Chapter 5.9
Integrating Multimedia Cues in E-Learning Documents for Enhanced Learning

Ankush Mittal
Indian Institute of Technology, India

Krishnan V. Pagalthivarthi
Indian Institute of Technology, India

Edward Altman
Institute for Infocomm Research, Singapore

INTRODUCTION

The digitization of educational content is radically transforming the learning environment of the student. A single lecture, as well as supporting reference material, textbook chapters, simulations, and threaded chat system archives, may be captured on one hour of video, a set of 20 or more slides, and ancillary text resources. A single course may contain 25 such lectures and a single department may have 30 distinct courses that have been digitized in a single year. If, while taking one course, a student wants to find a relevant definition, example, discussion, or illustration of a concept, then the student would potentially need to search as much as 750 hours of video, 15,000 slides, and a huge volume of text in order to find the desired information. Thus the online student is overwhelmed by a flood of multimedia data which inhibits the development of insight.

Insight is a key ingredient of education and is most often achieved by the manipulation of information through the discovery of new relationships, identification of hidden structures, or the construction of domain models. The methodology of instructional design may be used to anticipate the needs of the student in the controlled environment of a classroom with novice learners, but it is inadequate for the needs of a heterogeneous population of online learners. Intelligent tutoring systems provide an additional degree of flexibility for the independent learner, but are

difficult to produce and maintain. The advanced learner needs powerful search and organizational tools to support self-guided learning. In all three cases there is a need for content-based retrieval of multimedia resources ranging from simple indexing and navigation of lectures to ontology-based mining of information nuggets from large repositories of heterogeneous content.

The needs of the independent learner are particularly demanding due to the requirements for real-time, context-dependant, and precise retrieval of unstructured and incomplete information distributed across multiple media sources. Due to the large size of the corpus and the highly focused nature of the target information, the systematic labeling (either manually or automatically) of the media is not feasible. Instead, we propose a system for *media hot spotting*.

Media hot spotting is the process of finding *hot spots* within the text, audio, video, and other media content. A hot spot may be a distribution of key terms in a text document; a matching distribution of terms in the speech track of a lecture video; or a sequence of writing, emphasis, and gesture events in the video stream of a lecture, where an event is a spatial or temporal set of interrelated features. Individually, each piece of information does not convey sufficient semantic information to identify the informational content of the media. In combination, they provide significant evidence, for instance, that a Definition event has occurred in the media and that a particular term has most likely been defined within the context of this event. Thus, given a model for how information co-occurs across different media objects, hot spotting enables the rapid retrieval of candidate media content for further analysis and reference.

Content-based solutions are available for domains like sports and news, but have not yet been systematically explored for educational videos (Idris & Panchanathan, 1997; Mittal & Cheong, 2003; Woudstra et al., 1998). This chapter describes a new method of content-based retrieval for e-learning videos using camera motion cues, audio features, slide layout, and associated decision rules. Using this technique, we are able to separate the lecture videos into several component states and personalize the video from these states. For our experiments, we used 26 lecture videos from the Singapore-MIT Alliance, along with the associated PowerPoint slides.

This chapter is organized as follows. The next section presents a discussion of an existing distance learning program (SMA) and other related work. Then an overview of our approach is presented, followed by the section titled "Multimedia Indexing Features," where we show the framework for modeling multimedia information and present a list of the most useful features used in the video segmentation task. The section titled "Indexing of Lectures" elaborates upon the mapping of low-level features to lecture semantics. Finally, we discuss the experimental results and significance of this approach.

DISTANCE LEARNING PARADIGM

Singapore: MIT Alliance Educational Setup

The work presented here relates to the materials used in the Singapore-MIT Alliance (SMA)[1] development program. SMA is an innovative engineering education and research collaboration among the National University of Singapore (NUS), Nanyang Technological University (NTU), and the Massachusetts Institute of Technology (MIT). SMA classes are held in specially equipped classrooms at the Singaporean institutes and at MIT using live video transmission over the Internet. The synchronous transmission allows participants at both locations to see each other and speak normally. However, because of the 12-hour time zone difference, SMA has made a great effort to find and develop tools to enhance asynchronous learning.