An Engine for Content-Aware On-Line Video Adaptation*

Luis Herranz, Fabricio Tiburzi, and Jesús Bescós

Grupo de Tratamiento de Imágenes, Escuela Politécnica Superior
Universidad Autónoma de Madrid, E-28049 Madrid, Spain
{Luis.Herranz, Fabricio.Tiburzi, J.Bescos}@uam.es

Abstract. In this paper we show a practical implementation of an adaptation engine based on content inspection of video material. The localization of domain independent hints or features that allow to infer the non homogeneous distribution of semantically relevant information, allows to dramatically reduce the amount of adapted data while maintaining the meaningful information. The extraction of these features is performed on-line, via techniques that operate on the compressed domain, following an abstraction model that allows transparent adaptation of DCT based video and wavelets based scalable video.

Keywords: video summarization, slideshows, storyboards, video analysis.

1 Introduction

Increasing availability of low cost and high bandwidth connectivity, along with the amazing proliferation of devices with video recording and display capabilities and the subsequent deployment of services in this new scenario are some of the factors that have recently empowered the development of a set of technologies for video content processing and management.

A main topic in this area is the adaptation of audiovisual media to different terminals, networks and users. In this sense, a first challenge is to achieve adaptation (e.g., bitrate reduction, frame size modification, colour conversion) while maintaining most of the desired information. A second one is to do this in an efficient way.

Regarding to the first challenge, traditional content-blind adaptation is performed via transcoding: the original media is decoded and then encoded according to the target scenario. Some more efficient solutions, like transcoding without fully decoding[1] or simply extracting bitstream data for scalable formats, via Bitstream Syntax Description (BSD)[2], are similar from this point of view.

Video coders are currently focused on preserving perceptually relevant information, but do not almost consider semantic relevance; hence, the coding process is usually homogeneous regarding to this concept. However, relevant semantic

* Work partially supported by the European Commission under the 6th Framework Program (FP6-001765 - aceMedia Project) and by the Spanish Government under Project TIN2004-07860-C02-01.
information is highly concentrated as spatial or temporal events (i.e., when something happens) and shows very low variation elsewhere (i.e., in static scenes, under global and slow motion, or under fast tracking of a relatively static object): it is highly inhomogeneous. Therefore, any adaptation operation that considers semantic features can help reducing the final bitstream size without a significant content loss.

Semantically relevant information is usually highly dependent on the specific domain, ranging from the presence or absence of objects (e.g., people, faces, race cars) to more complex actions (e.g., left objects, people running, overtakes or accidents), which makes it difficult to follow a generic approach. Nevertheless, some meaningful situations can be considered quite domain independent (e.g., shot changes, changes in the camera motion scheme, moving objects relative to the background, etc.), and are hence specially adequate for content-aware domain-independent media adaptation. One of the objectives of the work presented is to extract these semantically relevant features.

Regarding to the second challenge, if we aim to carry out adaptation both over stored and on-line video content, the identification of relevant features has to be on-line performed for every adaptation operation. This avoids the requirement to store semantic hints which may be quite bulky (e.g., segmentation masks), and the need to perform an exhaustive feature extraction which for many adaptation requests would be useless. In order to achieve on-line operation, feature extraction in the presented engine is performed applying techniques that work on the video compressed domain. Some advantages are the direct availability of estimators for features that are hard to extract at a pixel level (e.g., the motion field), and a dramatic reduction of the dimensionality (e.g., by working over DC images); the main drawback is that these techniques are highly dependent on the coding standard and on the coding parameters. Here we present an abstraction model aimed at obtaining features mostly independent of the specific coding domain.

The general objective of this paper is to show a practical implementation of an adaptation engine based on content inspection of coded video material. Section 2 shows a general diagram of the adaptation engine, identifying three modules: Content Analysis, Media Generation, and Adaptation Control. Section 3 presents the analysis module, which integrates the feature extraction techniques into the abstraction model that allows to transparently operate over two notably different domains (DCT based video and wavelets based video). Section 4 introduces the approach followed in the content generation module and the problematic associated to the control of the adaptation process. Finally, Section 5 shows some experimental tests.

2 Overview

This section presents the engine architecture (see Fig. 1) and some design considerations proposed to perform content-driven adaptation of video material. The main difference between this approach and other reported architectures[3][4] for content-aware video adaptation is, first, that we focus on on-line extraction of the semantic features that control the process; and second, that the control loop used to manage the output bit-rate is based in our case on the variation of the number of included relevant frames.