TRACKING-OPTIMAL ERROR CONTROL SCHEMES FOR H.264 COMPRESSED VIDEO FOR VEHICLE SURVEILLANCE

Zhaofu Chen[1], *Eren Soyak*[2], *Sotirios A. Tsaftaris*[1,3], *Aggelos K. Katsaggelos*[1]

[1]Electrical Engineering and Computer Science Dept, Northwestern University, Evanston, IL, USA
[2]AirTies Wireless Networks, Istanbul, Turkey
[3]Dept. of Computer Science and Applications, IMT-Institutions Markets Technologies, Institute for Advanced Studies Lucca, Lucca, Italy

ABSTRACT

In this paper we present a transportation video coding and transmission system specifically tailored to automated vehicle tracking applications. By taking into account the video characteristics and the lossy nature of the wireless channels, we propose error control approaches to enhance tracking accuracy. The proposed system is shown to give performance improvement over the current state-of-the-art system and yields bitrate savings of up to 60%.

Index Terms— Transportation video, forward error control (FEC), error concealment, object tracking, H.264/AVC, surveillance centric coding

1. INTRODUCTION

Remote imaging sensors are commonly deployed for transportation monitoring and surveillance applications. Often the captured video needs to be transferred back to a central office for processing. The bandwidth limitation of the current wireless communication channels necessitates the use of video compression technologies at the remote sensors. Recently, H.264 started to be used in transportation video related applications, and has significantly reduced the bandwidth requirement. However, most of the systems currently in use are not specifically optimized for transportation videos and the automated analysis that might follow, and hence, the system performance as well as visual quality of the video will be severely affected.

Another challenge faced by the transportation video transmission system is the lossy nature of the wireless channels. The highly dependent H.264 bitstreams are sensitive to channel degradations, suffering error propagation due to predictive coding structure. There is significant interest in resource-distortion optimization given channel losses [1, 2]; however, these works focus on maximizing PSNR and do not consider the accuracy of object tracking. Recently, [3] proposed tracking-optimal modifications to H.264 compression that increase automated tracking accuracy in the receiver. Furthermore, in [4], certain aspects of equal error protection (EEP) were introduced for protecting H.264 bitstreams used in transportation monitoring applications.

In this paper, we propose a transportation video transmission system to consolidate bitrate on important video information and protect it with error control techniques. At the transmitter, we propose to use forward error control (FEC) with unequal protection levels to minimize the overall loss probability of the important packets while conserving the bandwidth resource. At the receiver side, we compare several error concealment (ERC) strategies, and propose the temporal-domain motion copy (MC) algorithm for transportation video decoding. The contributions of this paper are centered on optimizing:

- system behavior
- unequal error protection (UEP) approaches; and
- optimal concealment strategies,

from the viewpoint of maximizing tracking accuracy at the receiver's end. Although, other works dicuss video-quality rate tradeoffs [1, 2], this is the first treatment of identifying beneficial error mitigation and concealment strategies towards maximizing tracking accuracy while minimizing the computational load of the encoder.

The rest of this paper is organized as follows. In Section 2 we provide a brief overview as well as detailed explanations to the proposed system modules. In Section 3 we present experimental results using real-life test videos and demonstrate the effectiveness of our proposed system, which shows performance improvements compared with the state-of-the-art implementation and yields bitrate savings of up to 60%. Finally, the paper is concluded in Section 4.

2. PROPOSED METHODS

2.1. Performance Metric

In our system of transportation video surveillance, the ultimate application is to track (automatically) the objects (e.g. vehicles) in the video and to perform subsequent operations based on the tracking results. Therefore, objective metrics are

Thanks to Center for the Commercialization of Innovative Transportation Technology, Northwestern University for funding.

necessary to quantify and optimize the tracking performance of the overall system.

In [5], a review of the state-of-the-art for video surveillance performance metrics is presented. We choose the Overlap (OLAP), Precision (PREC) and Sensitivity (SENS) metrics presented in [3] due to their pertinence to the transportation tracking application. The tracking accuracy $\mathbf{A}$ is defined as a convex combination of the three components. Due to space limitation, we omit the details in this paper and the interested readers are referred to [3].

2.2. Overall System Description

The overall block diagram of the proposed transportation video transmission system is illustrated in Fig. 1. We denote

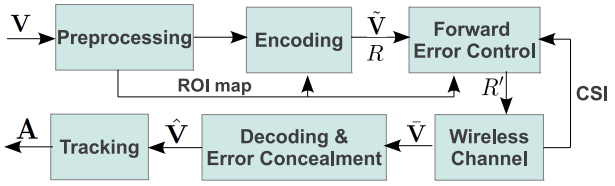


Fig. 1. System Diagram.

the input video as $\mathbf{V}$ and the output of encoder as $\tilde{\mathbf{V}}$. $\tilde{\mathbf{V}}$ is associated with the bitrate R and consists of video slices encapsulated into packets. The packets are protected with FEC schemes, which modify the bitrate to $R'(R, FEC)$. The protected packets are transmitted over the wireless channel characterized by a loss pattern. The received bitstream $\bar{\mathbf{V}}$ is then decoded with its possible losses concealed by the ERC module at the decoder. Finally, the decoded video $\hat{\mathbf{V}}$ is used for various applications such as object tracking.

As illustrated in Fig. 1, the preprocessing step filters the video for encoder as well as provides information for FEC. FEC utilizes the channel state information (CSI) feedback from the channel to determine the appropriate protection schemes. In this work we design the FEC at the transmitter and the ERC at the receiver to achieve the optimal balance between bitrate and accuracy, while always maintaining a low-computational profile at the encoder. In the following paragraphs, we will present our approach in detail and compare it with the current state-of-the-art implementation.

2.3. Preprocessing

The preprocessing step serves two purposes: (1) to filter the input video sequence to remove temporal noise for encoding and (2) to generate a region-of-interest (ROI) map for error protection. The filtering process is performed using the Temporal Deviation Thresholding (TDT) algorithm introduced in [3]. The TDT algorithm removes noise-like variations from the raw video before encoding, and re-inserts synthesized noise back to the decoded video prior to tracking. For further details on TDT, the reader is referred to [3]. As shown in [3], TDT allows for up to 90% reduction in bitrate required for a given level of tracking accuracy.

The ROI map provides a guidance to the FEC operation by classifying the video packets into a foreground group and a background group. Such classification is based on a non-parametric model of the temporal distribution of pixel intensities [6]. The goal is to isolate the regions showing events of high tracking interest (e.g., vehicles moving in streets) from regions undergoing constant changes (such as waving trees, water fountains, or light reflections). Specifically, let $f_t(n_1, n_2)$, denote intensity of pixel located at (n_1, n_2) in the t^{th} frame. In order to detect the ROI, we use the kurtosis of intensities for each pixel position over time, defined as:

$$\kappa(n_1, n_2) = \frac{\frac{1}{T}\sum_{t=0}^{T-1}(f_t(n_1, n_2) - \bar{f}(n_1, n_2))^4}{(\frac{1}{T}\sum_{t=0}^{T-1}(f_t(n_1, n_2) - \bar{f}(n_1, n_2))^2)^2} - 3, \tag{1}$$

where $\bar{f}(n_1, n_2)$ is the mean value of the intensities over the training length T. Note that T is less than the length of the entire sequence. In practice, if the scene is relatively fixed (e.g., the camera is mounted on a pole), a single ROI can be used for a relatively long time, until significant scene change occurs. The value of T can be calculated based on the statistical stationarity of the input video sequence. The training length should be selected to the balance between good statistical stability and computational complexity.

By definition, a Gaussian distribution has an excess kurtosis of 0. Furthermore, the additive property of kurtosis implies that a mixture of Gaussians also has an excess kurtosis of 0. Since the capture noise is modeled as additive Gaussian, and the constant movements of objects (such as trees) can be modeled as a mixture of Gaussians, they both can be characterized as having kurtosis of 0 [6]. The desired type of motion due to events such as moving vehicles can be modeled as a Poisson process, which has excess kurtosis of 6. Based on the above discussion, we can build the ROI map of pixels by thresholding their excess kurtosis values. For computational reasons, the threshold is set to a fixed value of 3, the middle point between the kurtosis values of the two models. Once the pixels are classified, the mapping of macroblocks can be done based on a majority vote rule. Specifically, the MB classification is determined by the majority class of pixels within that MB. The generated MB-level ROI map is fed into the FEC module, in which it is used to guide the assignment of protection levels, as explained in the next subsection.

2.4. Forward Error Control

The FEC module improves error resilience of the transmitted packets by adding redundancies in the encoded bitstreams. There exist various approaches for adding redundancies, including both intra-packet FEC and inter-packet FEC [7, 8]. Considering the limited computational resources at the remote node, we propose a simple yet effective channel protection methodology using redundant slices (RS) to minimize the overall packet loss probability in the wireless transmission.

In order to model the packet loss pattern, we consider a memoryless and uniformly distributed fading channel. Each unprotected packet is therefore subject to channel loss with

probability P_{unprot}. Let $i \in [0, I-1]$ be the packet (slice) index, and $c(i)$ be the total number of copies transmitted for packet i. Then the overall loss probability for packet i after FEC is $P_i = P_{\text{unprot}}^{c(i)}$, while the aggregated bitrate for packet i is $R_i' = R_i \cdot c(i)$, where R_i and R_i' are the bitrates before and after the FEC, respectively.

The assignment of protection levels (in terms of the number of copies transmitted for a packet) constitutes a trade-off between effective bitrate and the loss probability. Given the ROI map, we divide a video frame into slices in such a way that all the MBs in a single slice belong to the same group. We modified the H.264 encoder to enable this custom MB to slice mapping. For foreground slices, we assign the protection level $H = 1, 2, \cdots$, and for background slices, we assign the protection level $L \leq H$. The values of H and L can be selected based on the maximum supported bitrate $\hat{R}$ and a target overall loss probability P_{target}. Specifically, let $\mathcal{I}_H$ denote the set of foreground slices, and let $\mathcal{I}_L = \{0, 1, \cdots, I-1\} \setminus \mathcal{I}_H$, where "$\setminus$" is the set difference operator. The following procedures can be carried out to determine H and L:

$c_{\text{target}} = \left\lceil \log_{P_{\text{unprot}}}(P_{\text{target}}) \right\rceil$;
if $c_{\text{target}} \sum_{i \in \mathcal{I}_H} R_i + \sum_{i \in \mathcal{I}_L} R_i > \hat{R}$ **then**
  $H = \left\lfloor (\hat{R} - \sum_{i \in \mathcal{I}_L} R_i) / \sum_{i \in \mathcal{I}_H} R_i \right\rfloor$;
  $L = 1$;
else
  $H = c_{\text{target}}$;
  $L = \left\lfloor (\hat{R} - c_{\text{target}} \sum_{i \in \mathcal{I}_H} R_i) / \sum_{i \in \mathcal{I}_L} R_i \right\rfloor$;
end if

The underlying assumption here is that $\hat{R}$ can support at least one copy of each packet (in either group). If this assumption is violated, then we can prioritize importance packets and drop the less important packets first [9].

2.5. Error Concealment

In our application where FEC is utilized, a lost packet will not be retransmitted, and consequently the information contained in the lost packet must be estimated in the decoding process. The estimation of the lost video content using the reconstructed video content available at the decoder is known as error concealment (ERC).

In general, ERC works by utilizing the spatial or temporal correlation between the lost information and its neighbors [10]. A typical example of ERC scheme utilizing spatial correlation is the boundary matching algorithm (BMA) [11], which is implemented in the JM H.264 reference model. The algorithm works by interpolating the video content from reliably reconstructed spatial neighbors into the region with information loss. The interpolation option is selected to minimize the discrepancies of the boundaries surrounding the lost region.

Besides spatial ERC represented by the BMA, there are ERC schemes making explicit use of the temporal correlation. Two straightforward but intuitive examples in this category are the frame-copy (FC) algorithm and the motion-copy (MC) algorithms [12]. FC works by directly copying the co-located pixels from the previously reconstructed frames into the current frame. Similarly, the MC algorithm copies the motion information and then reconstructs the pixel values using such estimated motion information. In a simple case, the MC algorithm copies the reference picture index from the previously decoded frame, and then scales the motion vectors accordingly.

In transportation videos, it is reasonable to assume the objects of tracking interest exhibit smooth and consistent translational motion throughout consecutive frames. Therefore, the MC algorithm is potentially able to accurately estimate the motion information. The MC algorithm is particularly suitable for the preprocessed video, because the pixel variations not due to translational object motion have been suppressed. With the MC algorithm, the scaled motion vector and extrapolated reference frame index provide reliable reconstruction of the lost video information using its temporal predecessors. In the numerical examples below, we demonstrate that the MC algorithm indeed outperforms the other ERC schemes, and in particular the spatial BMA algorithm, in terms of tracking accuracy improvement.

3. NUMERICAL EXAMPLES

To verify the gains made possible by our proposed schemes, we conduct experiments using multiple sequences with different characteristics such as viewing angles, quality and type of observed vehicle traffic. Details of the sample implementation and experimental procedure with the test results are presented below.

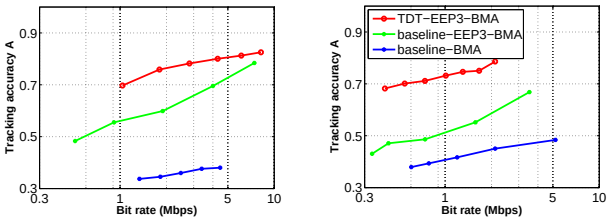
To implement the proposed system, the open-source JM 16.2 encoder with FMO enabled is used to read in the ROI map from the preprocessing step and to allow for packetization of the two different slice groups, and the JM decoder is modified to enable the FC and MC strategies. The JM decoder has a built-in BMA for error concealment, and it is used as a reference for performance evaluation. The open-source OpenCV "blobtrack" module is used as the object tracker which relies on the mean shift object tracking algorithm [13].

The following video sequences are used for the experiments. The "Camera6" sequence used under the NGSIM license courtesy of the US FHWA shows an intersection with light traffic, with trees swaying in the wind and buildings casting reflections of passing cars as part of the scene. The "dt_passat" sequence by courtesy of KOGS/IAKS Universität Karlsruhe shows a busy intersection with steady traffic interrupted by a traffic signal and a light urban rail crossing. Both sequences contain significant capture noise.

3.1. Effect of TDT

In this subsection, we demonstrate the effect of the TDT preprocessing step and compare its performance with that of the baseline H.264 implementation. The packet loss probability is set to be 0.1. The implementation with TDT and EEP at the transmitter and with BMA at the receiver is referred to as

the "Reference System", and is denoted by the red curves in Fig. 2. The baseline H.264 implementations are denoted by the green (with EEP) and blue (without EEP) curves in Fig. 2. The effect of the TDT preprocessing is significant; it reduces the bitrate by up to 90% for the same tracking accuracy.



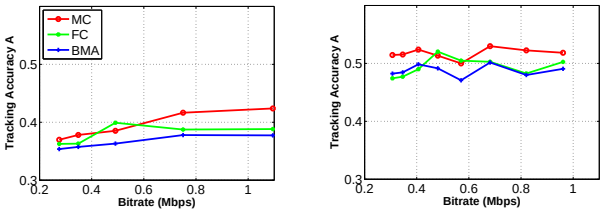
(a) Camera6 (b) dt_passat

Fig. 2. Effect of TDT preprocessing.

3.2. Comparison of Error Concealment Strategies on Tracking Accuracy

Here, we compare the BMA, FC, and MC algorithms in terms of their effectiveness in recovering the lost information and maintaining tracking accuracy. The channel model remains the same as in the previous subsection. The videos are encoded with various quantization parameters (QPs) and for each QP eight random channel realizations are obtained. In order to obtain a fair comparison, for each realization, the same lossy bitstreams are decoded using the various ERC schemes, and the final tracking accuracy results of each realization are averaged.

As is evidenced by Fig. 3, the spatial-domain scheme represented by the BMA in general performs worse than its temporal-domain counterparts. This highlights the characteristics of transportation video, and contrasts the unique requirements of a tracking application to the conventional viewing-oriented application. Comparing the MC and FC algorithms, the MC shows some advantage because it makes better use of the available motion information embedded in the previously decoded frames. The experimental results verify our theoretical reasoning and intuition made in the previous section.

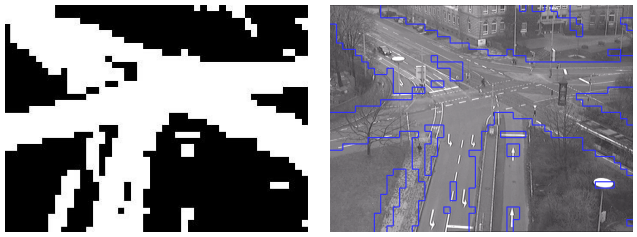


(a) Camera6 (b) dt_passat

Fig. 3. ERC comparison (packet loss probability of 0.1.)

3.3. ROI Extraction

In this subsection we illustrate the ROI map generated from the preprocessing step. After analysis of the video statistics (excess kurtosis), the preprocessing step identifies the regions in the video where motion of tracking interest is likely to occur, and generates an ROI map that is used to divide the video frame into different slice groups. The ROI map for the "dt_passat" sequence and the slice group mapping are shown in Fig. 4 (a) and (b), respectively. ROI maps for other sequences are similar and are omitted for brevity.



(a) ROI mapping (b) Frame w/ slice group mapping

Fig. 4. Example of ROI mapping. ("dt_passat" sequence)

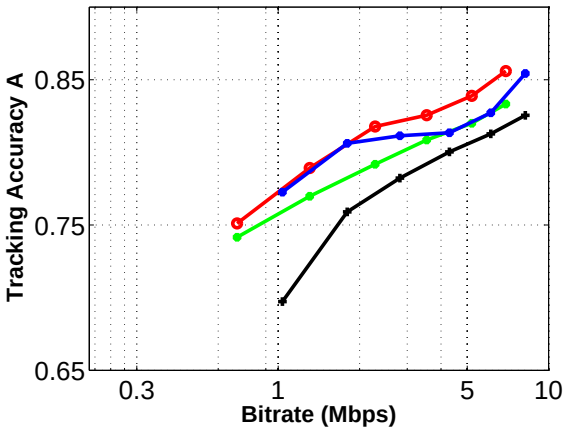
3.4. System Comparison

Finally, we integrate the proposed preprocessing, FEC and ERC modules together into a complete system (referred to as the "Proposed System" in what follows), and compare its performance with that of the current state-of-the-art implementation (referred to as the "Reference System"). In the FEC module, the Reference System applies EEP with protection level of 3 (this value gave empirically the optimal performance). In contrast, the Proposed System uses the protection levels $H = 3$ and $L = 2$, respectively. At the receiver, the Reference System uses the default ERC scheme (BMA) while for the Proposed System, we modify the JM decoder and implement the MC algorithm as an ERC module.

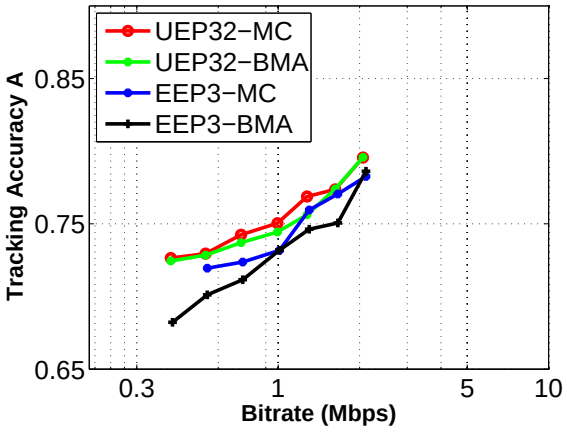
The performance comparisons are shown in Fig. 5. The Proposed System denoted as "UEP32-MC" is plotted in red curves, while the Reference System denoted as "EEP3-BMA" is plotted in black curves. Two intermediate implementations are also included for comparison. By including the "UEP32-BMA" and "EEP3-MC' implementations, we demonstrate that it is the combination of UEP (at the FEC module) and the MC (at the ERC module) that contribute to the overall system performance improvement. As can be seen from the figures, the Proposed System exhibits uniformly better performance than the Reference System. Quantitatively, the Proposed System provides a performance improvement of 40% to 60% bitrate reduction given the same tracking accuracy.

4. CONCLUSIONS

In this paper we presented a video coding and transmission system specifically tailored to automated vehicle surveillance and monitoring. The characteristics of transportation video and the lossy nature of the wireless channels were considered when designing the system. To mitigate the negative effects of channel losses to automated tracking of vehicles, we combined forward error correction (FEC) with unequal error protection (UEP) at the transmitter and an error concealment (ERC) module at the receiver. The effectiveness of the proposed system was demonstrated using real-life video sequences, and the performance improvement over the cur-



(a) Camera6



(b) dt_passat

Fig. 5. Comparison of System performance.

rent state-of-the-art system in maximizing tracking accuracy shown.

5. REFERENCES

[1] A. Katsaggelos, Y. Eisenberg, F. Zhai, R. Berry, and T. Pappas, "Advances in efficient resource allocation for packet-based real-time video transmission," *IEEE Proceedings*, vol. 93, pp. 135–147, January 2005.

[2] P. Baccichet, S. Rane, A. Chimienti, and B. Girod, "Robust low-delay video transmission using H.264/AVC redundant slices and flexible macroblock ordering," in *Proc. of IEEE Int'l Conf. on Image Proc.*, vol. 4, pp. 93–96, July 2007.

[3] E. Soyak, S. Tsaftaris, and A. Katsaggelos, "Low-complexity video compression for automated transportation surveillance," *IEEE Trans. on Circ. and Sys. for Video Tech., Special Issue on Video Analysis on Resource-Limited Sys.*, vol. 21, no. 10, pp. 1378–1389, 2011.

[4] E. Soyak, S. Tsaftaris, and A. Katsaggelos, "Channel protection for H.264 compression in transportation surveillance applications," in *Proc. of IEEE Int'l Conf. on Image Proc.*, pp. 2285–2288, 2011.

[5] A. Baumann, M. Boltz, J. Ebling, M. Koenig, H. Loos, M. Merkel, W. Niem, J. Warzelhan, and J. Yu, "A review and comparison of measures for automatic video surveillance systems," *EURASIP Journal on Image and Video Proc.*, vol. 2008, no. 1, 2008.

[6] E. Soyak, S. Tsaftaris, and A. Katsaggelos, "Content-aware H.264 encoding for traffic video tracking applications," in *Proc. of IEEE Int'l Conf. on Acoustics Speech and Signal Proc.*, pp. 730–733, March 2010.

[7] F. Zhai, Y. Eisenberg, T. Pappas, R. Berry, and A. Katsaggelos, "Joint source-channel coding and power adaptation for energy efficient wireless video communications," *Signal Proc.: Image Comm.*, vol. 20, pp. 371–387, March 2005.

[8] Y. Wang and Q. Zhu, "Error control and concealment for video communication: A review," *IEEE Trans. on Circ. and Sys. for Video Tech.*, vol. 86, no. 5, pp. 974–997, 1998.

[9] P. Pahalawatta, R. Berry, T. Pappas, and A. Katsaggelos, "Content-aware resource allocation and packet scheduling for video transmission over wireless networks," *IEEE Journal on Selected Areas in Comm., Special Issue on Cross-Layer Opt. for Wireless Multimedia Comm.*, vol. 25, pp. 749–759, 2007.

[10] Y. Wang, S. Wenger, J. Wen, and A. Katsaggelos, "Error resilient video coding techniques," *IEEE Signal Proc. Mag.*, vol. 17, pp. 61–82, July 2000.

[11] W. Lam, A. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," in *Proc. of IEEE Int'l Conf. on Acoustics, Speech, and Signal Proc.*, vol. 5, pp. 418–420, April 1993.

[12] S. Bandyopadhyay, Z. Wu, P. Pandit, and J. Boyce, "An error concealment scheme for entire frame losses for H.264/AVC," in *IEEE Sarnoff Symposium*, pp. 1–4, March 2006.

[13] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. 142–149, 2000.