \approx 1.8\beta \tag{10}$$

The variable β can be regarded as the effect fact of motion, which is in inverse proportion to the similarity of successive video frames. We just classify the motion of the video into three types: *Low, Middle*, and *High*.

And the simulation results of rate allocation for single source video (foreman sequence for experiment) are shown below:

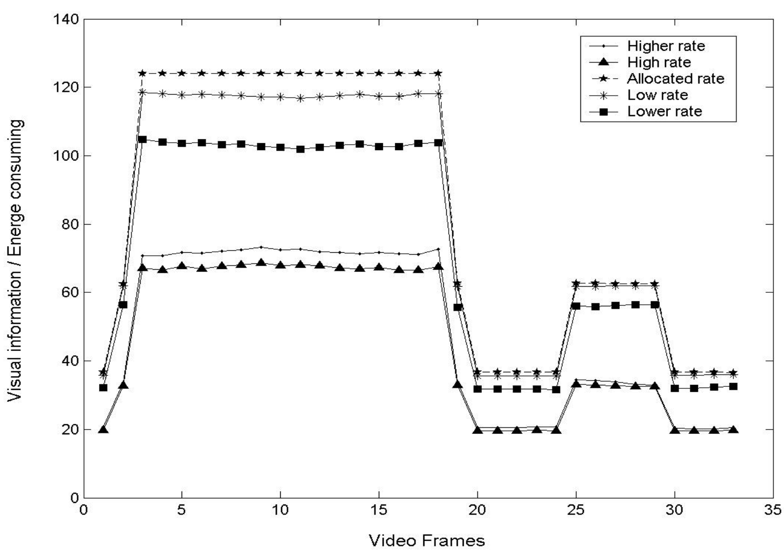

Fig. 5. Rate allocation simulations

As shown in the Figure5, the rate allocated by the proposed algorithm can achieve the highest efficiency of the system.

Since visual information of a video sequence is determined by the motion of the video sequence and the variance of the video frames. Therefore, the higher motion the video sequence has, the lower efficiency (which we defined as the ratio of visual information achieved to the energy consumed) achieved. We can see the simulation results shown in the figure above, in the low motion area (from the 3th frame to the 17th frame), high efficiency is achieved. On the other hand, in the high motion area (from the 20th frame to the 24th frame), efficiency is lowered down.

If a very low rate is allocated to encode the video frames, low visual information is achieved though the encoding rate is low, which results in low system efficiency. When the rate is increasing, the system efficiency is getting higher. As shown in the figure above, the optimal system efficiency is achieved when the rate is allocated using $R \approx 1.8\beta$.

5 Conclusions

We investigate optimization of system performance in the scenario of capturing and transmitting videos by single or multiple video sensors using distributed video coding (DVC) over ad hoc networks. We propose a joint optimization between the decoding quality and network lifetime using a quantitative metric of system performance, which is defined as the amount of collected visual information during the operational time of the video sensor. In the simulation, we classify the motion of the video into three types: Low, Middle, and High. Different types have different β, which is regarded as the effect fact of motion. The simulations show that the rate allocated by the proposed algorithm can achieve the highest efficiency of the system. For the future work, we will research on allocating the optimal encoding rate for the to-be-encoded video frame accurately and propose a rate allocation strategy for multi-source video streams from the sensors to achieve the optimal overall system performance.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under Grant No. 60503063, National Basic Research Program of China (973) under Grant No.2006CB303103 and the Natural Science Foundation of Beijing under Grant No. 4063037. The authors would like to thank Prof. Zhihai He and Prof. Dapeng Wu for their helping on visual information model and power consumption model.

References

1. B. Girod, A. Aaron, S. Rane and D. Rebollo-Monedero , "Distributed video coding," *Proceedings of the IEEE*, Special Issue on Video Coding and Delivery, vol. 93, no. 1, pp. 71-83, January 2005.
2. S. S. Pradhan and K. Ramchandran, “Distributed source coding using syndromes (DISCUS): Design and construction,” in Proc. IEEE Data Compression Conference, Snowbird, UT, Mar. 1999, pp. 158 –167.

3. A. Aaron, R. Zhang, and B. Girod, “Wyner-Ziv coding of motion video,” in Proc. Asilomar Conference on Signals and Systems, Pacific Grove, CA, Nov. 2002.
4. A. Aaron, S. Rane, R. Zhang, and B. Girod, “Wyner-Ziv coding for video: Applications to compression and error resilience,” in Proc. IEEE Data Compression Conference, Snowbird, UT, Mar. 2003, pp. 93–102.
5. Zhihai He; Dapeng Wu, Accumulative visual information in wireless video sensor network: definition and analysis, IEEE International Conference on Communications (ICC 2005), Seoul, Korea, May 16 - 20, 2005 Page(s):1205 - 1208 Vol. 2
6. J. D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” IEEE Transactions on Information Theory, vol. IT-19, pp. 471– 480, July 1973.
7. A. D. Wyner, “Recent Results in the Shannon Theory,” IEEE Transactions on Information Theory, vol. 20, no. 1, pp. 2–10, Jan. 1974.
8. A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Trans. Info. Theory, vol. IT-22, pp. 1 - 10, January 1976.
9. D. Rebollo-Monedero and B. Girod, “A generalization of the rate-distortion function for Wyner-Ziv coding of noisy sources in the quadratic-Gaussian case,” in *Proc. IEEE Data Compression Conferaence, DCC-2005*, Snowbird, UT, Mar. 2005.
10. Wendi Rabiner Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan, “Energy-Efficient Communication Protocol for Wireless Microsensor Networks,” Proc. Hawaiian Int'l Conf. on Systems Science, January 2000.
11. Wu-chi Feng, Brian Code, Ed Kaiser, Mike Shea, Wu-chang Feng, Louis Bavoil, Panoptes: Scalable Low-Power Video Sensor Networking Technologies, ACM Transactions on Multimedia Computing, Communications and Applications, January 2005.
12. Zhihai He, Yongfang Liang, Lulin Chen, Ishfaq Ahmad, and Dapeng Wu, Power-rate-distortion analysis for wireless video communication under energy constraint, IEEE Trans. Circuits Syst. Video Technol., vol. 15, no. 5, pp. 645–658,. May. 2005.