Principles and Applications of Content-Aware Video Communication

Shih-Fu Chang and Paul Bocheck

Department of Electrical Engineering
Columbia University, New York, U.S.A.

Abstract

Most traditional video communication systems consider videos as low-level bit streams, ignoring the underlying visual content. Content-aware video communication is a new framework that explores the strong correlation between video content, resource (bit rate), and utility (quality). Such a framework facilitates new ways of quality modeling and resource allocation in multimedia communication. We demonstrate advantages of the content-aware approaches in two applications. First, content-aware models were developed for predicting video traffic for live video streams. The video traffic models were evaluated in a dynamic network resource allocation system. Our simulations have shown that, compared to existing techniques, significant reduction (55% to 70%) in required network resources can be achieved. Second, we have used the content-aware principle for automatic generation of utility function (subjective quality vs. bit rate) for live video. Our results indicate that high accuracy in estimating utility functions can be achieved. Such utility functions can be applied to optimal transcoding and media scaling in distributed network environments.

1 Introduction

Pervasive multimedia communication poses significant technical challenges in quality of service guarantee and efficient resource management. Over the past several years, it is recognized that end-to-end quality requirements of pervasive multimedia applications can be reasonably achieved only by integrative study of advanced networking and content processing techniques. Among successful ones are joint source-channel coding [1], adaptive media scaling and resilience coding [2].

However, most existing integration techniques stop at the bit stream level, ignoring a deeper understanding of the media content. For example, video traffic models are constructed to approximate the statistical parameters of the low-level bit rate time series. The encoded video quality is modeled by the relationship between the bit rate and the resulting signal distortion. Yet, the underlying visual content of the video stream contains a vast amount of information that can be used to predict the bit-rate or quality more accurately.

In the content-aware framework, media content is extracted automatically and used to predict video quality under various manipulations (e.g., transcoding) and network resource requirements. We refer the media content to the high-level multimedia features that can analyzed by the machine. Examples include visual features (e.g., motion, complexity, size, and spatio-temporal relationships) of the scenes or objects. These features can be systematically analyzed and are very likely to be present in the future multimedia content representations such as MPEG-4 [3] and MPEG-7 [4].

The content-aware principle can be used for many applications. In dynamic resource allocation (DRA) it can be used for real-time video traffic prediction. Alternatively, it can be used for utility function generation to facilitate the network-wise scalability. It can be used in selecting the optimal transcoding architecture and content filtering in a pervasive computing environment. The media object scalability through the use of utility functions has been included in object description schemes for Universal Multimedia Access (UMA) of MPEG-7 [5].

In this paper, we first articulate the content-aware principle, discuss the practical issues in extracting content features, and then demonstrate the applications in predicting bandwidth requirement and utility functions related to video quality.

2 Content-aware principle

The content-aware framework is based on the recognition of strong correlation among video content, required network resources (bandwidth), and the resulting video quality (utility function). Such correlation between the video content and the traffic has been reported in our prior work [6], in which a conceptual model for content-based traffic prediction has been proposed. In the model, an *activity period* (AP) was defined as an elementary content unit. AP is an video segment that is bounded by abrupt or substantial change in video content. AP's are assumed content-homogeneous and described by a set of content features. In addition, it is assumed that bandwidth requirements of activity periods can be described by relatively simple traffic descriptors.

The detection of activity periods is closely related to the detection of changes in video content. The content of activ-

content feature	category
camera operation	static, panning, zoom
complexity	smooth, medium, cluttered
video object size	small, medium, large
video object speed	low, medium, fast

Table 1: Content features describing activity period and objects.

ity periods is described by *activity period content descriptors* (APC). APC contains a single global scene descriptor G and a set of object descriptors O_I, each one corresponding to particular video object.

$$APC(ap) \triangleq \{G, O_1, O_2, \cdots, O_I\} \tag{1}$$

Examples of global features may be types of camera operations, global motion speed, etc. Object features may include complexity, motion, and size of objects.

The content-aware principle recognizes the importance of using content features in estimating the traffic models. The content features are used as the bridge in linking the activities in visual content to bandwidth requirements. For example, we use the content features to classify the content of each activity period into a set of *activity classes*, and then associate each activity class with representative traffic models. The activity classes are formed based on content features, categorized into limited set of discrete values. This model can be formally described as follows:

Denote $\mathcal{C}$ content-based classifier consisting of g activity classes $\{AC_i \mid i = 1, 2, \cdots, g\}$. The content-based classifier is used to relate activity periods *ap* described by a set of content features $CF(ap)$ into a finite number of activity classes AC.

$$CF(ap) \overset{\mathcal{C}}{\rightarrow} AC_i \Rightarrow TM_i \tag{2}$$

where "⇒" denotes mapping between activity class AC_i and traffic model TM_i.

We verified the content-aware principle by confirming the correlation between content features and video traffic using real MPEG-2 video sequences. In this experiment, we first manually categorized visual features of various activity periods in video and then studied similarities of video traffic among activity periods with similar features. The effect of feature extraction errors in automatic processes will be discussed later in this paper. Table 1 summarizes content features and their categories. At most two video objects (including background) were identified at each activity period.

Figure 1 illustrates the results obtained using the above model. It depicts the difference in bandwidth requirements (D-BIND descriptors [7]) among nine activity classes based on two content features (complexity and motion) categorized into three categories each. *A, B, and C* are primary categories relating to spatial complexity (i.e., smooth, medium, cluttered) and *slow, medium, fast* are secondary categories relating to motion. Each D-BIND descriptor, was computed as an average of D-BIND descriptors of activity periods that were classified into the same category. In essence, the figure reveals the existence of strong correlation between the video content and traffic.

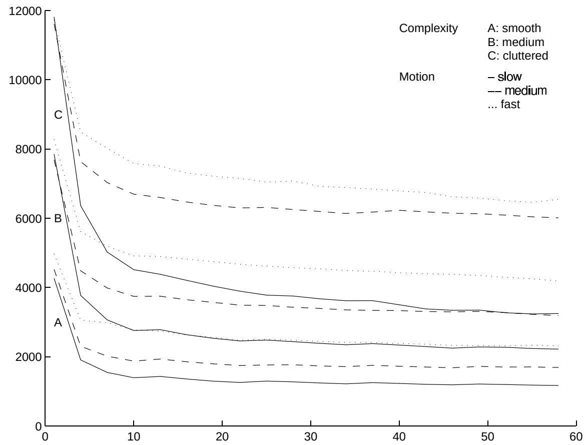

Figure 1: D-BIND traffic resource modeling based on complexity and global motion.

Ideally, different content features result in different traffic models. However, in practice, it is possible that activity periods of different content will result in similar traffic model descriptors. This property is also used later in our automatic content analysis and classification.

3 Issues in video content analysis

Content features are extracted from video streams by the *content analyzer*. The estimation of content features can be done in either uncompressed or compressed domains. Processing in the compressed domain reduces computation because frames do not need to be converted back to the uncompressed (original) domain [8].

We have implemented a real-time content analyzer in the compressed domain that is based on fully automated methods of content analysis [6]. MPEG-2 content analyzer consists of four modules that are invoked sequentially. In the first module, the activity period is detected. In the second module, video objects are detected in each activity period. In the third module, activity period content descriptors are estimated. The last module implements the content-based traffic classifier.

Because our automatic analyzer operates directly in the compressed domain, the decoder was simplified to contain only parts that are necessary for activity period detection, video object detection and content feature estimation. In particular, the computationally intensive DCT function was omitted. This simplification resulted in a real-time performance on a general-purpose workstation. For example, on a SUN SPARCstation 5 it was possible to analyze each video frame for its content in less then 10 ms.

4 Practical issues in bandwidth prediction

One example of use of content-aware framework is for prediction of bandwidth requirements for live video. The *traffic prediction* module is a critical element in adaptive networking systems such as dynamic resource allocation (DRA). It predicts resource requirements for the current activity period. In Section 2, we described a conceptual approach based on two assumptions. First, recognition of distinctive classes of activities can be achieved by content feature analysis. Second, members of each activity class have consistent traffic models. As illustrated in Figure 1, this approach is effective when subjective content analysis is available.

However, in practical applications, content features are extracted by automatic processes, which may cause some errors. In addition, videos of different activity types may produce similar traffic traces. The goal is to predict the traffic based on automatically extracted features. Distinction of activity types is not needed explicitly. Therefore, in automated system, each activity period is directly mapped to a traffic class, without categorizing its activity type (like in Eq. 2), in the following way.

Denote $\mathcal{A}$ a content-based classifier. Assume that each activity period can be classified into one of I *traffic classes* $TC = \{TC_i \mid i = 1, 2, \cdots, I\}$. Denote CHT_i *characteristics traffic descriptor* describing the traffic model that is associated with each traffic class TC_i. Then, traffic model corresponding to each activity period can be obtained as follows:

$$APC(ap) \xrightarrow{\mathcal{A}} TC_i \Rightarrow CHT_i \tag{3}$$

where "$\Rightarrow$" denotes mapping between TC_i and CHT_i.

In the actual application, the content-based classifier operates in two modes: training and selection. In the training mode, (i) traffic classes, TC, are determined, (ii) content-based classifier parameters are learned and (iii) characteristics traffic descriptors, CHT, are estimated. In the selection mode, content classifier is used to predict resource requirements for each activity period.

In our experiments, publicly available machine learning tools were used to simulate the above-described content-based classifier [9]. The time spent for clustering operation is not critical, because it is performed during off-line training or on-line adaptation and does not affect real-time performance.

Accuracy of the content-based classifier is important. Misclassification occurs when an activity period is not mapped to the traffic class with the most accurate traffic descriptor. We have simulated generation of parameters of the content-based classifier using one half of activity periods identified in the video and verified its accuracy on the remaining subset. In our experiments, a high classification accuracy of 86.14 % was achieved.

Performance study of content-based DRA was based on trace-driven simulations. A trace-driven simulator was developed for that purpose. Results were obtained using a single 54000-frame-long trace (30 minutes) of an MPEG-2 encoded movie.

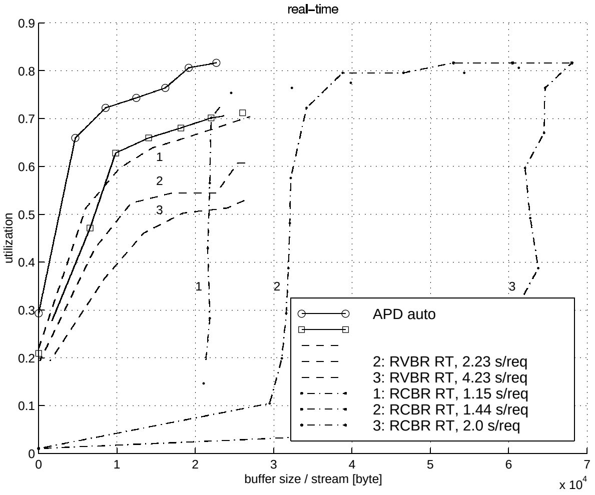

Figure 2: Performance comparison of content-based with existing DRA algorithms.

Figure 2 depicts the main result. The first curve, *APD auto*, represents a performance upper bound which uses automatic activity period detection but also uses precise traffic descriptors measured offline. The second curve (auto CA/C RT) uses the automatic activity period detection as well as the traffic descriptors obtained by automatic content mapping. Network simulations revealed that both the content-based approaches achieve better performance (in terms of link utilization) than other existing schemes (*RVBR RT*). The link utilization achieved by *RVBR RT* was substantially less than utilization achieved using the *APD auto* scheme (about 55% - 70% difference). In addition, *auto CA/C RT* achieved better utilization than existing schemes except for a special case.

The superior performance of the content-based dynamic resource allocation, based on the content-aware framework, can be attributed to its two distinguished features. First, to the fact that it is able to track changes in visual content (and therefore changes in bit-rate). This is accomplished mainly by detecting discontinuities in visual content. Second, content-aware models facilitate the use of effective content classification methods which improve resources prediction accuracy. This is in contrast to traditional prediction schemes that use only bit rate and network buffer occupancy in their heuristics segmentation and resource prediction algorithms.

5 Utility function estimation

Utility functions represent a powerful framework for characterizing the ability of applications to adapt to varying network conditions. Specifically, in the context of bandwidth allocation, utility functions indicate a media object's quality

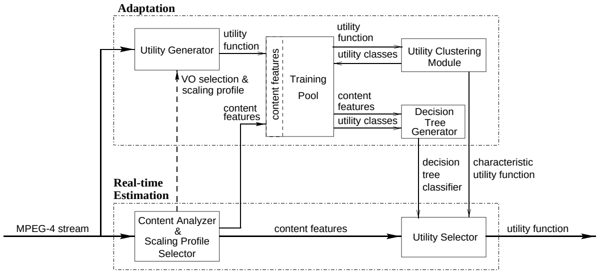

Figure 3: Content-based utility function estimator.

as a function of available bandwidth.

In practice, estimation of utility functions requires repetitive computation of quality metrics with different encoder parameters such as quantization or dynamic rate shaping (DRS) parameters. Process of repetitive estimation is main cause of large amount of calculations required for utility function evaluation. To the best of our knowledge, a system that allows efficient estimation of subjective utility function in real time does not exist. We demonstrated a new system for speedup of generation of utility function based on content-based classification technique that allows estimation of utility in real-time [9].

The acceleration technique does not explicitly compute utility functions for each video object. Rather, the content-aware principle is applied and machine learning techniques are used. The system uses video content, represented by a limited set of content features, to determine the *utility class* of an object. Because video content can be dynamically extracted from compressed video streams, this technique is suitable for real-time applications.

Figure 3 illustrates the architecture of a content-based utility function estimator that can support a variety of compression schemes (e.g., MPEG-1, MPEG-2, MPEG-4, H.263, etc.) The system architecture comprises two main components: a real-time estimation module and an adaptation module.

The system allows smooth adaptations during continuous operation and over various video content types. The adaptation module, which is computationally intensive, is decoupled from the real-time estimation module. Both modules operate asynchronously. During the normal operation, based on content features extracted online by the content analyzer, the utility selector dynamically determines utility class and corresponding *characteristic utility function* for each video object. The characteristic utility function is used as an estimator of real utility function that is not explicitly computed for each video object. An adaptation module is activated to periodically re-compute the decision tree parameters used to initialize the utility selector. In this manner, by avoiding explicit per-object generation of utility functions the system facilitates operation in real-time.

Accuracy of the MPEG-2 and MPEG-4 content-based utility estimator was evaluated in [9]. For the experiment using MPEG-2 video (17 utility classes), the classification accuracy of the whole set of utility functions was 91%; that is to say, among the total 734 video frames, 91% of them were classified correctly using content features. Similarly high classification accuracy of 80% - 85% was achieved using MPEG-4 traces.

6 Conclusion

In this paper, we presented a new content-aware framework for video communication and presented two applications. First, content-based bandwidth prediction was used for dynamic resource allocation. Second, the content-aware framework was used for real-time generation of the utility function. The importance of the content-aware framework is that it can be effectively used for media scaling in future low bandwidth and wireless networks. Our results indicated the feasibility and relatively high accuracy of such systems.

References

[1] A. Ortega and M. Khansari, "Rate Control for Video Coding over Variable Bit Rate Channels with Applications to Wireless Transmission", *Proceedings IEEE ICIP*, October 1995.

[2] G. de los Reyes, A. R. Reibman, J. C.-I. Chuang, and S.-F. Chang, "Video Transcoding for Resilience in Wireless Channels", *IEEE International Conference on Image Processing (ICIP'98)*, October 1998.

[3] ISO/IEC 14496-2 CD, "MPEG-4 Visual", October 1997.

[4] ISO/IEC JTC1/SC29/WG11, "MPEG-7 Requirements Document, Coding of Moving Pictures and Audio", Vancouver, July 1999.

[5] J. R. Smith, Ch.-S. Li, A. Puri, Ch. Christopoulos, A. B. Benitez, P. Bocheck, and S.-F. Chang "Content Description for Universal Multimedia Access", *International Organisation for Standardisation ISO/IEC JTC1/SC29/WG11 MPEG99*, Vancouver, BC, July 1999.

[6] P. Bocheck and S.-F. Chang "Content-based Video Traffic Modeling and its Application to Dynamic Resource Allocation", *Submitted to ACM/IEEE Transactions on Networking*, 1999.

[7] E. W. Knightly and H. Zhang, "D-BIND: An Accurate Traffic Model for Providing QoS Guarantees to VBR Traffic", *IEEE/ACM Transactions on Networking*, Vol. 5, No. 2, April 1997, pp. 219-231.

[8] J. Meng and S.-F. Chang, "Tools for Compressed-Domain Video Indexing and Editing", *Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Database*, Vol. 2670, San Jose, February 1996.

[9] R. Liao, P. Bocheck, A. Campbell, and S.-F. Chang "Content-aware Network Adaptation for MPEG-4", *Proceedings of NOSSDAV'99*, June 1999.