Video Quality Monitoring for Mobile Multicast Peers Using Distributed Source Coding

Yao-Chung Lin, David Varodayan, and Bernd Girod
Information Systems Laboratory
Electrical Engineering Department, Stanford University
Stanford, CA, United States
{yao-chung.lin, varodayan, bgirod}@stanford.edu

ABSTRACT

We consider a peer-to-peer multicast video streaming system in which untrusted intermediaries transcode video streams for heterogeneous mobile peers. Many different legitimate versions of the video might exist. However, there is the risk that the untrusted intermediaries might tamper with the video content. Quality estimation and tampering detection are important in this scenario.

We propose that each mobile peer sends a digest of its received video to a quality monitoring server which has access to the original video. The digest is a Slepian-Wolf coded projection of the received video. Distributed source coding provides rate-efficient encoding of the projection by exploiting the correlation between the projections of the original and received videos. Two different projections are designed for quality estimation and tampering detection, respectively. We show that the projections can be encoded at a low rate of just a few kilobits per second. Compared to the ITU-T J.240 Recommendation for remote PSNR monitoring, our scheme achieves a bit-rate which is lower by at least one order of magnitude.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Design, Algorithms

Keywords

distributed source coding, video quality estimation, tampering detection, transcoding, peer-to-peer multicast video streaming

1. INTRODUCTION

Peer-to-peer streaming can efficiently deliver media content to a large population [3, 7, 9, 13]. A recent extension applies video multicasting to mobile peers with different decoding and display capabilities [7, 8]. A key feature of this system is that intermediaries transcode the video stream to accommodate the heterogeneous capabilities of mobile peers [8]. Therefore, many differently encoded versions of video exist. On the other hand, the intermediaries might tamper with the video for many reasons, such as inserting unauthorized advertisements or piggybacking unauthentic contents. Quality estimation and tampering detection are important tasks and this paper addresses how to efficiently exchange the information to achieve these goals.

In prior work, we have proposed media authentication systems in which a server provides Slepian-Wolf coded projections of original media as authentication data to users so that they can distinguish legitimate encodings from tampered copies [5, 12]. In an extension, the users can further estimate the received image quality [4] using a projection derived from the J.240 recommendation [1]. These approaches are suitable when the received and original videos are at the same frame rate and resolution, but a more rate-efficient approach is possible when the mobile peers receive videos that are at lower resolutions or lower frame rates.

We propose that each mobile peer sends the Slepian-Wolf coded projection of its received video to a quality monitoring server via a feedback channel, instead of the server sending the Slepian-Wolf coded projection as in prior work [5, 12, 4]. The quality monitoring server decodes the feedback data using the original video as side information. This architecture is advantageous for two reasons. The mobile peers are responsible for Slepian-Wolf encoding, which is much less computationally demanding than Slepian-Wolf decoding. Moreover, only the quality monitoring server has access to the full resolution video. Decoding the Slepian-Wolf coded projections of lower resolution video using full resolution side information is more efficient than decoding the coded projections of full resolution video using lower resolution side information.

In Section 2, we will describe the proposed quality monitoring schemes and the two classes of projections for quality estimation and tampering detection, respectively. Simulation results in Section 3 will demonstrate the tradeoffs between performance and feedback data rates for both cases.

2. VIDEO QUALITY MONITORING

Fig. 1 depicts the proposed quality monitoring system using distributed source coding. We denote the original video as

Digital Object Identifier: 10.4108/ICST.MOBIMEDIA2009.7532
http://dx.doi.org/10.4108/ICST.MOBIMEDIA2009.7532

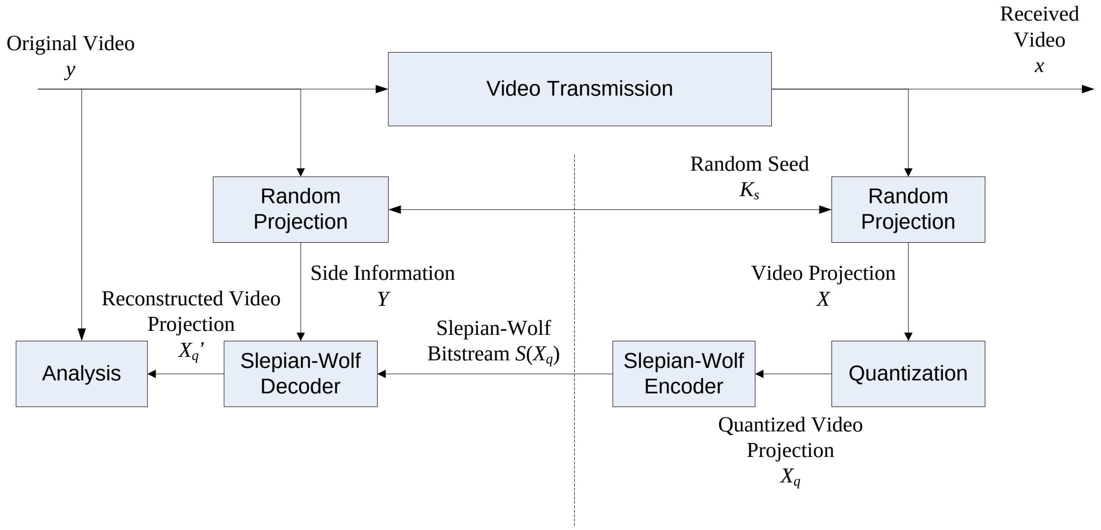


Figure 1: Proposed video quality monitoring scheme using distributed source coding.

y and the received video as x. Each mobile peer provides feedback data consisting of a Slepian-Wolf coded random projection of the received videos. The quality monitoring server uses the projection of the original video as side information to decode the feedback data. It then analyzes the projections either to estimate the quality in terms of reconstruction Peak Signal to Noise Ratio (PSNR) or to detect possible tampering by an adversary. We first describe the overall operation of the system, leaving the details of the random projections and analysis methods to Subsection 2.1 for quality estimation and to Subsection 2.2 for tampering detection.

The right-hand side of Fig. 1 shows the mobile peer. It applies a pseudorandom projection (based on a randomly drawn seed K_s) to its received video x and quantizes the projection coefficients X to yield X_q. These quantized coefficients are then coded by a Slepian-Wolf encoder based on low-density-parity check (LDPC) codes [6]. The mobile peer sends the Slepian-Wolf bitstream $S(X_q)$ as feedback data back to the quality monitoring server (shown on the left-hand side of Fig. 1) through a secure channel.

The mobile peer pseudorandomly generates a projection as a 16x16 block $\mathbf{P}$ according to a seed K_s. The seed changes for each frame and is communicated to the quality monitoring server along with the Slepian-Wolf bitstream. This prevents a malicious attack which simply confines tampering to the nullspace of the projection. For each 16x16 nonoverlapping block $\mathbf{B}_i$ of x, the inner product $\langle \mathbf{B}_i, \mathbf{P} \rangle$ is quantized into an element of X_q. The rate R of Slepian-Wolf bitstream $S(X_q)$ is determined by the joint statistics of X_q and Y. If the conditional entropy $H(X_q|Y)$ exceeds the rate R, then X_q can no longer be correctly decoded [10]. Therefore, we choose the rate R to be just sufficient to decode given x at the worst permissible quality.

Upon receiving the feedback data, the quality monitoring server first projects the original video y into Y using the same projections as at the mobile peer. A Slepian-Wolf decoder reconstructs X_q' from the Slepian-Wolf bitstream using Y as side information. Decoding is via LDPC belief propagation [6] initialized according to the statistics of the worst permissible degradation for the given original video. Finally, the quality monitoring server analyzes the reconstructed projection X_q' and the projection Y of the original video either to estimate video quality in terms of reconstruction PSNR or to detect tampering.

2.1 Quality Estimation

For quality estimation, we use the projection defined in the feature extraction module (shown in Fig. 2) of the J.240 recommendation [1]. Each 16x16 block $\mathbf{B}_i$ is whitened in both spatial and Walsh-Hadamard Transform (WHT) domains using pseudorandom number (PN) sequences $\mathbf{s}$ and $\mathbf{t}$, respectively, to yield 16x16 block $\mathbf{F}_i$. From this block, a single feature pixel $\mathbf{F}_i(k)$ is selected. Casting $\mathbf{B}_i$ and $\mathbf{F}_i$ as 1-D vectors, we can write

$$\mathbf{F}_i = \underbrace{\mathbf{H}^{-1}\mathbf{THS}}_{\mathbf{G}}\mathbf{B}_i$$

where $\mathbf{H}$ is the WHT matrix (casted from the 2-D WHT), and $\mathbf{S}$ and $\mathbf{T}$ are diagonal whitening matrices with entries $\mathbf{s}$ and $\mathbf{t}$, respectively. The projection $\mathbf{P}$ that produces $\mathbf{F}_i(k)$ is the k^{th} row of $\mathbf{G}$.

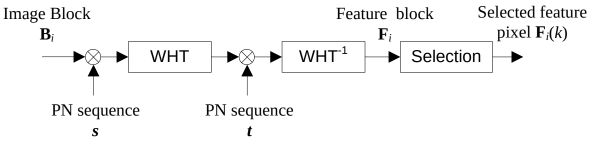


Figure 2: Random projection of J.240 feature extraction module.

Digital Object Identifier: 10.4108/ICST.MOBIMEDIA2009.7532
http://dx.doi.org/10.4108/ICST.MOBIMEDIA2009.7532

The analysis method for quality estimation is similar to that of J.240. In J.240 estimated PSNR (ePSNR) between x and y is computed as follows:

$$\mathrm{eMSE}_{\mathrm{J240}} = \frac{Q_s^2}{N}\sum_{i=1}^{N}(X_q(i) - Y_q(i))^2$$

$$\mathrm{ePSNR}_{\mathrm{J240}} = 10\log_{10}\frac{255^2}{\mathrm{eMSE}_{\mathrm{J240}}},$$

where N is the number of samples, Y_q is quantized version of Y, and Q_s is the quantization step size of Y_q and X_q. But in our system, the quality monitoring server has complete information about Y, and we propose maximum likelihood estimation of the MSE between x and y as follows:

$$\mathrm{eMSE}_{\mathrm{ml}} = \frac{1}{N}\sum_{i=1}^{N} E[(X - Y(i))^2 | Y(i), X_q(i)]$$

$$\mathrm{ePSNR}_{\mathrm{ml}} = 10\log_{10}\frac{255^2}{\mathrm{eMSE}_{\mathrm{ml}}}$$

We will compare the quality estimation performance of these two estimators in Section 3.

Finally, we argue that compression of X_q using distributed source coding is much more efficient then using conventional coding. Fig. 3 depicts the distributions of X and $X - Y$ of the first 100 frames of *Foreman* sequence at CIF resolution, and shows that X has a large variance whereas X and Y are highly correlated. We model $X|Y$ as a Gaussian with mean Y and variance σ_z^2, which is unknown at the decoder but can be estimated.

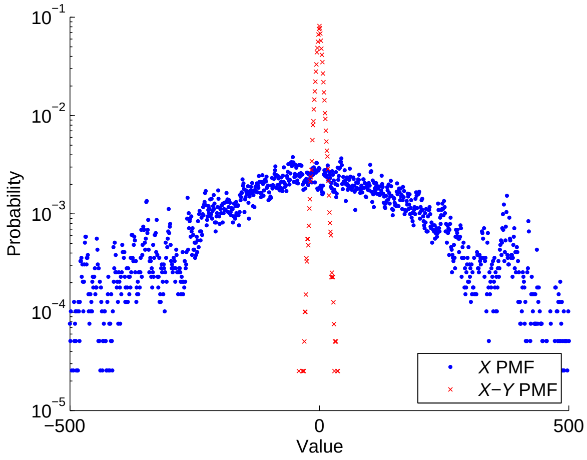


Figure 3: Distributions of X and $X - Y$ of the first 100 frames of *Foreman* sequence at CIF resolution

2.2 Tampering Detection

We model the tampering scenario by way of two-state lossy channel shown in Fig. 4. In the legitimate state, the channel performs lossy compression, transcoding and reconstruction using (for example) H.264/AVC, with PSNR of 30dB or better. In the tampered state, it additionally includes a malicious attack. Fig. 5 compares a sample input and two outputs of this channel. The source video y is *Foreman* in CIF. In the legitimate state, the channel is H.264 compression and reconstruction at 30dB PSNR. In the tampered state, a further malicious attack is applied: a 15x125 pixel text banner is overlaid on the reconstructed video.

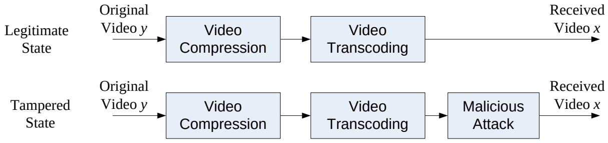
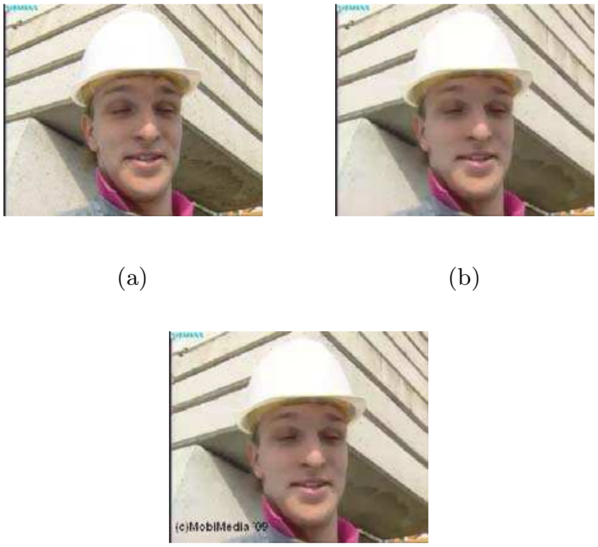


Figure 4: Two-state lossy channel



(a)

(b)

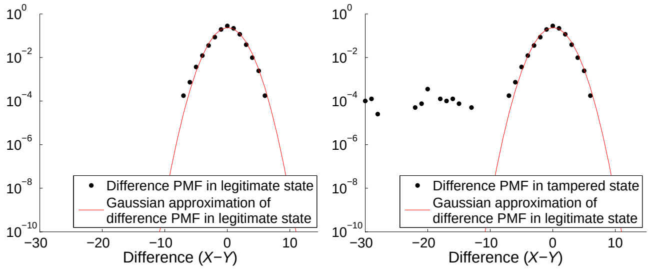
(c)

Figure 5: The first picture of *Foreman* sequence (a) y original, (b) x at output of legitimate channel, (c) x at output of tampered channel.

The legitimate channel introduces quantization errors of limited magnitude due to lossy compression, while the tampered channel additionally introduces much larger deviations such that joint statistics of X and Y vary depending on the state of the channel. We illustrate this by plotting in Fig. 6 the histogram of the difference $Z = X - Y$, where X and Y are projections of x and y in Fig. 5, respectively. We found that using block-wise mean as the projection $\mathbf{P}$ works well. To avoid null-space attacks, the elements in $\mathbf{P}$ are drawn from a Gaussian distribution $\mathcal{N}(1, \sigma^2)$ and normalized so that $||\mathbf{P}||_2 = 1$.

The analysis method for tampering detection is based on hypothesis testing. The null hypothesis H_0 that the channel is legitimate is tested against the hypothesis H_1 that the channel is tampered. We use the likelihood ratio test for the decision: $\prod_i \frac{P_0(X_q(i)'|Y(i))}{P_1(X_q(i)'|Y(i))} \lessgtr T$, where P_0 is the integral of a Gaussian at mean $Y(i)$ over quantization intervals of $X_q(i)'$, and P_1 is a convex combination of P_0 and the uniform distribution, i.e. $P_1 = (1-\alpha)P_0 + \alpha U_q$, for some α in [0,1]. U_q is the integral of a uniform distribution over quantization intervals of X_q. T is a fixed decision threshold.

Digital Object Identifier: 10.4108/ICST.MOBIMEDIA2009.7532
http://dx.doi.org/10.4108/ICST.MOBIMEDIA2009.7532



(a) (b)

Figure 6: The difference distributions between X and Y (a) in legitimate state, (b) in tampered state.

3. SIMULATION RESULTS

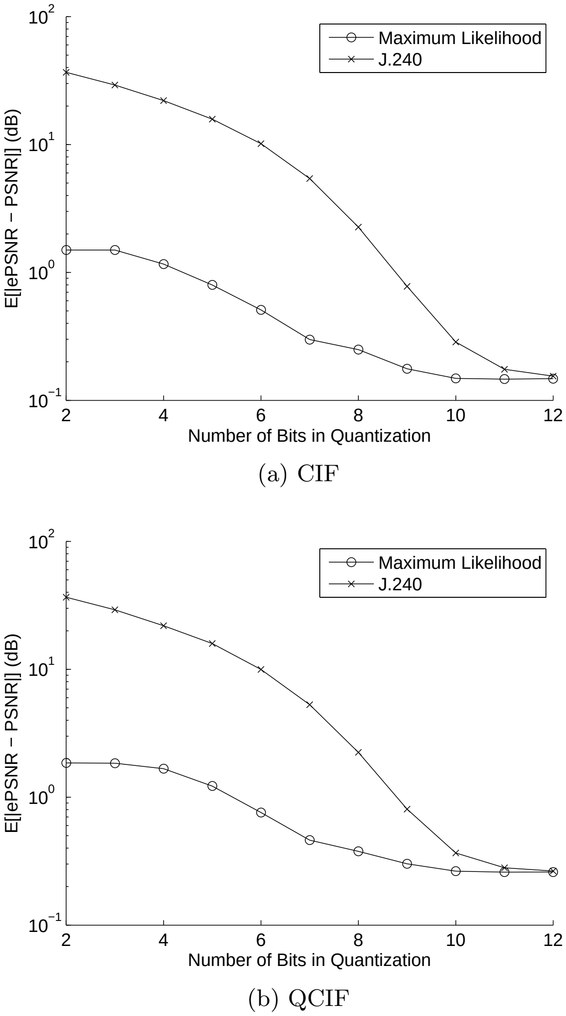
We use original videos consisting of the first 160 frames of *Foreman*, *Football*, *News*, *Mobile*, and *Coastguard* CIF video sequences at 30 frame per second (fps) for simulation. To create untampered received videos, the video sequences are first compressed and reconstructed by H.264 with quantization parameters (QP) 21, 24 and 26, for I-, P- and B-pictures, respectively. The group of picture (GOP) coding structure is IBBPBBP and GOP size is 16 frames. Then the compressed video is transcoded into CIF or QCIF resolution with GOP structure IPPP, GOP size 16 frames, and QP at most 38. The reconstruction yields the untampered received videos. For the experiments on tampering detection only, we additionally overlay a 15x125 text banner of random luminance from [0,255] at a random location of each frame to create the tampered videos.

At the mobile peer, a feedback unit consists of 16 frames. In the simulations, we vary the quantization of the random projection coefficients to different numbers of bitplanes. Each bitplane is coded at the Slepian-Wolf encoder using LDPC Accumulate (LDPCA) codes [11] with block size of 6336 bits for each CIF bitplane and 1584 bits for each QCIF bitplane. At the quality monitoring server, the bitplanes are conditionally decoded as in [2].

3.1 Quality Estimation Performance

Fig. 7 shows the average PSNR estimation error as we vary the number of bits in quantization, comparing the maximum likelihood and J.240 PSNR estimation methods. Each point represents the average PSNR estimation error |ePSNR − PSNR| of the luminance component over 350 measurements using 5 video sequences, 10 GOPs per sequence, and 7 transcoding QPs from the set (26, 28, ..., 38). Fig. 7 indicates we can get PSNR estimation error of just 0.3 dB with maximum likelihood estimation using 7-bit and 8-bit quantization for CIF and QCIF, respectively. This compares favorably with using 10-bit quantization for 0.3 dB PSNR estimation error with J.240 estimation.

Fig. 8 compares the rate of the Slepian-Wolf coded feedback data and the entropy of the quantized coefficients X_q, as we vary the number of bits in quantization. Each rate point represents the rate R required to decode X'_q given x at the worst permissible quality (transcoding QP 38) for all GOPs in all test video sequences. These results indicate that, at a rate of 0.006 bits per pixel of the received video, the mobile peers can send X in 7-bit precision using distributed source coding, but in less than 2-bit precision using conventional coding.



(a) CIF

(b) QCIF

Figure 7: Average PSNR estimation error versus the number of bits in quantization.

Fig. 9 combines the results of Figs. 7 and 8, comparing different combinations of estimation and coding methods. We plot the average PSNR estimation error versus the feedback data rate in kilobits per second (kbps), for videos at 30 fps. At average PSNR estimation error of 0.3 dB, maximum likelihood estimation and distributed source coding can reduce the feedback data rate up to 85% compared to J.240. This means that a feedback data packet of size 1500 bytes can provide PSNR estimation with less than 0.3 dB error for up to 0.7 second of CIF or 1.1 seconds of QCIF video at 30 fps.

3.2 Tampering Detection Performance

The tampering detection performance is evaluated per frame using a 5600 legitimate frames and 5600 tampered frames. We measure the false acceptance rate (the chance that a

Digital Object Identifier: 10.4108/ICST.MOBIMEDIA2009.7532
http://dx.doi.org/10.4108/ICST.MOBIMEDIA2009.7532

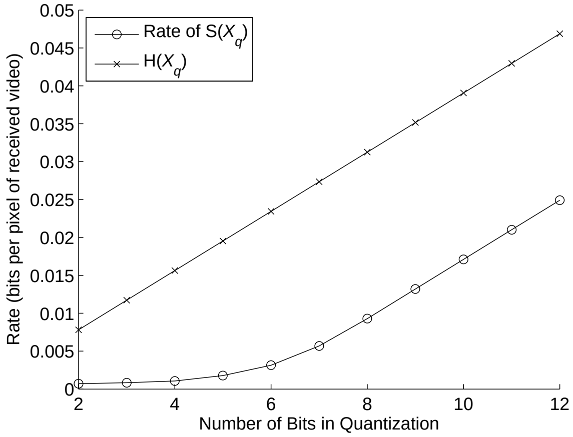


Figure 8: Minimum decodable rates versus the number of bits in quantization for X at the worst permissible quality.

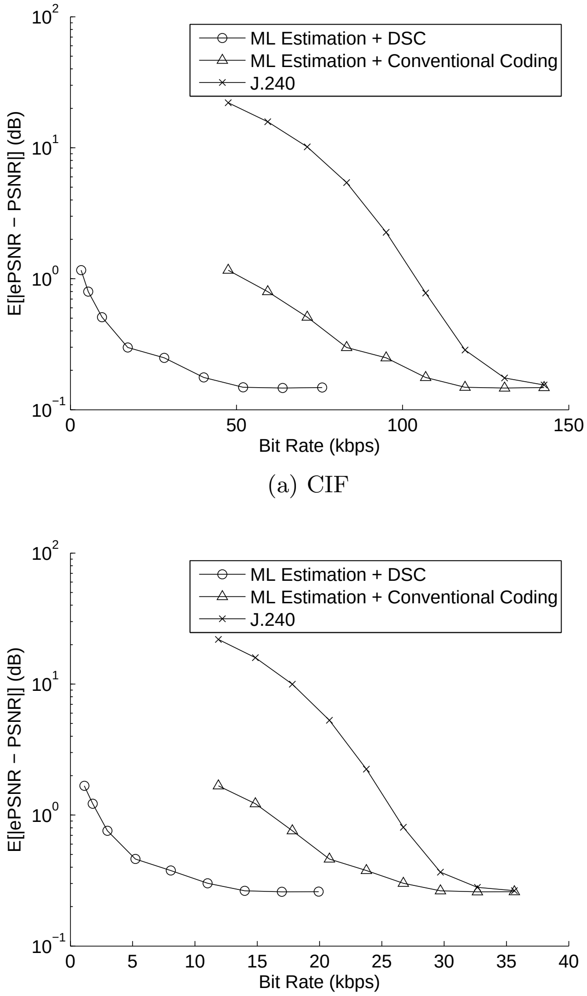
tampered frame is falsely accepted as a legitimate one) and the false rejection rate (the chance that a legitimate frame is falsely detected as tampered one.)

Fig. 10 compares the receiver operating characteristic (ROC) curves for tampering detection with different numbers of bits in quantization by sweeping the decision threshold T in the likelihood ratio test or the likelihood test. In the likelihood ratio test, we set the variance of the Gaussian in P_0 to be 2 and α (the convex combination parameter in P_1) to be 0.01. The likelihood test, which ignores the statistics of the alternative hypothesis H_1, makes a decision based on $\prod_i P_0(X_q(i)'|Y(i))$ only. The results justify our choice to use the likelihood ratio test.

Fig. 10 also indicates that higher quantization precision offers better detection performance, but this comes at the cost of higher feedback data rate. Fig. 11 plots the ROC equal error rate versus the feedback data rate in kbps (for video at 30 fps) for different combinations of hypothesis testing and coding methods. The equal error rates are interpolated from the ROC curves as the points where the false acceptance rate equals the false rejection rate. Distributed source coding reduces the feedback data rate by 75% to 83% compared to conventional source coding at the same ROC equal error rates less than 10%.

4. CONCLUSIONS

We developed a rate-efficient quality monitoring scheme for mobile peers in peer-to-peer multicast video streaming using distributed source coding. In our scheme, each mobile peer sends a Slepian-Wolf coded projection of its received video to a quality monitoring server. We designed projections and analysis methods for the server to perform one of two tasks: quality estimation in terms of PSNR and tampering detection. Distributed source coding offers up to 85% feedback data rate savings compared to conventional source coding at the same performance. This means that a feedback data packet of size 1500 bytes can accurately estimate reconstructed PSNR or verify the integrity of at least 0.7 second of CIF video at 30 fps.



(a) CIF

(b) QCIF

Figure 9: Average PSNR estimation error versus feedback data rates for videos at 30 fps.

5. REFERENCES

[1] ITU-T Recommendation J.240: Framework for remote monitoring of transmitted picture signal-to-noise ratio using spread-spectrum and orthogonal transform, June 2004.

[2] A. Aaron, S. Rane, E. Setton, and B. Girod. Transform-domain Wyner-Ziv codec for video. In *SPIE Visual Communications and Image Processing Conference*, San Jose, CA, Jan. 2004.

[3] M. Castro, P. Druschel, A.-M. Kermarrec, A. Nandi, A. Rowstron, and A. Singh. Splitstream: High-bandwidth content distribution in a cooperative environment. In *International workshop on Peer-To-Peer Systems*, Berkeley, CA, Feb. 2003.

[4] K. Chono, Y.-C. Lin, D. Varodayan, and B. Girod. Reduced-reference image quality estimation using distributed source coding. In *International Conference on Multimedia and Expo*, Hannover, Germany, June 2008.

Digital Object Identifier: 10.4108/ICST.MOBIMEDIA2009.7532
http://dx.doi.org/10.4108/ICST.MOBIMEDIA2009.7532

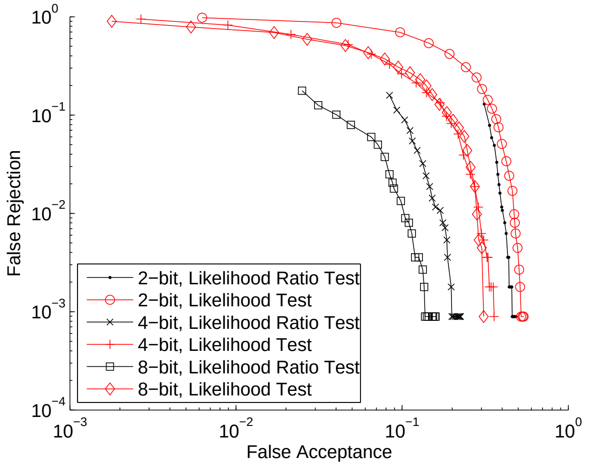


Figure 10: Receiver operating characteristic curves of tampering detection with different number of bits in quantization of X for CIF videos.

[5] Y.-C. Lin, D. Varodayan, and B. Girod. Image authentication based on distributed source coding. In *IEEE International Conference on Image Processing*, San Antonio, TX, Sep. 2007.

[6] A. Liveris, Z. Xiong, and C. Georghiades. Compression of binary sources with side information at the decoder using LDPC codes. *IEEE Communications Letters*, 6(10):440–442, Oct. 2002.

[7] J. Noh, P. Baccichet, F. Hartung, A. Mavlankar, and B. Girod. Stanford peer-to-peer multicast (SPPM) - overview and recent extensions. In *Picture Coding Symposium*, Chicago, IL, May. 2009.

[8] J. Noh, M. Makar, and B. Girod. Streaming to mobile users in a peer-to-peer network. In *International Mobile Multimedia Communications Conference*, London, England, Sep. 2009.

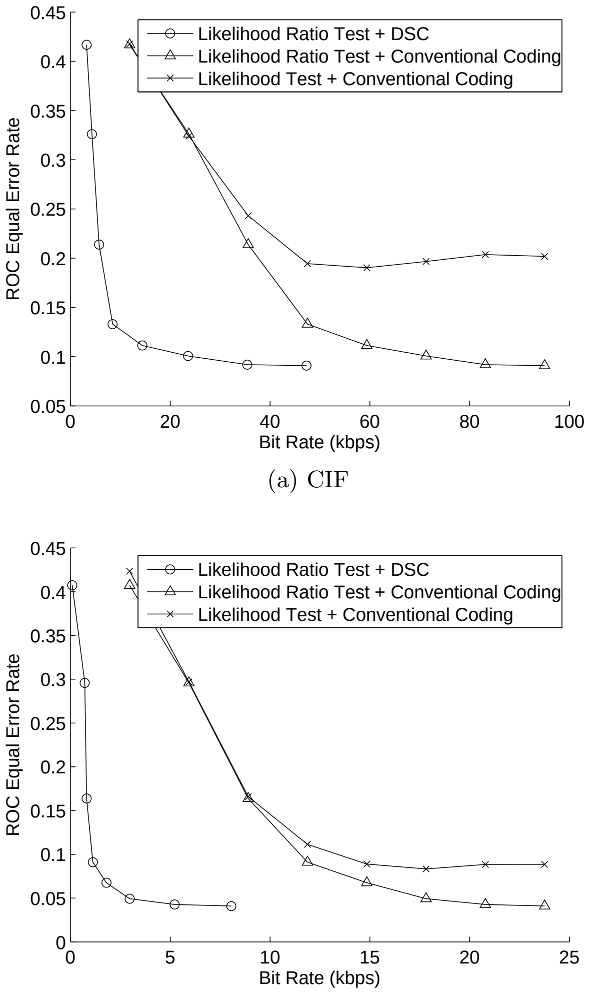
[9] V. N. Padmanabhan, H. J. Wang, P. A. Chou, and K. Sripanidkulchai. Distributing streaming media content using cooperative networking. In *International workshop on Network and operating systems support for digital audio and video*, Miami, FL, 2002. ACM.

[10] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, IT-19(4):471–480, July 1973.

[11] D. Varodayan, A. Aaron, and B. Girod. Rate-adaptive codes for distributed source coding. *EURASIP Signal Processing Journal, Special Section on Distributed Source Coding*, 86(11):3123–3130, Nov. 2006.

[12] D. Varodayan, Y.-C. Lin, and B. Girod. Audio authentication based on distributed source coding. In *IEEE International Conference on Acoustics, Speech, and Singal Processing*, Las Vegas, NV, April 2008.

[13] X. Zhang, J. Liu, B. Li, and Y.-S. Yum. Coolstreaming/donet: a data-driven overlay network for peer-to-peer live media streaming. In *IEEE INFOCOM 2005*, Miami, FL, March 2005.



(a) CIF

(b) QCIF

Figure 11: ROC equal error rates versus feedback data rates for videos at 30 fps.

Digital Object Identifier: 10.4108/ICST.MOBIMEDIA2009.7532
http://dx.doi.org/10.4108/ICST.MOBIMEDIA2009.7532