J. Vis. Commun. Image R. 23 (2012) 569–585

Contents lists available at SciVerse ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Low-complexity video coding via power–rate–distortion optimization ☆

Li-Wei Kang [a], Chun-Shien Lu [a], Chih-Yang Lin [b,*]

[a] Institute of Information Science, Academia Sinica, 128, Sec. 2, Academia Rd, Nankang, Taipei 11529, Taiwan
[b] Department of Computer Science and Information Engineering, Asia University, 500, Lioufeng Rd, Wufeng, Taichung 41354, Taiwan

ARTICLE INFO

Article history:
Received 22 March 2011
Accepted 30 January 2012
Available online 6 February 2012

Keywords:
Hash
Low-complexity video coding
Motion estimation
Rate–distortion
Power–rate–distortion optimization
Multiview video coding
Wireless multimedia sensor networks
Low-power and power-aware video coding

ABSTRACT

Wireless multimedia sensor networks (WMSNs) have been potentially applicable for several emerging applications. The resources, *i.e.*, power and bandwidth available to visual sensors in a WMSN are, however, very limited. Hence, it is important but challenging to achieve efficient resource allocation and optimal video data compression while maximizing the overall network lifetime. In this paper, a power–rate–distortion (PRD) optimized resource-scalable low-complexity multiview video encoding scheme is proposed. In our video encoder, both the temporal and interview information can be exploited based on the comparisons of extracted media hashes without performing motion and disparity estimations, which are known to be time-consuming. We present a PRD model to characterize the relationship between the available resources and the RD performance of our encoder. More specifically, an RD function in terms of the percentages for different coding modes of blocks and the target bit rate under the available resource constraints is derived for optimal coding mode decision. The major goal here is to design a PRD model to optimize a "motion estimation-free" low-complexity video encoder for applications with resource-limited devices, instead of designing a general-purpose video codec to compete compression performance against current compression standards (*e.g.*, H.264/AVC). Analytic results verify the accuracy of our PRD model, which can provide a theoretical guideline for performance optimization under limited resource constraints. Simulation results on joint RD performance and power consumption (measured in terms of encoding time) demonstrate the applicability of our video coding scheme for WMSNs.

1. Introduction

1.1. Background

With the availability of low-cost hardware, wireless multimedia sensor networks (WMSNs) have been potentially applicable for several emerging applications [1,2]. A WMSN is a network of several wireless embedded devices supporting to retrieve visual, acoustic, and scalar data from a monitored physical environment. Here, a WMSN consisting of several battery-powered visual sensor nodes (VSNs) scattered in several sensor fields is considered. Each VSN equipped with a low-cost camera (*e.g.*, complementary metal-oxide-semiconductor, *i.e.*, CMOS camera) can capture and encode visual information along with delivering the compressed video data to the aggregation and forwarding node (AFN). The AFNs aggregate and forward the video data to the remote control unit (RCU), usually supporting a powerful decoder for video decoding. Compared with traditional network systems, WMSN operates under several resource constraints (*e.g.*, lower computational capability, limited power supply, and narrow transmission bandwidth). This will pose an emergent demand for simultaneously minimizing the power consumption and optimizing the video compression performance for each VSN in a WMSN while maximizing the overall network lifetime [3].

Based on the experimental analysis presented in [4], in typical scenarios of wireless video communication among visual sensor devices, the video encoding process consumes a significant portion (about 40–60%) of the total power consumption for a video device. For example, for a well-known WMSN hardware platform, called Crossbow Stargate [2], equipped with a low-power USB video camera, the video encoder consumes about 48% of the total power while the wireless transmission consumes about 11% of the total power. In this paper, we focus on optimal allocation of the power consumed for the video encoding component. Usually, the wireless transmission component is only enabled to transmit video data via request from AFN. If the power consumption for video encoding can be minimized while optimizing video compression efficiency, more power can be reserved for wireless transmission. Hence, the operational lifetime of each VSN can be prolonged while maximizing the overall network lifetime. The power consumption of

☆ This work was supported by National Science Council, Taiwan, under Grants NSC 100-2218-E-001-007-MY3, NSC 100-2811-E-001-005, and NSC 100-2221-E-468-021.

* Corresponding author.
E-mail address: andrewlin@asia.edu.tw (C.-Y. Lin).

1047-3203/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jvcir.2012.02.001

the remaining components depends on specific system design and cannot be easily controlled from a video encoding perspective [4].

To reduce the power consumption of both the video encoder and wireless video transmission for a VSN, low-complexity and high-efficiency video encoding is essential. If the video encoding complexity can be minimized while certain compression efficiency can be kept, the power consumption for both video encoding and video data transmission can be simultaneously decreased. Nevertheless, current video coding standards (*e.g.*, H.264/AVC [5]) usually perform complex interframe encoding (*e.g.*, motion estimation with high computational complexity for exploiting temporal correlation of successive frames). On the other hand, to sufficiently exploit correlations among adjacent VSNs in a WMSN, current multiview video coding (MVC) schemes (*e.g.*, H.264/AVC-based MVC, *i.e.*, joint multiview video model) [6,7] usually perform both interview (*e.g.*, disparity estimation) and temporal (*e.g.*, motion estimation) predictions at the encoder with high complexity. In addition, to perform interview prediction, uncompressed frames must be exchanged among VSNs, which is prohibitive in a WMSN.

1.2. Related works

To reduce the power consumption of a video encoder, several approaches have been proposed, including motion estimation-based low-complexity video encoder designs [8,9], hardware-based video encoder designs [10], and joint video encoder/decoder and hardware adaptation schemes [11,12]. It is claimed that the existing approaches, so far, focused on reduction of encoder complexity and power consumption through heuristic adaption, instead of systematic power optimization, due to their lack of an analytic model to characterize the optimal trade-off between power consumption and video encoding efficiency. Hence, in [4,13,14], a power–rate–distortion (PRD) optimization framework is proposed for optimal resource allocation of wireless video communication among visual sensor devices. Based on the PRD model, the minimum distortion that a video encoder can achieve under current power and rate constraints can be derived. The major core is to adjust the number of SAD (sum of absolute difference) operations used by motion estimation under the constraints of current encoding power and bit rates due to motion estimation is a major computation burden of general video encoders. This power-aware motion estimation concept is also reviewed in [15].

Nevertheless, the above-mentioned approaches [4,8–16] are all based on a standard or motion estimation-based video encoder, which is essentially with high encoding complexity. Recently, several "motion estimation-free" low-complexity video encoding schemes have been proposed, which can be roughly classified into three categories shown as follows.

1.3. Still image-based or standard codec-based low-complexity video coding [17–20]

Without performing motion estimation, the most straightforward video encoder is to apply still-image/intraframe encoding to each frame individually. In [17], an innovative image-pixel-position information based resource allocation scheme was proposed to optimize wavelet-based coded image transmission quality with strict energy budget constraint for image applications in WMSNs by exploring these uniquely different importance levels among image data streams. Network resources can be optimally allocated cross PHY, MAC and APP layers regarding inter-segment dependency, and energy efficiency is assured while the image transmission quality is optimized. In [18], a video-based sensor networking architecture, called Panoptes was presented, where two video sensor platforms that can deliver high-quality video over 802.11 networks with a power requirement less than 5 W were described. In their experiments, JPEG was employed as the compression format on the Panoptes platform. It is claimed that JPEG cannot allow for temporal compression of data, however, it saves on computational complexity and power consumption. In [19], a distributed image compression scheme is proposed to overcome the computation and energy limitation of individual nodes in a resource-constrained multihop wireless network by sharing the processing of tasks, where the overall network lifetime can be extended by distributing the computational load among otherwise idle processors. The authors proposed two design alternatives for energy efficient distributed JPEG-2000 image compression with respect to energy consumption and image quality. In [20], the H.264/AVC intraframe encoder was used as baseline benchmarking to evaluate their low-complexity encoding scheme. To further exploit temporal correlation, H.264/AVC interframe coding with no motion, where all the motion vectors are set to zeros, has been shown to be very efficient and difficult to be defeated [20].

1.4. Collaborative video coding and transmission [21,22]

To further increase coding efficiency, interview correlation among VSNs can be exploited via collaborative video coding and transmission. While transmitting the intra-encoded frames from adjacent VSNs toward the AFN through the same intermediate node, this node can perform an image matching procedure to detect the similar regions, which can be encoded once only for these frames. However, image matching is usually a complex task, but [21,22] usually assume that they can be easily detected or already known as prior knowledge.

1.5. Distributed video coding (DVC) [20,23–29]

The unique characteristic of DVC is that individual frames are encoded independently, but decoded jointly. The major encoding burden, *i.e.*, motion estimation, can be shifted to the decoder while preserving a certain coding efficiency. At the encoder, the compression of W (called Wyner-Ziv frame) can be achieved by transmitting only part of the parity (Wyner-Ziv) bits derived from the channel-encoded version of W. The decoder uses the received Wyner-Ziv bits and the side information Y derived from previous decoded video signals to perform channel decoding for the reconstruction of W. The side information Y can be generated by decoder-side motion estimation exploiting the temporal and/or interview correlations, respectively, from current VSN and adjacent VSNs.

1.6. Motivations and overview of the proposed scheme

So far, the existing "motion estimation-free" video encoding schemes lack an analytic model to characterize the optimal trade-off between power consumption and video encoding efficiency. Hence, in this paper, we focus on presenting a PRD optimized "motion estimation-free" low-complexity multiview video encoding scheme.

In our method, both the temporal and interview predictive coding can be achieved by extracting the significant differences between a frame and its reference frames, respectively, from the same VSN and the adjacent VSNs based on comparing the extracted media hashes without performing motion and disparity estimations. To exploit interview correlation, limited inter-VSN communications [21,22] during the encoding process are allowed to exchange hash information of relatively small size.

For simplicity, it is assumed that the overall available power for a VSN has been optimally allocated to video encoding, wireless transmission, and other components, and the transmitted video data can be correctly received at the decoder. We will focus on the optimal allocation of the available resources including power