Modeling Packet-Loss Visibility in MPEG-2 Video

Sandeep Kanumuri, *Student Member, IEEE*, Pamela C. Cosman, *Senior Member, IEEE*, Amy R. Reibman, *Senior Member, IEEE*, and Vinay A. Vaishampayan, *Senior Member, IEEE*

***Abstract*—We consider the problem of predicting packet loss visibility in MPEG-2 video. We use two modeling approaches: CART and GLM. The former classifies each packet loss as visible or not; the latter predicts the probability that a packet loss is visible. For each modeling approach, we develop three methods, which differ in the amount of information available to them. A reduced reference method has access to limited information based on the video at the encoder's side and has access to the video at the decoder's side. A no-reference pixel-based method has access to the video at the decoder's side but lacks access to information at the encoder's side. A no-reference bitstream-based method does not have access to the decoded video either; it has access only to the compressed video bitstream, potentially affected by packet losses. We design our models using the results of a subjective test based on 1080 packet losses in 72 minutes of video.**

***Index Terms*—Packet-loss visibility, perceptual quality metrics, subjective testing, video quality.**

I. Introduction

When sending compressed video across today's communication networks, packet losses may occur. Network service providers would like to provision their network to keep the packet loss rate below an acceptable level, and monitor the traffic on their network to assure continued acceptable video quality. Traditional approaches to video quality assume that all packet losses affect quality equally. In reality, packet losses have different visual impacts. For example, one may last for a single frame while another may last for many; one may occur in the midst of an active scene while another is in a motionless area. Not all such packet losses are visible to the average human viewer. Thus, the problem of evaluating video quality given packet losses is challenging. As a first step toward developing a quality metric for video affected by packet losses, we address the problem of packet loss visibility in our current work.

Our long-term goal is to develop a quality monitor that is accurate, real-time, can operate on every stream in the network and answers the question, "How are the losses present in this particular stream impacting its visual quality?" In this paper, we focus on predicting the visibility of packet losses in MPEG-2 compressed video streams. Toward this goal, we develop statistical models to predict the visibility of a packet loss. We use a well known statistical tool called Classification and Regression Trees (CART) to classify each packet loss as visible or invisible. We use a generalized linear model (GLM) to predict the probability that a packet loss will be visible to an average viewer. The input to these models consists of parameters that can be easily extracted from the video near the location of the loss. We will also show how our GLM can be used to classify each loss as visible or invisible and compare its performance to that of CART.

We designed and conducted a subjective test that explores the impact of each packet loss individually. Viewers are shown MPEG-2 video with injected packet losses, and asked to indicate when they see an artifact in the displayed video. Data is gathered for a total of 1080 packet losses over 72 minutes of MPEG-2 video. Ground truth for the visibility of packet losses is defined by the results of our subjective tests. The frequency of visible packet losses will have a significant influence on the overall perceived quality; however, in this study, we do not explore this issue.

Fig. 1 illustrates different methods for quality assessment based on locations for measuring networked video. Full-Reference (FR) methods have access to the exact pixel values at both the encoder and decoder. Reduced-Reference (RR) methods have access to only certain key parameters extracted from the video at the encoder, but they have access to the exact pixel values at the decoder. If information at the encoder's side has to be transmitted reliably to the location of the quality monitor, FR methods will need much larger bandwidth compared to RR methods. No-Reference (NR) methods do not have access to any measurements at the encoder. There are two types of NR methods: NR-Pixel (NR-P) and NR-Bitstream (NR-B) methods. NR-P methods can measure the decoded video at the pixel level, while NR-B methods can measure only the compressed bitstream, not the decoded pixels. FR methods might give the highest accuracy, but NR-B methods are the best choice for network-based quality monitoring [1]. They can be deployed at different points in the network without the additional complexity of a decoder for every stream. In this paper, we explore RR, NR-P, and NR-B methods using both CART and GLM.

This paper is organized as follows. Section II reviews related work done previously. Section III gives an overview of MPEG-2 packet losses and their impact. Section IV describes our subjective test. Section V gives an introduction to our modeling approaches, CART and GLM. Section VI describes the objective factors (parameters) that we believe should be included in our

Manuscript received October 21, 2004; revised May 20, 2005. This work was supported in part by the National Science Foundation, the Center for Wireless Communications at UCSD and the UC Discovery Program of the State of California. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Pascal Frossard.

S. Kanumuri and P. C. Cosman are with the Department of Electrical and Computer Engineering, University of California-San Diego (UCSD), La Jolla, CA 92093-0407 USA (e-mail: skanumur@code.ucsd.edu; pcosman@code.ucsd.edu).

A. R. Reibman and V. Vaishampayan are with AT&T Labs—Research, Florham Park, NJ 07932-0971 USA (e-mail: amy@research.att.com; vinay@research.att.com).

Digital Object Identifier 10.1109/TMM.2005.864343

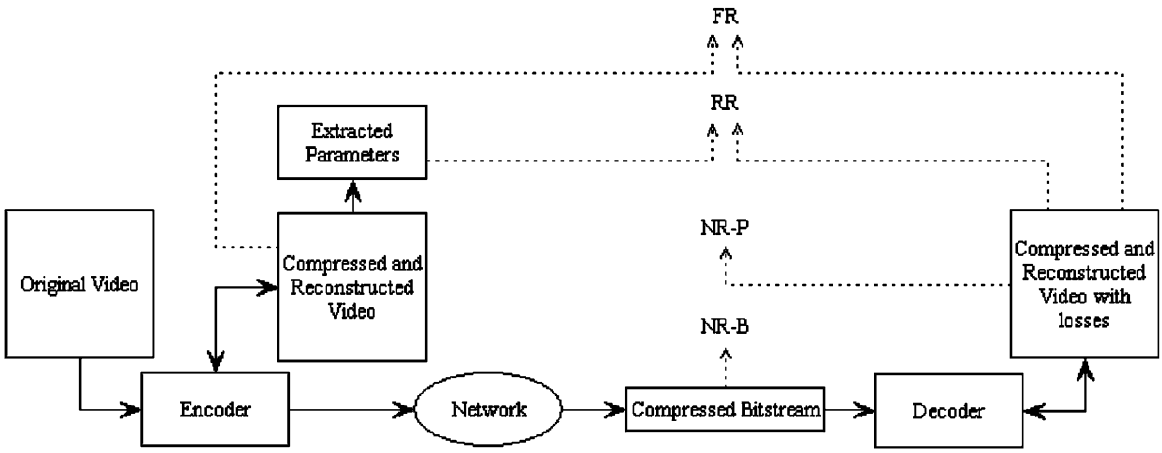

Fig. 1. Illustration of FR, RR, and NR methods.

models. Sections VII and VIII describe the results obtained by using CART and GLM respectively. Section IX describes classification based on GLM and compares the classification performance of CART with that of GLM. Section X concludes.

II. Related Work

Considerable research has been done on developing objective perceptual quality metrics for compressed video not affected by network losses. For example, an FR metric based on a multichannel model of the human visual system was proposed in [2] that generates continuous estimates of perceived quality for low bit rate compressed video. Gastaldo *et al.* [3] used a Neural Network approach to design an objective quality assessment algorithm for MPEG-2 video streams without decoding. While these metrics are designed to predict the quality degradation caused by compression artifacts, they are not tailored to handle the degradation caused by network impairments.

The joint impact of encoding rate and ATM cell losses on MPEG-2 video quality was studied in [4], [5]. They show the existence of an optimal coding rate for a given loss ratio. A similar study [6] on quality-of-service (QoS) performance of end-to-end video transmission also showed that an increase in video bit-rate may improve video quality only when cell loss ratio is below a certain level. In both cases, the quality of video is judged based on an existing picture quality model and not based on subjective tests. A framework for employing objective perceptual quality assessment methods, evaluating the quality of audio, video and multimedia signals, to model network performance is demonstrated in [7]. In their paper, they focus on modeling network performance of multiplexed VOIP calls using the perceptual analysis approach.

Much of the effort to understand the visual impact of packet losses [8]–[11] has focused on the goal of understanding the average quality of typical videos subjected to average packet loss rates (PLR). Video conferencing is studied in [8] using the average judgment of consumer observers to examine the relative importance of bandwidth, latency and packet loss. The impact of packet loss on the mean opinion score (MOS) of real-time streaming media was studied in [9] for Microsoft Windows Media encoder 9 (beta version) video. A neural network was trained in [10] to viewer responses on the ITU-R 9-point quality scale, when a single 10-s sequence was subjected to different bandwidth, frame rate, packet loss rate, and I-block refresh rate.

Hughes *et al.* [11] use MOS to evaluate the subjective quality of VBR video subjected to ATM cell loss over a 10-s period. They show that performance is sensitive not only to the magnitude of the bursts, but also to their frequency. "Very different" results were obtained for different sequences. Other challenges identified by these authors [11] were a) many different realizations of both packet loss and video content are necessary to reduce the variability of viewer responses; b) very low PLRs are difficult to explore because the typical test period (10 s) is so short that typical realizations may have no packet losses; and c) the "forgiveness effect" causes viewers to rank a long video based on more recently viewed information.

In [12], two different subjective testing procedures, namely Single Stimulus Continuous Quality Evaluation (SSCQE) and Double Stimulus Impairment Scale (DSIS), were compared. The first procedure shows one stimulus to the subjects, the second two. The data obtained with these procedures was found to be highly correlated and of comparable prediction accuracy. Further, blockiness, blurriness, and jerkiness metrics were not able to accurately predict viewers' subjective responses to packet losses.

In [13], subjective tests were conducted to validate the usefulness of an existing spatio-temporal model for predicting quality in the presence of packet losses. Both one- and two-layer encodings were studied. According to the study, the model examined did not have a significant advantage over PSNR.

Typically, these studies [8]–[13] use subjective tests to evaluate quality using MOS. However, the MOS quality rating methodology has a number of difficulties, as detailed in [14]. First, the impairment (or quality) scales are generally not interpreted by subjects as having equal step-size, and labels in different languages are interpreted differently. Second, subjects tend to avoid the end-points of the scales. Third, the term "quality" itself is actually not a single variable, but has many dimensions.

Instead of asking viewers to respond with a scaled rating (i.e., MOS), viewers in recent subjective tests have been asked simpler questions. For example, in [15], [16], viewers were asked to indicate when an artifact was visible. In [17] and [18], viewers were asked to adjust the artifact strength until it became just

visible. In these studies, the artifacts were imposed on natural, not synthetic images and videos. Answers from the subjective viewers were then analyzed to obtain a deeper understanding of the factors that affect visual quality [15]–[18]. Our subjective tests in this work were designed with similar motivation.

III. Effect of a Packet Loss

Video is typically packetized in one of two ways: it can be segmented and packetized into small fixed-size packets (such as ATM cells or MPEG-2 Transport Stream packets), or a variable-sized packet can contain one or more slices. In both cases, a packet loss will cause the loss of one or more slices. Typical scenarios for fixed-size packetization are a) a packet contains part of one slice, b) a packet contains the end of one slice and the beginning of another, and c) a packet contains a frame header. These will cause the loss of a) one slice, b) two slices, and c) an entire frame. Therefore, in this paper, we focus on exploring the impact of these three situations.

The initial error caused by a packet loss propagates in space and time as a result of the video decoding algorithm. The exact error due to the packet loss can be completely described by a) the initial error for each macroblock in the lost packet, b) the macroblock type, and c) motion information for subsequently received macroblocks [1]. The latter two control the temporal duration and spatial spread of the error.

The initial error induced by a packet loss depends on the error concealment strategy used by the decoder. A typical concealment strategy, used here, is zero-motion concealment, in which a lost macroblock is concealed using the macroblock in the same spatial location from the closest reference frame. In this case, the initial error is simply the difference between the current encoded frame and the closest reference frame for the affected macroblocks.

We expect the visibility of a loss to depend on a complex interaction of its location, the video encoding parameters, and the underlying characteristics of the video signal itself. For example, the texture and motion of the underlying signal may potentially mask the error. To isolate the impact of the various parameters, one approach could be to inject different error amplitudes against an identical signal background, as was done in [16] for blocky, blurry and noisy artifacts. However, for packet losses, the error itself is highly correlated with the underlying signal and so we do not have control over the amplitude of the error. Therefore, we must take a different approach.

When choosing the packet losses to inject for our subjective tests, we have independent control over the location, initial spatial extent and temporal duration of each loss we inject. The other factors depend on the signal. Thus, we choose whether to lose a single slice, double slice or an entire frame. We also choose the loss to be in a B-frame (which would last a single frame) or in a reference frame (which will last until the next I-frame). In choosing the vertical location of the loss, we uniformly distribute the losses within the frame.

IV. Subjective Tests

For the subjective tests, we can conduct either a single-stimulus test or a double-stimulus test. In a single-stimulus test, only the video being evaluated (here, video with packet losses) is shown. The reference or original video is not shown. In a double-stimulus test, both videos are shown. We conducted a single-stimulus test because the test mimics the perceptual response of a viewer who does not have access to the original video, which is a natural setting for most applications. The viewer bases his/her judgment on the lossy video only.

In the test, the viewers' task is to indicate when they saw an artifact, where an artifact is defined simply as a glitch or abnormality. We wanted viewers to be immersed in the viewing process and not scrutinizing the video for any possible impairment. Thus we chose DVD-quality MPEG-2 video from travel documentaries. Audio was not presented. The video sequences had a resolution of 720×480 pixels and had 30 frames per second. The average bitrate for the sequences varied from 3.5 Mbps to 4.4 Mbps. In our encoding structure, we had two consecutive B-frames between reference frames and we had an I-frame every 13 frames. Zero-motion error concealment using the closest reference frame was used whenever there was a packet loss. This presumes a minimum amount of intelligence on the part of the decoder. Decoders that use sophisticated error concealment methods may have fewer visible packet losses. However, since we would like to predict the visibility of packet losses in the network, without necessarily knowing which decoder the viewer is using, we assume only this minimal error concealment strategy.

The video sequences we chose contain a wide variety of scenes with different types of camera motion (panning, zooming) and different types of scenes with varying types of motion. The high motion scenes included bike racing, bull fighting, dancing and flowing water. The low motion scenes included showing maps, historic buildings and structures. The videos also had scenes with varying spatial content such as a bird's eye view of a city, a crowded market, portraits, sky and still water, etc.

We chose 12 6-min video sequences, for a combined length of 72 min. We grouped the sequences into four sets, each consisting of three sequences. This limited a viewing session to 18 min so as not to tire or bore the viewers. During each session, a viewer evaluated a set of video sequences with a short break after each sequence. Some viewers participated in more than one viewing session, although never on the same day. Each set of video sequences (and hence each packet loss) was evaluated by 12 viewers.

The age of the viewers varied from 25 to 60 years. All the viewers had either normal or corrected-to-normal vision. None of the subjects had previous experience in video quality except for one expert subject, who evaluated all the four sets of video sequences. The profession of the viewers was either technical or secretarial.

Viewers were told that the videos they were watching would have impairments caused by packet losses, and that when they saw something unexpected in the video like a glitch, they should respond by pressing the space bar. They were asked to keep their finger on the space bar so they would not be distracted by that task. All the tests were conducted in a well lit room using the same monitor and settings. Viewers were positioned approximately six picture heights from the screen. We observed that

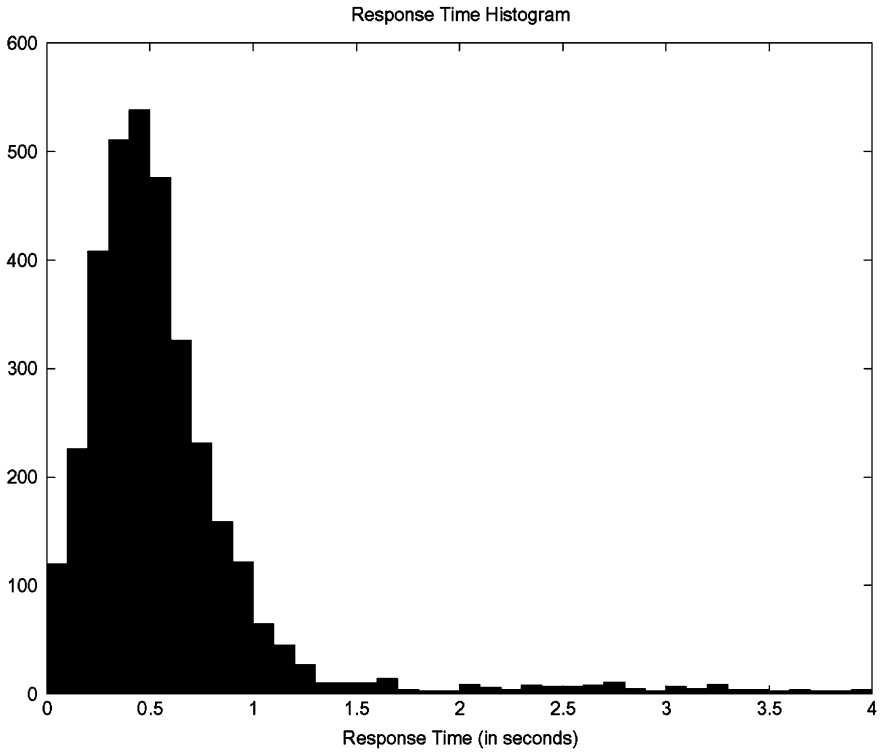

Fig. 2. Histogram of response times.

the viewers were able to perform the task without any difficulty although they were untrained.

A total of 1080 packet losses were randomly injected into these videos. We are not trying to simulate a typical packet loss scenario, which may include bursty losses, but are instead trying to answer the question "What causes a single packet loss to result in a visible artifact?" Therefore, we introduce isolated losses randomly into each nonoverlapping 4-s interval. To ensure viewers have adequate time for responding, we randomly inject a packet loss in the first three seconds of each interval and allow a one-second guard interval during which the decoder can recover and the user can respond.

We distributed the losses such that 30% affected an entire frame, 10% affected two adjacent slices, and 60% affected a single slice. Here we consider a slice to be one horizontal row of macroblocks. Further, we chose to have 30% of the losses be in B-frames (and hence have a temporal duration of one frame), and the remaining 70% evenly distributed across the available P- and I-frames in the 3-s interval.

The output of the subjective test was a set of files containing the times that the viewer pressed the space bar relative to the start of the video. We processed these to create a matrix with 1080 rows and 12 columns, whose entries indicate whether a viewer responded to a packet loss or not. If a viewer pressed the space bar within two seconds after a packet loss occurred, he/she is considered to have responded to that packet loss. Otherwise, he/she is considered not to have responded to the packet loss. Fig. 2 shows a histogram of the response times (time difference between the packet loss and the key press). From the histogram, we can see that 91% of the responses occur within one second of the packet loss, and 97% of them occur within 2 s. The average response time is 0.6 s. We believe that a viewer who saw a packet loss should be able to respond within two seconds and we consider the responses that come after two seconds to be false alarms. The ground truth for the probabilities of visibility of a packet loss was defined from these viewers' responses. The probabilities were calculated as the number of viewers who saw the packet loss divided by 12.

Viewers were not told the pattern of injected packet losses. There is a concern, however, that while viewing the video they might infer that a packet loss occurs in every 4-s interval. If viewers were able to predict this, it might bias their responses. To analyze this, we examined the time between adjacent packet losses (Inter Packet loss Interval), and the time between adjacent responses of a viewer (Inter Response Interval). Fig. 3 shows that the density of *Inter Packet loss Interval* (IPI) is triangular with a minimum, mean, and maximum of 1, 1, and 7 s, as expected. Also shown in Fig. 3 is the density of *Inter Response Interval* (IRI). Its long tail out to 150 s is not shown; instead all the samples larger than 40 s have been assigned to the last bin of the histogram which explains the spike in the tail. This density has a peak near four seconds. However, only a small percentage (11.3%) of the IRI samples are between 3.5 and 4.5 s, which indicates that viewers did not infer that a packet loss occurs in every 4-s interval and begin to anticipate an artifact.

V. Statistical Background—CART and GLM

Our first goal is to classify packet losses as visible or invisible. We use a well known statistical software called CART. Our second problem is a regression problem: we want to predict the probability of packet loss visibility. We use logistic regression, a GLM, to solve this problem. This section gives a brief introduction to these approaches.

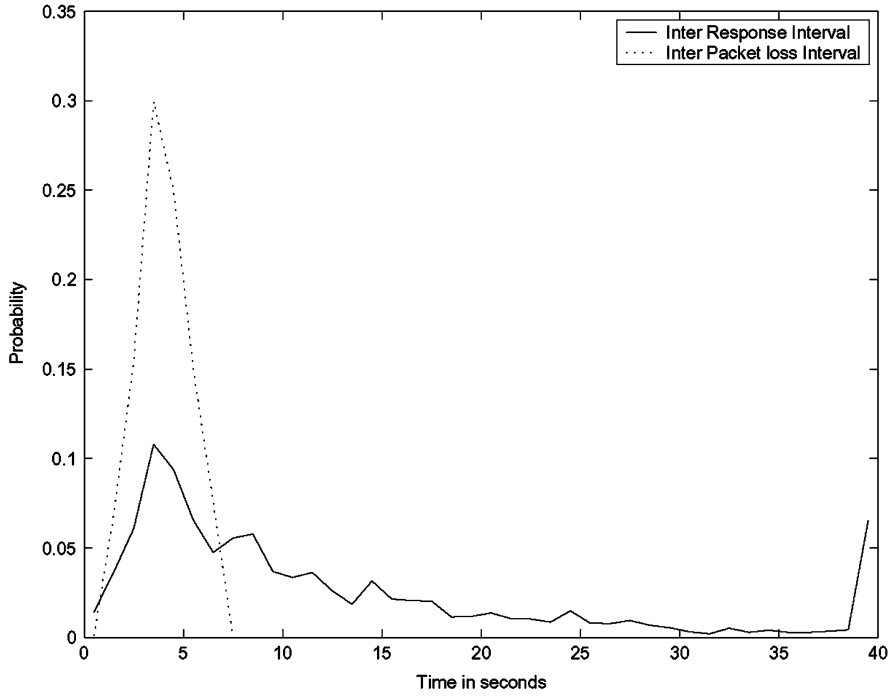

Fig. 3. Histogram of times between adjacent losses.

A. CART

CART is a tool for tree structured data analysis introduced by Breiman *et al.* [21]. CART generates its results in the form of decision trees. This allows CART to handle massively complex data while producing classifiers that are easy to understand. The decision criteria give us insight into what causes a packet loss to be visible, and can be compared with intuition. Other approaches such as artificial neural networks tend to be harder to interpret, since they may involve weighted sums of large numbers of input parameters, whose individual effects cannot be discerned.

CART uses binary recursive partitioning. The process is binary because parent nodes are split into exactly two child nodes. It is recursive because the process can be repeated by treating each child node as a parent. The key elements of a CART analysis are a set of rules for 1) splitting each node in a tree, 2) deciding when a tree is complete, and 3) assigning each terminal node to a class outcome (or predicted value for regression).

We wish to select each split of a subset so that the data in each of the descendant subsets are "more pure" than the data in the parent subset. CART usually splits data based on a threshold applied to the value of a variable. At each node, CART searches through all possible thresholds for all variables and picks the variable and the threshold that give the best split for that node. The best split is based on a purity criterion, such as the Gini index of diversity [21].

CART continues to split until all the elements in a node belong to the same class or the number of elements in a node is less than a predetermined threshold. Using this process, CART forms the largest possible tree which is later pruned to get the final tree, the one that gives the best cross-validation accuracy among all the pruned trees.

For each terminal node, CART assigns a class that minimizes the misclassification cost incurred on account of this assignment. If the misclassification costs for all the classes are the same, then the class with highest representation in the terminal node will be assigned as the class for the terminal node.

B. GLM

We model the probability of visibility using a GLM, which is an extension of classical linear models [22]. Logistic regression is a type of GLM which is a natural model to predict the parameter p of a binomial distribution [22]. Let $y_1, y_2, \ldots, y_N$ be a realization of independent random variables $Y_1, Y_2, \ldots, Y_N$ where Y_i has binomial distribution with index m_i and parameter p_i. Let $\mathbf{y}, \mathbf{Y}$ and $\mathbf{p}$ denote the N-dimensional vectors represented by y_i, Y_i, and p_i respectively. We will model the parameter p as a function of P factors. Let $\mathbf{X}$ represent a $N \times P$ matrix, where each row i contains the P factors influencing the corresponding parameter p_i.

A generalized linear model between $\mathbf{p}$ and $\mathbf{X}$ can be represented as

$$g(\mathbf{p}) = \gamma + \sum_{j=1}^{P} \mathbf{x_j} \beta_j \tag{1}$$

where $g(.)$ is called the link function, which is typically nonlinear, $\mathbf{x_j}$ is the jth column of $\mathbf{X}$ and $\beta_1, \beta_2, \ldots, \beta_P$ are the coefficients of the factors. Coefficients $\boldsymbol{\beta}$ and the constant term γ are usually unknown and need to be estimated from the data. For logistic regression, the link function is the logit function, which is the canonical link function for the binomial distribution. The logit function is defined as

$$g(p) = \log\left(\frac{p}{1-p}\right). \tag{2}$$

Given N observations, we can fit models using up to N parameters. The simplest model (Null model) has only one parameter: the constant γ. At the other extreme, it is possible to have a model (Full model) with as many parameters as there are observations. The goodness of fit for a generalized linear model can be characterized by its deviance, defined in [22]. By definition, the deviance for the Full model is zero and the deviance for all other models is positive. A smaller deviance means a better model fit. The deviance can be shown to be asymptotically distributed as $\chi^2_{n-(P+1)}$, where $(P+1)$ is the total number of parameters fitted for the model [22]. Furthermore, the difference in deviance between two models is also known to be approximately distributed as χ^2_k where k is the difference in the number of parameters estimated for each model. This is very useful in determining the significance of different factors.

We use the statistical software R [23] for our model fitting and analysis. To obtain the model parameters, R uses an iteratively re-weighted least-squares technique to generate a maximum-likelihood estimate. After fitting a particular model, the importance of each factor in the model can be evaluated by the resultant increase in deviance when we remove that factor from the model. This increase can be compared with the appropriate χ^2 statistic to compute the p-value for this factor. If the p-value is less than 0.05, then the factor is significant at the 95% level. We represent the observed probability of visibility as $\tilde{p}$ and the predicted probability of visibility as $\hat{p}$.

VI. Factors Affecting Visibility

In this section, we describe the objective factors that we believe will be useful in modeling the visibility of a packet loss. We focus primarily on factors that are easily extracted from the video, as our main goal is to develop an NR-B method for evaluating video quality within a network. In the following sections, we will explore the usefulness of these factors in our models.

These objective factors can be classified into two types: content-independent factors and content-specific factors. Content-independent factors depend on the location of the packet loss in the MPEG-2 bitstream, but do not depend on the content of the video. Content-independent factors can therefore be calculated exactly from the lossy bitstream itself. Content-specific factors depend on the content of the video at the location of the packet loss. Content-specific factors can be calculated exactly at the encoder side, by using the original bitstream without losses. However, these content-specific factors cannot be exactly obtained from a bitstream in which packets are already lost.

The first content-independent factor we consider characterizes the duration of time an error persists. We start by using the temporal duration (TMDR), which represents the maximum number of frames that may be affected by the packet loss. In our data, this varies from one to thirteen because of the encoder's prediction structure. An error in a B-frame lasts a single frame. An error in a reference frame may propagate in time but will always be removed by the next I-frame. Our previous research [24] showed that if $\mathrm{TMDR} = 1$, the packet loss is almost always invisible. However, the correlation coefficient between the number of viewers who saw a packet loss and TMDR is only 0.051, which is very low.

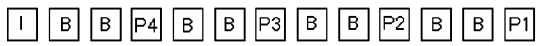

Fig. 4. FRAMETYPE value for different frames in a GOP.

Thus, for GLM, we also explore two alternate ways to represent the temporal duration. The first is the boolean variable, BFRAME, which is set whenever the packet loss occurs in a B-frame. The second is the categorical variable FRAMETYPE, which has six levels depending on the type of frame in which the packet loss occurs. These six levels correspond to a B-frame, four P-frames with a different distance to the next I-frame, and an I-frame. We call these levels B, P1, P2, P3, P4, and I. Fig. 4 illustrates how these frames occur in the GOP structure of our videos. FRAMETYPE captures all the information in the temporal duration of a packet loss. For example, a packet loss in a P3 frame will have a temporal duration of 9. In the GLM, a categorical variable with N levels is treated as a vector of $N-1$ boolean variables. (The N-th level is represented by setting all $N-1$ boolean variables to zero.) Thus for FRAMETYPE, we considered five boolean variables: FRAMETYPE-P1, FRAMETYPE-P2, FRAMETYPE-P3, FRAMETYPE-P4, and FRAMETYPE-I. FRAMETYPE-B is considered default and its effect is included in the constant term.

The second content-independent factor we consider is spatial extent (SPTXNT) which represents the number of slices affected by the packet loss. In our case, it is either 1, 2, or 30 corresponding to single slice, double slice or frame loss respectively. SPTXNT can be treated as an ordinal variable, taking on values 1, 2, and 30, or as a categorical variable with three levels to distinguish the cases of single slice, double slice, and frame loss errors. In the remainder of the paper, SPTXNT refers to the ordinal variable except where categorical is stated. For SPTXNT (categorical), in the context of GLM, we consider two boolean variables: SPTXNT-2 and SPTXNT-30. SPTXNT-1 is considered default.

The third content-independent factor we consider is the vertical position (HGT) of the error induced by the packet loss. HGT is the number of the topmost slice affected by the packet loss, where the slices are numbered from 0 to 29 from top to bottom. This factor captures the varying attention viewers have on different regions in the frame. In our study, the values of each of the content-independent factors can be controlled at the time of choosing which losses to introduce. Since the content-independent factors can be extracted exactly from the lossy bitstream, they are identical across our RR, NR-P, and NR-B models.

Content-specific factors must include some description of motion. We use the MPEG-2 bitstream to extract the received motion vectors corresponding to each frame, so no motion estimation step is needed by the quality monitor. These motion vectors are scaled down by the distance to the reference frame such that the scaled motion vectors represent the motion per frame. Since MPEG-2 uses x and y directional motion vectors, the most immediate way to express motion for purposes of predicting visibility is to use these vector components. MOTX and MOTY represent the scaled motion vector components in x and y directions respectively, averaged across all macroblocks

TABLE I
DESCRIPTION OF FACTORS AFFECTING VISIBILITY

Factor Name	Description
TMDR	Time Duration: Maximum number of frames affected by the packet loss
BFRAME	Boolean variable set if packet loss occurs in a B-frame
FRAMETYPE	Type of frame in which packet loss occurred
SPTXNT	Spatial Extent: Number of slices lost
SPTXNT (categorical)	Number of slices lost, categorized as single, double or frame loss
HGT	Height: Number of the topmost slice lost
MOTX	Average motion in x-direction
MOTY	Average motion in y-direction
VARMX	Variance of motion in x-direction
VARMY	Variance of motion in y-direction
MOTM	Magnitude of overall motion
MOTA	Angle of overall motion
VARM	Variance in overall motion
HIGHMOT	Boolean variable set when MOTM > 0.707
RSENGY	Average residual energy per pixel after motion compensation
IMSE	Mean square error per lost pixel

initially affected by the loss. The motion vector variance is denoted by VARMX and VARMY respectively.

To explore whether visibility might be governed more by total motion than by x and y directional motion, we introduce MOTM to represent the magnitude, MOTA to represent the angle, and VARM to represent the variance in overall motion. We calculate them as follows:

$$\mathrm{MOTM} = \sqrt{\mathrm{MOTX}^2 + \mathrm{MOTY}^2} \tag{3}$$

$$\mathrm{MOTA} = \arctan\left(\frac{\mathrm{MOTY}}{\mathrm{MOTX}}\right) \tag{4}$$

$$\mathrm{VARM} = \mathrm{VARMX} + \mathrm{VARMY} \tag{5}$$

We also define HIGHMOT, a boolean variable, to be set when $\mathrm{MOTM} > 0.707$. This threshold was set to correspond to motion that is greater than half a pixel per frame in both x and y directions.

Content-specific factors also include residual energy (RSENGY) and initial mean square error (IMSE). RSENGY denotes the average residual energy per pixel after motion compensation for the lost slices. IMSE is the MSE per pixel, after loss and error concealment, in the frame affected by the packet loss averaged over only those pixels in the lost slices. This involves a slight modification of our definition of IMSE from our previous work [24]. Previously, IMSE was averaged over all the pixels in the frame (whether lost or not) and so IMSE for single and double slices was very small (by a factor of 30 approximately) compared to IMSE for a frame loss. With our new definition, IMSE has the same range of values for single, double and frame losses and is not dependent on SPTXNT. Table I summarizes the descriptions of all the factors.

The content-specific factors described above can be extracted exactly using both the complete bitstream (available at the encoder) and the decoded pixels. For the RR method, the content-specific factors can be extracted at the encoder for all slices, and this information can be made available to the quality monitor via reliable means. This information is then combined with the knowledge of which slices are lost to generate the content-specific factors for lost slices. These factors can be exactly obtained only for an RR method. NR-P and NR-B methods must estimate these factors for the missing slices. Further, to compute IMSE, decoded pixels are necessary; however, these are unavailable to the NR-B method since an NR-B method has access only to the compressed bitstream and not the decoded pixels.

For the NR-P and NR-B methods, the parameters MOTX, MOTY, VARMX, VARMY (and thereby MOTM, MOTA, VARM, and HIGHMOT), and RSENGY are extracted directly from the bitstream for all *received* slices. Parameters for the missing slices are then estimated using one of two approaches. The first approach estimates the parameter using co-located slices in the previous frame. The second approach estimates the factor using spatially neighboring slices in the same frame. We tried each approach on one video sequence and found that the first approach performed best for all the above mentioned parameters.

For the NR-P case, IMSE is computed for all received slices, where IMSE for received slices is defined to be the IMSE that would have resulted if the slice had been lost. The second approach above was found to be more effective for estimating the IMSE of the missing slices. For the NR-B method, neither of the above two approaches can be used to estimate IMSE since the decoded pixels are not available. Thus, for the NR-B case, we use the approach described in [1], which extracts and estimates additional parameters (such as mean, spatial correlation, spatial variance) using the DCT coefficients from the received slices, to estimate IMSE for the missing slices.

VII. CART Results

For classification purposes, we define a packet loss to be *visible* if 75% or more viewers responded to it. Similarly, a packet loss is *invisible* if 25% or fewer viewers responded to it. The remaining losses are *indeterminate*. Of the 1080 total packet losses shown to viewers, 732 were invisible, 195 were visible and 153 were indeterminate. For the classification problem, we do not concentrate on the 14% of losses that were indeterminate, but instead focus on understanding the 927 visible and invisible losses.

We compare two objective classifiers that classify each packet loss to be visible or invisible to an "average" human observer. Both the classifiers are based on those used in our previous work [24]. Each classifier is a decision tree; the classifier traverses a tree where the path at each node depends on a binary decision using one of the factors discussed in Section VI. During the formation of the tree, a node is split to minimize the probability of misclassification.

The first classifier we consider is $\mathbf{Root{-}tree + CART}$. $\mathbf{Root{-}tree}$ is based on our earlier observations [24] in the RR case regarding the impact of short temporal duration, small motion and spatial extent. Of all packet losses with $\mathrm{TMDR} = 1$, only one is visible and the remaining 119 are invisible. Only 11 out of 330 packet losses with both MOTX and MOTY less than 0.5 (half a pixel) are visible. Of the full-frame packet losses, 39% are visible, while only 13% of the single- and double-slice losses are visible.

$\mathbf{Root{-}tree}$ then consists of the following decisions. First, all packet losses with temporal duration of one frame $(\mathrm{TMDR} = 1)$ are classified as *invisible*. Second, all packet losses with small motion, defined by $(\mathrm{MOTX} \leq 0.5$ & $\mathrm{MOTY} \leq 0.5)$, are classified as *invisible*. The application of $\mathbf{Root{-}tree}$ results in 12 and 35 misclassifications out of 450 and 604 cases for the RR and both NR cases respectively. Both the NR cases have the same result with the $\mathbf{Root{-}tree}$ since the variables involved have the same values for both NR-P and NR-B cases. Next, we split the tree based on the initial spatial extent $(\mathrm{SPTXNT} < 15)$ without making any decision. The threshold for the split on SPTXNT could be any value between 3 and 30; the goal of the split is merely to separate slice losses (single and double) from frame losses. At this stage, we apply CART to classify the data in each of the two nodes.

The second classifier we consider is CART itself. This classifier is designed by applying CART to the entire data set using the factors TMDR, SPTXNT, HGT, MOTX, MOTY, VARMX, VARMY, RSENGY, and IMSE.

We also consider the two classifiers described above with modified motion variables MOTM and VARM instead of MOTX, MOTY, VARMX, and VARMY. The $\mathbf{Root{-}tree}$ is slightly modified to incorporate the MOTM variable instead of MOTX and MOTY. Now, packet losses with small motion are defined by $\mathrm{MOTM} < 0.707$ instead of $\mathrm{MOTX} \leq 0.5$ & $\mathrm{MOTY} \leq 0.5$. The threshold for MOTM is obtained by using the thresholds for MOTX and MOTY in the equation for MOTM.

Fig. 5 shows the cross-validation classification accuracy for the two classifiers with and without using modified motion variables under different methods RR, NR-P and NR-B. As we can see, both the classifiers $\mathbf{Root{-}tree + CART}$ and CART show an overall improved performance with this treatment of motion variables. CART has a slight edge over $\mathbf{Root{-}tree + CART}$. As expected, the RR method always performs better than the NR methods, but the improvement in performance is not large. The maximum improvement in performance observed is 3.6% which occurs over the NR-P method, with the $\mathbf{Root{-}tree}+\mathrm{CART}$ classifier when MOTX and MOTY variables are used.

Fig. 6 shows the classification tree obtained using CART for the NR-B case. The terminal nodes are represented by ovals and the internal nodes are represented by rectangles. Each internal node is split on the variable shown in its rectangle. If a terminal node is marked VIS, then all packet losses that fall into this node are classified as visible. Similarly, if a terminal node is marked INV, all packet losses that fall into this node are classified as invisible. The tree has splits based on MOTM and IMSE at the top, which shows that they are very important factors. This classifier tree performs the best in the NR-B case with a cross validation accuracy of 91.2%. Most of the splits of the tree are intuitively reasonable; for example, branches on the left half of the tree with lower IMSE or SPTXNT lead to terminal nodes labeled INV. However, we find some counter-intuitive splits close to some terminal nodes. One of these is in the left half of the tree, where the data set is split on TMDR with a threshold of 12.5, and then it is immediately split again with a threshold of 2. Both the data less than 2 and the data greater than 12.5 are classified as invisible, which is counter-intuitive. The other occurrence of a counter-intuitive split is in the lower right of the tree, where there is a repeated split on HGT. We will only incur an additional ten classification errors out of 927 (approximately 1%) during resubstitution if we remove these counter-intuitive splits. Since these splits classify a very small fraction of the data, they do not affect the overall performance of the classifier significantly. We believe that these spurious splits are a result of few available data points to judge the split and can be rectified with a larger data set.

VIII. Logistic Regression Results

In this section, we apply logistic regression, a type of GLM, to the problem of estimating the probability that a packet loss is visible to an average viewer. We use the factors extracted from our RR, NR-P, and NR-B methods to derive a separate model for each case.

We use the word "model" to characterize the set of factors which comprise the matrix $\mathbf{X}$, introduced in Section V-B. We note that for each "model," we actually consider three: one for each of the RR, NR-P, and NR-B cases. The distinction among the three lies in whether the content-specific factors are extracted exactly, or estimated as described in Section VI.

We explored a number of models with different sets of factors, to determine the best way to characterize sequence motion and loss-duration for our objective. Our final model, denoted Model 3, uses the factors FRAMETYPE, SPTXNT (categorical), MOTM, HIGHMOT, VARM, RSENGY, IMSE, and HGT to predict the probability of visibility of a packet loss.

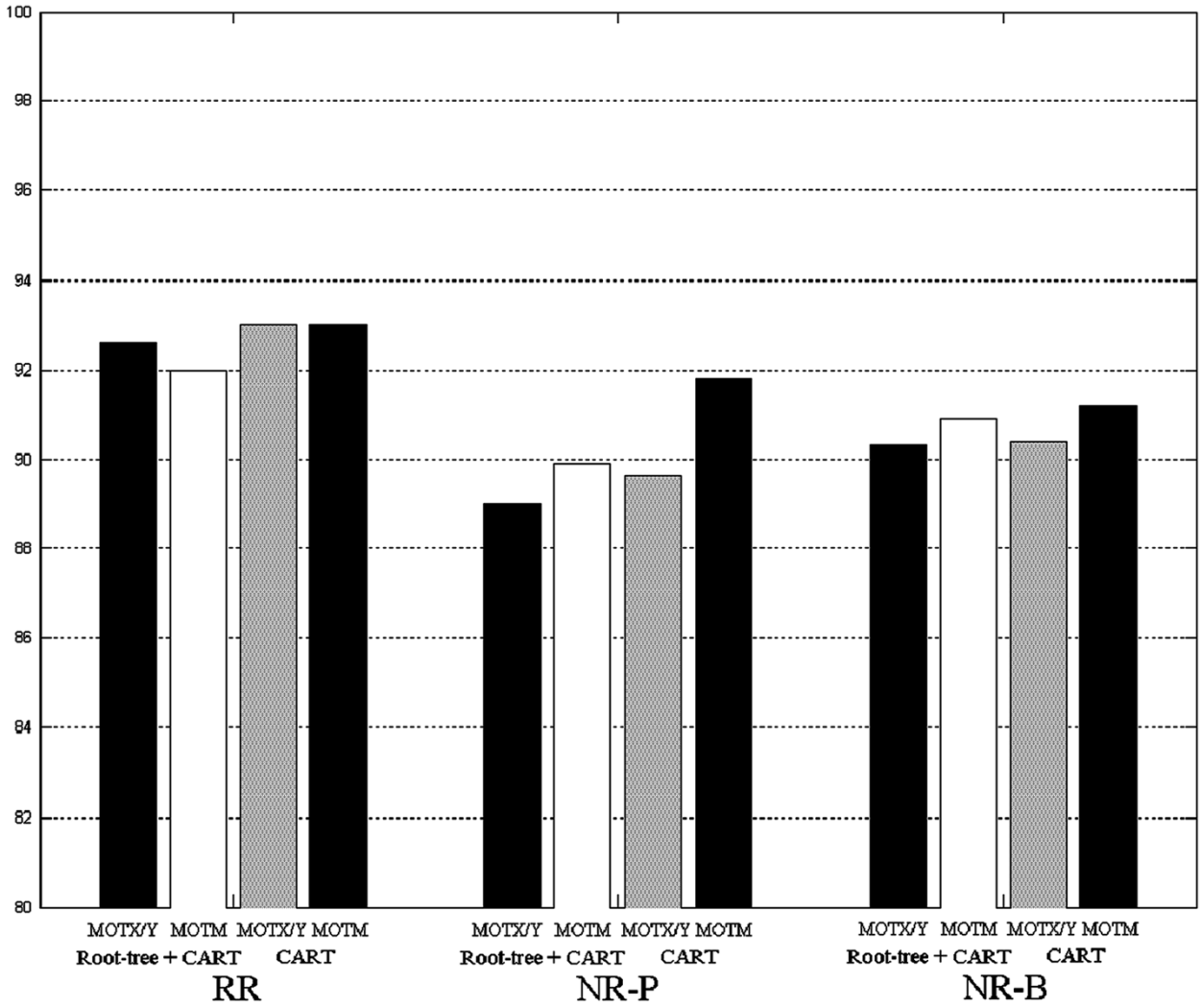

Fig. 5. Comparison of classification accuracy. For example, the first bar on the left shows that the RR method using $\mathrm{Root-tree+CART}$, with motion information expressed as x and y directional motion, achieves a cross-validated correct classification of 92.6%.

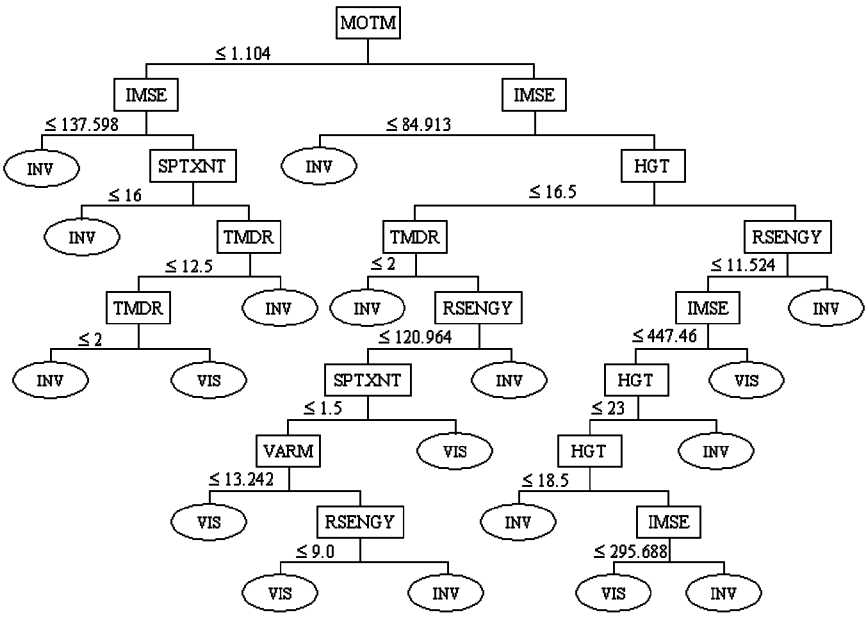

Fig. 6. CART classifier tree in the NR-B case.

TABLE II
COEFFICIENTS FOR MODEL 3 IN NR-B

factor	coefficient
constant γ	-4.53
FRAMETYPE-P1	2.116
FRAMETYPE-P2	2.104
FRAMETYPE-P3	2.117
FRAMETYPE-P4	2.188
FRAMETYPE-I	5.326e-01
SPTXNT-2	7.161e-01
SPTXNT-30	1.54
MOTM	4.212e-01
HIGHMOT	1.398
VARM	-1.144e-02
RSENGY	-6.902e-03
IMSE	9.890e-04
HGT	-2.797e-02

The deviances obtained with this model for the RR, NR-P and NR-B cases are 4797.6, 5106.7, and 5115.7 respectively with 1066 degrees of freedom for the χ^2 distribution, while the deviance for the null model (Model 0) is 9254.8 with 1079 degrees of freedom. The MSE obtained between actual probability $\tilde{p}$ and predicted probability $\hat{p}$ is 0.0565 for RR, 0.0608 for NR-P, and 0.0611 for the NR-B case.

To verify the applicability of this model to new data, we perform a 4-fold cross-validation procedure. For this, we use the data from three out of the four sets of video as a training set. The data from the remaining set is used for testing. We repeat this process four times, each time choosing a different set for the testing set. Thus we have a predicted probability, $\hat{p}$, for each packet loss obtained when the packet loss was not used for training. The MSE obtained between $\tilde{p}$ and $\hat{p}$ during cross-validation for Model 3 is 0.0627 for RR, 0.065 for NR-P, and 0.0647 for the NR-B case. This shows that the model continues to perform well when encountering new data.

The coefficients (γ and βs) for the final model (Model 3) in the NR-B case are tabulated in Table II. The values of the coefficients do not necessarily convey the importance of corresponding factors because these factors have different variances and ranges. However, the sign of the coefficients is important and informs whether a packet loss is more visible with a high or low value for a factor. We can make the following conclusions based on the coefficient values:

High values of MOTM and IMSE cause a packet loss to be more visible. Visibility of losses in I, P, and B frames decreases in that order. A large spatial extent (SPTXNT) increases the visibility of a packet loss. High values of VARM and RSENGY cause a packet loss to be less visible. As the physical location of the packet loss is shifted from the top to the bottom of a frame, the visibility of a packet loss decreases.

The significance of different factors in the model can be understood by the increase in the deviance that results if each factor is individually removed from the model. Fig. 7 shows the increase in deviance for each factor, for the RR, NR-P, and NR-B cases. From the figure, we see that FRAMETYPE, SPTXNT (categorical), MOTM, HIGHMOT, and IMSE are very significant factors affecting visibility. Since HIGHMOT depends completely on MOTM, we can attribute its importance also to MOTM. Considered this way, MOTM becomes the most significant factor affecting visibility.

A. Other Models

Before arriving at the final model, we explored different models to find the one with the best performance (smallest deviance). We began with Model 1 using factors TMDR, SPTXNT (categorical), MOTX, MOTY, VARMX, VARMY, RSENGY, IMSE, and HGT. Model 2a drops the four factors related to directional motion, and adds the three factors for overall motion, and it consists of factors TMDR, SPTXNT (categorical), MOTM, MOTA, VARM, RSENGY, IMSE, and HGT. We observed that MOTA is insignificant (95% level). This is not surprising, since intuitively there does not seem to be a reason why motion to the left or to the upper right should make a packet loss more visible. MOTA was dropped (Model 2b) and then HIGHMOT was added (Model 2). Model 3a uses BFRAME instead of TMDR. Our final model (Model 3) uses FRAMETYPE instead of BFRAME.

The improvement in models from the null model (Model 0) to the final model (Model 3) can be summarized by the plot of deviance, shown in Fig. 8, for all three cases (RR, NR-P, and NR-B). There is a huge drop in deviance from the null model to the initial model (Model 1), which is expected. When we improve the treatment of the motion variables and also reduce the model order (Model 2), we see a decrease in deviance indicating a better fit. Also, we see a further decrease in deviance from Model 2 to Model 3 when we treat the time-duration information using a Boolean structure.

IX. GLM FOR CLASSIFICATION

A. Classification Based on Probability of Visibility

Until now, we have used GLM to predict the probability of visibility only. In this subsection, we describe one way to use the GLM model for classifying packet losses, and we analyze the results.

For this study, we classify a packet loss to be visible, invisible, or indeterminate, based on its probability of visibility. We divide the interval [0, 1] into three regions, using the parameter α:

$$\begin{aligned} &[0, 0.5-\alpha] \quad \text{Invisible region} \\ &(0.5-\alpha, 0.5+\alpha) \quad \text{Indeterminate region} \\ &[0.5+\alpha, 1] \quad \text{Visible region.} \end{aligned}$$

The only exception is that when $\alpha = 0$, a probability of 0.5 is considered to be indeterminate and the invisible and visible regions are half open intervals. Our classifier takes as input the extracted parameters, and applies initial (Model 1) and final (Model 3) models. If the resulting probability of visibility does not fall in the indeterminate region, we classify the packet loss to be visible or invisible appropriately.

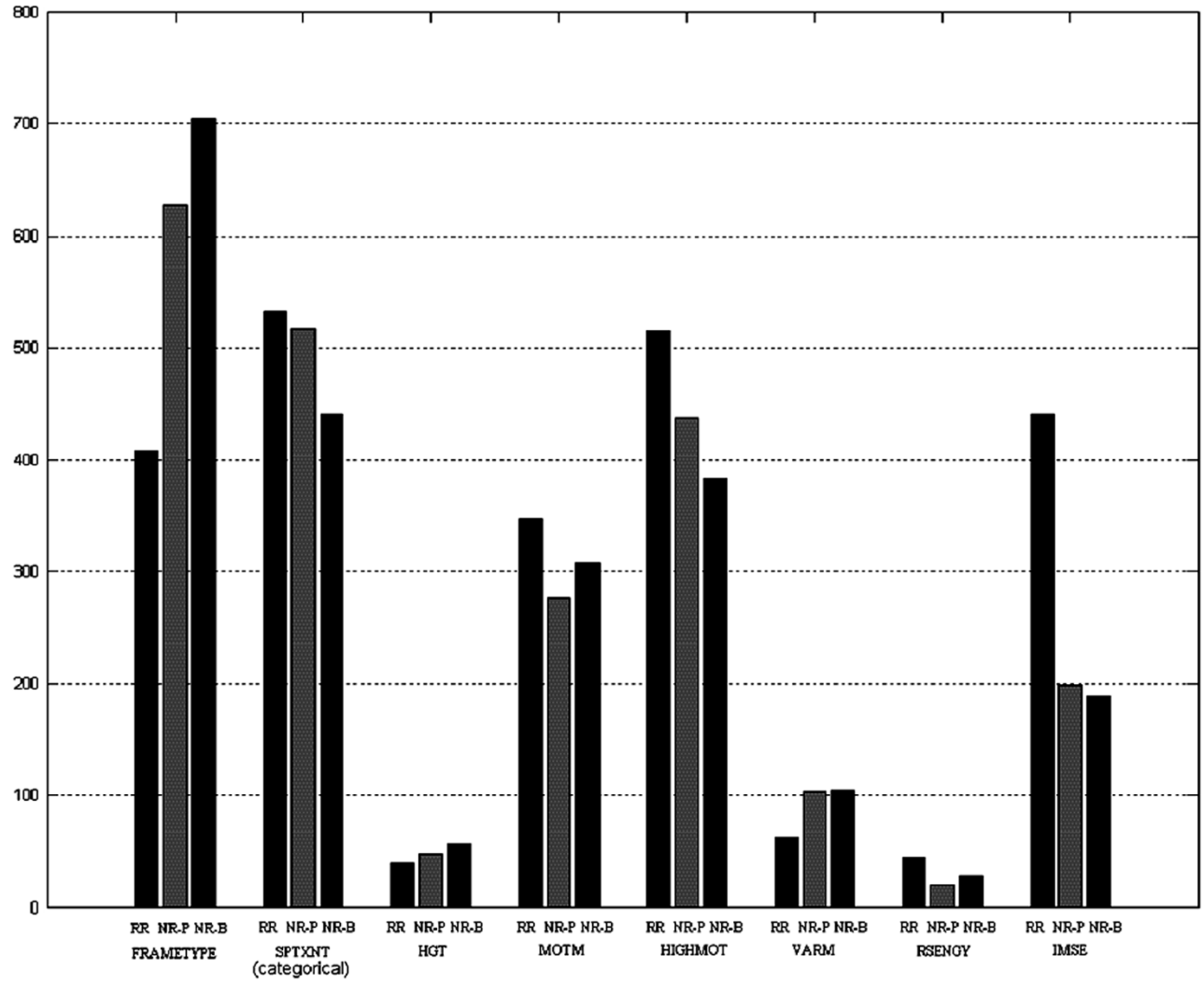

Fig. 7. Factor Significance: Plot showing increase in deviance that results if each factor is individually removed from the model.

To evaluate the accuracy of these models for classification purposes, we compute the ground truth regarding visibility using the results of the subjective test. Further, for the evaluation process, we only consider those packet losses where the ground truth regarding visibility is not indeterminate. Thus, we only consider those cases where both $\tilde{p}$ and $\hat{p}$ do not fall into the indeterminate region. A decision is correct if $\tilde{p}$ and $\hat{p}$ both fall into the visible region or the invisible region. A decision is wrong if $\tilde{p}$ falls in the invisible region and $\hat{p}$ falls in the visible region or vice-versa. Here, we assign zero cost to classifying an invisible/visible packet loss as an indetermnate packet loss, and unit cost for each wrong decision described above.

We vary α from 0 to 0.45 in steps of 0.05 and calculate the accuracy of the model for each value of α. Fig. 9 shows the variation of cross-validation accuracy with α for the initial and final models using the NR-B method. RR and NR-P methods also exhibit similar variation of accuracy with α. The final model is more accurate than the initial model for all three methods.

Fig. 10 compares the accuracy of the RR, NR-P and NR-B methods using the final model for different values of α, and Fig. 11 shows the corresponding number of decisions in each case. Clearly, all three methods (RR, NR-P and NR-B) perform very similarly for different values of α. In particular, our NR-B method performs almost as well as our RR method. As expected, fewer decisions are made as the size of the indeterminate region (2α) increases, but accuracy of classification increases. If we choose a large value of α, we will obtain high accuracy but fewer decisions. On the other hand, a small value of α allows us to make more decisions, but with lower accuracy.

B. Comparison—CART and GLM

We now compare CART and GLM in terms of classification performance. In order to do this and be consistent, we need to compare their classification performance on the same set of packet losses. Using CART, we are able to classify 927 packet losses labeled as visible or invisible based on the ground truth of visibility defined in Section VII. When we apply the classification procedure described in Section IX-A, we do not classify the same set of packet losses as CART does, even when α is equal to 0.25 (CART classifies packet losses when $\tilde{p}$ is not in

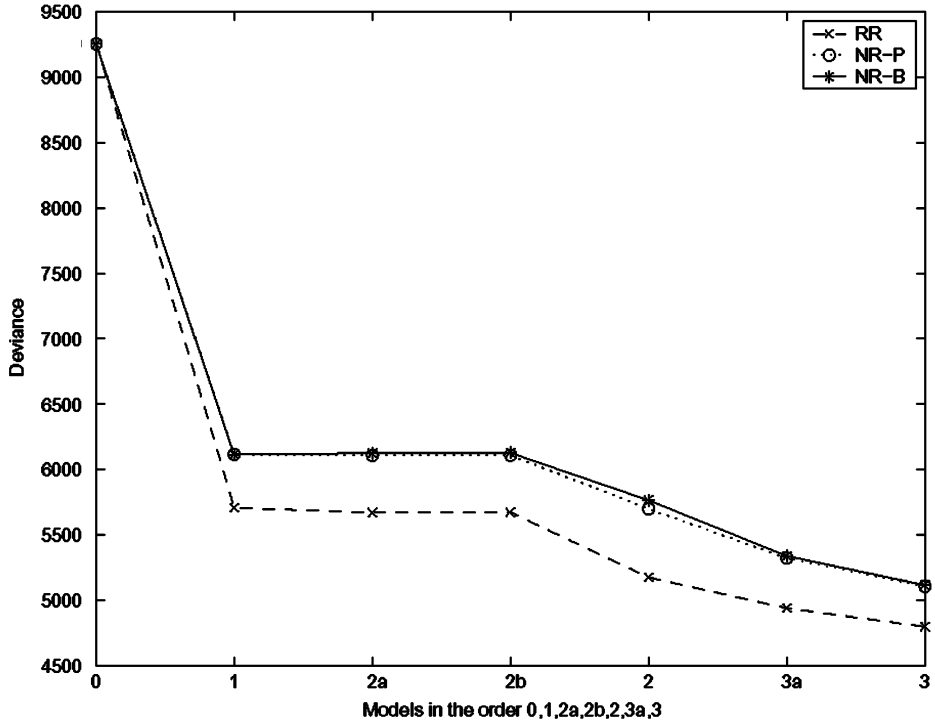

Fig. 8. Plot of Deviance for models considered.

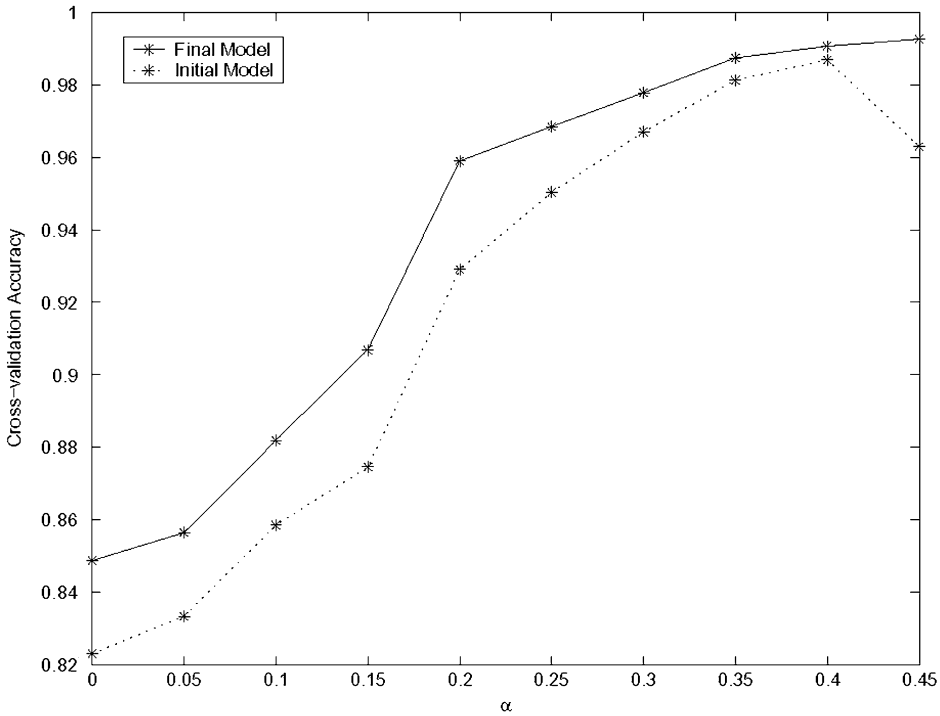

Fig. 9. NR-B: Cross-validation accuracy versus α for GLM classifier.

the indeterminate region, but GLM classifies only when both $\tilde{p}$ and $\hat{p}$ are not in the indeterminate region).

For comparison purposes, we restrict our data set to the 927 visible and invisible packet losses. We use this restricted set with their actual probabilities of visibility, $\tilde{p}$, to train the GLM. When the predicted probability, $\hat{p}$, is greater 0.5, we classify the packet loss as *visible*. Otherwise, we classify the packet loss as *invisible*. Fig. 12 shows a bar plot comparing the cross validation classification accuracy of GLM to that of the two classifiers based on CART.

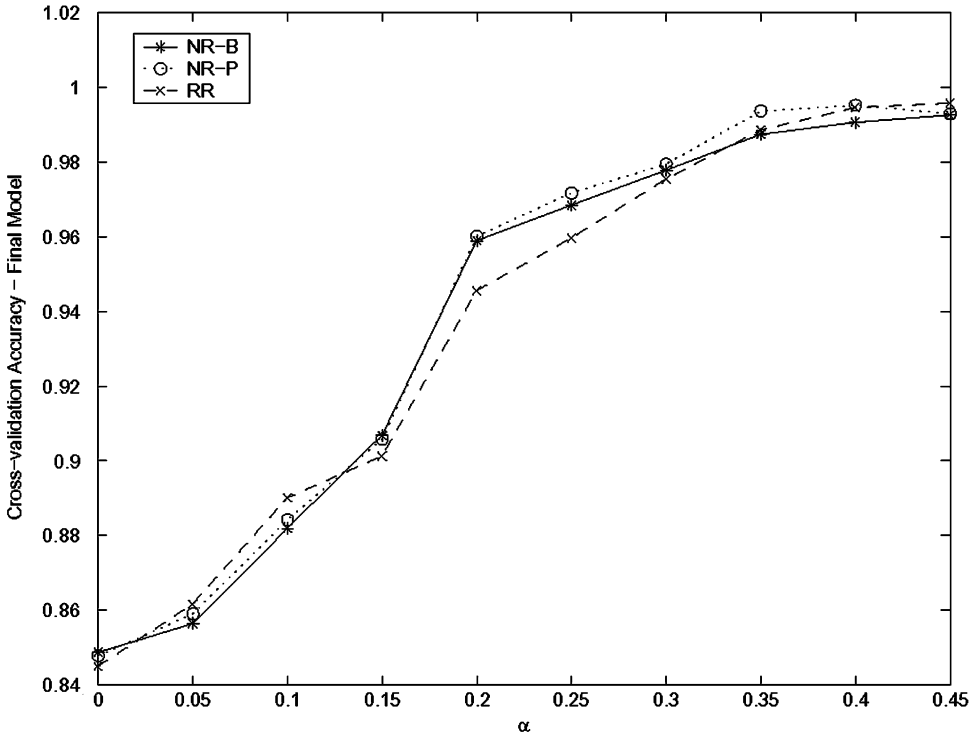

Fig. 10. GLM classifiers: Comparision of RR, NR-P, and NR-B methods.

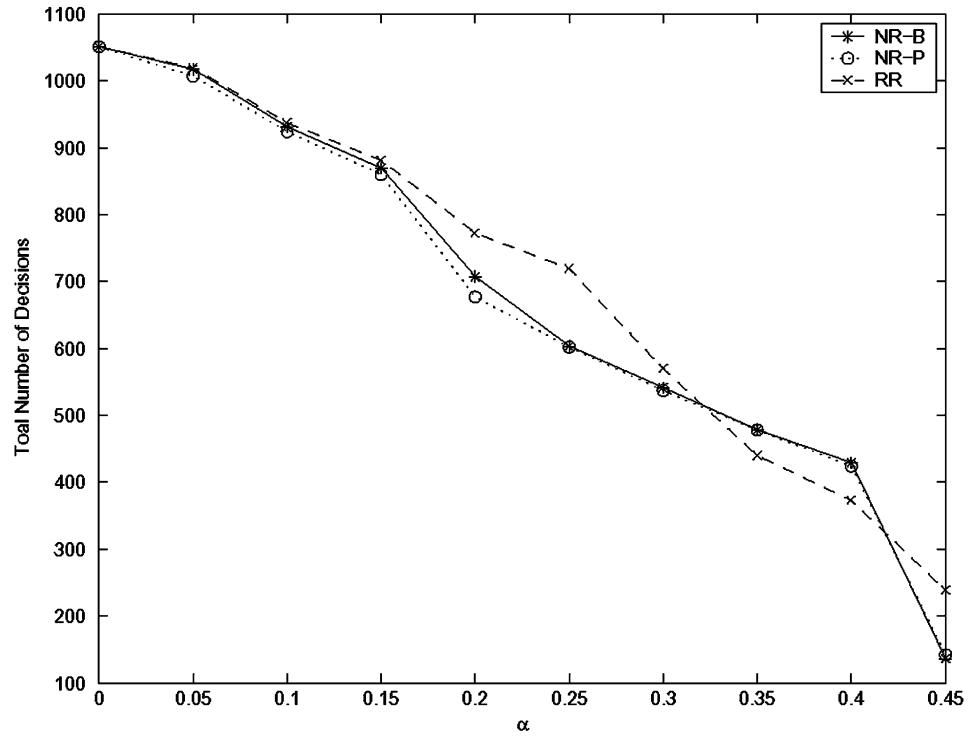

Fig. 11. GLM classifiers: Number of decisions made versus α.

As we can see, the classifiers based on CART outperform the GLM based classifier. From these observations, we can conclude that CART is still a better model for classification purposes, while GLM gives more information in the form of probability of visibility of a packet loss.

X. Conclusion

In this paper, we considered two problems. The first problem is to classify each packet loss as visible or invisible. We used CART to design a tree that solves this problem. The second problem is to assign a probability to each packet loss that it is visible to an

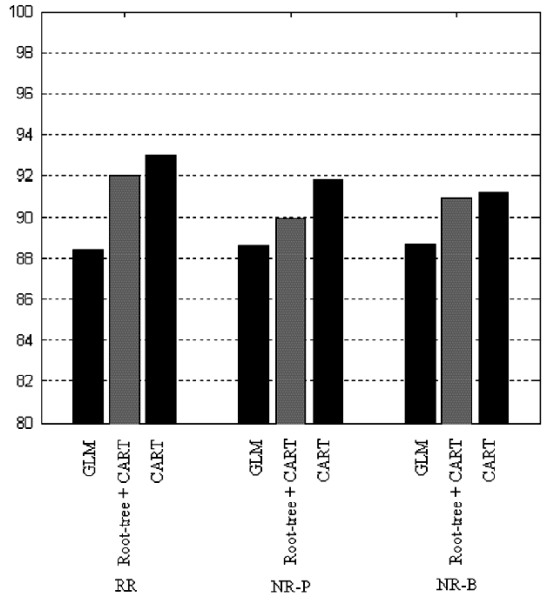

Fig. 12. Comparison: Classification accuracy of GLM and CART based classifiers. For example, the first bar on the left shows that the RR method using GLM achieves a cross-validated correct classification of 88.4%.

average human viewer. We used a logistic regression model to fit the data from subjective tests. We applied these two models when different constraints are active regarding how much data is available to the quality monitor. We considered three cases: RR, NR-P, and NR-B. We examined how to describe pertinent factors such as motion to best predict visibility. As a result, we used total motion instead of x and y directional motion, FRAMETYPE instead of TMDR, and we dropped insignificant factors such as the angle of motion. We described how each factor affects the visibility of a packet loss. We identified the importance of each factor using deviance values. Finally, we used the predicted probabilities to decide whether a packet loss is visible or not, and compared the performance to that of CART.

Our models may be useful in scenarios other than measuring video quality inside the network. For example, we can set thresholds for acceptable quality on visible packet loss rate (VPLR), the rate at which visible packet losses occur, rather than using PLR. Further, our GLM model could be used to prioritize packets within the network based on their probability of visibility, so as to achieve visually optimal streaming.

ACKNOWLEDGMENT

The authors would like to thank Prof. C. Berry at UCSD for his support and guidance with statistical methods and Y. Sermadevi for the software instrumental in conducting the subjective test.

REFERENCES

[1] A. R. Reibman, V. Vaishampayan, and Y. Sermadevi, "Quality monitoring of video over a packet network," *IEEE Trans. Multimedia*, vol. 6, no. 2, pp. 327–334, Apr. 2004.

[2] M. Masry and S. Hemami, "A metric for continuous quality evaluation of compressed video with severe distortions," in *Signal Process.: Image Commun.*, vol. 2, Feb. 2004. Special issue on Video Quality.

[3] P. Gastaldo, S. Rovetta, and R. Zunino, "Objective quality assessment of MPEG-2 video streams by using CBP neural networks," *IEEE Trans. Neural Networks*, vol. 13, pp. 939–947, Jul. 2002.

[4] O. Verscheure, P. Frossard, and M. Hamdi, "Joint impact of MPEG-2 encoding rate and ATM cell losses on video quality," in *Global Telecommunications Conf. (GLOBECOM)*, vol. 1, Nov. 1998, pp. 71–76.

[5] ——, "User-oriented QoS analysis in MPEG-2 video delivery," *J. Real-Time Imag.*, vol. 5, pp. 305–314, Oct. 1999.

[6] J. Lu, M. Chatterjee, M. D. Schwartz, M. K. Ravel, and W. M. Osberger, "Measuring ATM video quality of service using an objective picture quality model," in *Proc. SPIE, Multimedia Systems and Applications II*, vol. 3845, Sept. 1999, pp. 290–297.

[7] A. E. Conway and Y. Zhu, "Applying objective perceptual quality assessment methods in network performance modeling," in *Proc. Eleventh Int Conf. on Computer Communications and Networks*, Oct. 2002, pp. 116–223.

[8] G. W. Cermak, "Videoconferencing Service Quality as a Function of Bandwidth, Latency, and Packet Loss," Verizon Laboratories, T1A1.3/2003-026, May 2003.

[9] B. Chen and J. Francis, "Multimedia performance evaluation," *AT&T Tech. Mem.*, Feb. 2003.

[10] S. Mohamed and G. Rubino, "A study of real-time packet video quality using random neural networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 12, pp. 1071–1083, Dec. 2002.

[11] C. J. Hughes, M. Ghanbari, D. E. Pearson, V. Seferidis, and J. Xiong, "Modeling and subjective assessment of cell discard in ATM video," in *IEEE Trans. Image Process.*, vol. 2, Apr. 1993, pp. 212–222.

[12] S. Winkler and R. Campos, "Video quality evaluation for internet streaming applications," in *Proc. SPIE, Human Vision and Electronic Imaging VIII*, vol. 5007, Jan. 2003, pp. 104–115.

[13] K. Brunnstrom and B. N. Schenkman, "Quality of video affected by packet loss distortion, compared to the predictions of a spatio-temporal model," in *Proc. SPIE, Human Vision and Electronic Imaging VII*, vol. 4662, Jan. 2002, pp. 149–158.

[14] A. Watson and M. A. Sasse, "Measuring perceived quality of speech and video in multimedia conferencing applications," in *ACM Int. Conf. Multimedia*, Apr. 1998, pp. 55–60.

[15] M. S. Moore, J. M. Foley, and S. K. Mitra, "Detectability and annoyance value of MPEG-2 artifacts inserted into uncompressed video sequences," in *Proc. SPIE, Human Vision and Electronic Imaging V*, vol. 3959, San Jose, CA, Jan. 2000, pp. 99–110.

[16] M. S. Moore, S. K. Mitra, and J. M. Foley, "Defect visibility and content importance implications for the design of an objective video fidelity metric," in *Proc. IEEE ICIP*, vol. 3, Jun. 2002, pp. 45–48.

[17] M. G. Ramos and S. S. Hemami, "Suprathreshold wavelet coefficient quantization in complex stimuli: Psychophysical evaluation and analysis," *J. Opt. Soc. Amer. A*, vol. 18, no. 10, pp. 2385–2397, Oct. 2001.

[18] D. Chandler and S. S. Hemami, "Effects of natural images on the detectability of simple and compound wavelet subband quantization distortion," *J. Opt. Soc. Amer.*, vol. 20, no. 7, pp. 1164–1180, July 2003.

[19] Z. Yu, H. R. Wu, S. Winkler, and T. Chen, "Vision-model-based impairment metric to evaluate blocking artifacts in digital video," in *Proc. IEEE*, vol. 90, Jan. 2002, pp. 154–169.

[20] S. Wolf and M. Pinson, *In-Service Performance Metrics for MPEG-2 Video Systems*. Montreux, Switzerland: IAB, Nov. 1998.

[21] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Pacific Grove, CA: Wadsworth, 1984.

[22] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, 2nd ed. London, U.K.: Chapman & Hall.

[23] [Online]. Available: The Website of R Project, http://www.r-project.org/

[24] A. R. Reibman, S. Kanumuri, V. Vaishampayan, and P. C. Cosman, "Visibility of individual packet losses in MPEG-2 video," in *Proc. IEEE ICIP*, Oct. 2004.

[25] S. Kanumuri, P. C. Cosman, and A. R. Reibman, "A generalized linear model for MPEG-2 packet-loss visibility," in *Int Packet Video Workshop*, Dec. 2004.

Sandeep Kanumuri (S'04) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Madras, India, in 2002 and the M.S. degree in electrical and computer engineering from University of California-San Diego (UCSD), La Jolla, in 2004. He is currently pursuing the Ph.D. degree at UCSD.

During the summer of 2003, he was a Visiting Researcher at AT&T Labs, Florham Park, NJ. During the summer of 2005, he worked in the MediaFLO division of Qualcomm, San Diego, CA. His research interests include video compression techniques, video quality estimation and streaming video.

Pamela C. Cosman (S'88–M'93–SM'00) received the B.S. degree (with honors) in electrical engineering from the California Institute of Technology, Pasadena, in 1987, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University in 1989 and 1993, respectively.

She was an NSF Postdoctoral Fellow at Stanford University and a Visiting Professor at the University of Minnesota, Minneapolis, during 1993–1995. In 1995, she joined the faculty of the Department of Electrical and Computer Engineering at the University of California-San Diego, La Jolla, where she is currently a Professor and Co-Director of the Center for Wireless Communications. Her research interests are in the areas of image and video compression and processing.

Dr. Cosman is the recipient of the ECE Departmental Graduate Teaching Award (1996), a Career Award from the National Science Foundation (1996–1999), and a Powell Faculty Fellowship (1997–1998). She was an associate editor of the IEEE COMMUNICATIONS LETTERS(1998–2001), a guest editor of the June 2000 IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (JSAC) special issue on "Error-resilient image and video coding," and was the Technical Program Chair of the 1998 Information Theory Workshop in San Diego. She was an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS (2002–2005). She was a senior editor (2003–2005) and is currently the Editor-in-Chief of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. She is a member of Tau Beta Pi and Sigma Xi.

Amy R. Reibman (SM'01) received the Ph.D. degree in electrical engineering from Duke University, Durham, NC, in 1987.

From 1987 to 1991, she taught electrical engineering at Princeton University, Princeton, NJ. In 1991, she joined AT&T and is currently a Technology Consultant in the Communication Research Sciences Department at AT&T Labs–Research, Florham Park, NJ. Her research interests include video compression systems for transport over packet and wireless networks, and video quality metrics.

Dr. Reibman won the IEEE Communications Society Leonard G. Abraham Prize Paper Award in 1998. She was the Technical co-chair of the IEEE International Conference on Image Processing in 2002.

Vinay A. Vaishampayan (SM'03) received the B.Tech degree from the Indian Institute of Technology, Delhi, in 1981, and the M.S. and Ph.D. degrees from the University of Maryland, College Park, in 1986 and 1989, respectively.

He works on research problems in the areas of communications, signal processing, statistics, and information theory, and has a strong interest in geometric aspects of such problems. He is currently with AT&T Labs–Research, Florham Park, NJ, where he heads the Communications Sciences Research Department. He has also worked in academia, teaching electrical engineering at Texas A&M University, College Station, and in the oil industry for Schlumberger Technical Services.