MIRACLE: Multimedia Information Retrieval by Analysing Content and Learning from Examples

Zhibin Lei  S. Kicha Ganapathy  Robert J. Safranek
Bell Laboratories, Lucent Technologies, 700 Mountain Ave., NJ 07090
zbl@research.bell-labs.com

Abstract

We present a novel application for Content Based Retrieval Techniques in Electronic Commerce using Interactive Multimedia, in which visual queries play a central role. We address some of the technical issues in representing and analysing image primitive features that are the building blocks of the MIRACLE system. These can be generalized into a much broader range of applications as well. We believe that Electronic Commerce is a very promising area with much potential for growth and that it covers a spectrum of rich research topics.

1. Introduction

Multimedia Information Systems are experiencing a tremendous growth as a direct consequence of the popularity and pervasive use of world wide web. As a consequence, it is becoming increasingly important to provide efficient and flexible solutions for accessing and retrieving multimedia data. Images and video are emerging as significant data types in multimedia systems. And yet, most commercial systems are still text and key-word based and do not fully exploit the image content of these systems.

There are certain items such as plumbing supplies, where the user can seldom identify the part by name but can sketch it or can show an image of it. The presence of a digital camera presents even more interesting opportunities. For instance, if one wanted to coordinate a color scheme for matching carpet textures with sofa colors/texture, there is, as yet, no convenient vocabulary to describe verbally the choice of patterns. These are cases where "Visual Similarities" may be the only way to do it. This retrieval technique combined with graphical techniques for visualization can be the difference to induce a buyer to either buy online or at the very least make a decision to buy from that store. Clearly we also need technologies to search different databases of various retailers. They may or may not use the same text to describe some patterns. This is another reason to use an image to illustrate what one needs as opposed to text.

If one can develop better interface technologies for enhanced presentation and allow multimedia queries of the kind involved in an actual store, then it is possible that the increased comfort level and the helpful interactions will prompt more consumers to close transaction online. Thus the two thrusts of our research are (1) Multimedia queries that analyse the content of the merchants’ databases to make more informed decisions and (2) Multimodal Interfaces to interact with the database to provide ease of use.

2. Architecture

The system under development allows the user to query using a scanned image, hand drawn sketch, or a live image or any combination of these. Additionally, the system uses a speech activated browser as front end so that the user can compose a compound multimedia query that combines speech, images and text. Additionally the system produces speech output using a very sophisticated text-to-speech synthesizer. The system uses speech and images as input and uses speech and images (graphics and video) as output. Our goal throughout has been to build a true multimedia system that uses multimedia queries for examining content and multimodal interfaces for interaction. The overall architecture of the system and the research modules is best illustrated using Figure 1. Since our system is geared to an individual user, we expect machine learning using relevance feedback to develop user profiles. We are also developing techniques to increase user satisfaction by using edit distances to better align system recommendations with user choices. It is our belief that machine learning techniques will be just as important, if not more, than feature based retrieval techniques that merely analyse content.

There are many components involved in the design and implementation of such a system. Several key steps involved in a typical interactive multimedia content based manipulation and retrieval application:
1. Acquire an image (sketch, scan, camera, database)
2. Select object manually/automatically (segmentation, knowledge of spatial layouts)
3. Compose Query (interactive editing, speech, dialog)
4. Process query and retrieve matches (content based retrieval)
5. Interaction/User experience (relevance feedback, learning)
6. Synthesize new environment (object based multi-layered rendering)
7. Loop through to refine search

Figure 2 shows a typical interior design example. It involves many ongoing research topics. The first row is an image of study room with several pieces of furniture: armchair, carpet, small table and bookshelf. Second row shows the results of synthetically composed scene with the furniture chosen by the user as shown in the third row: carpet, plant and window and curtain. Visual query engine retrieves database items which match user’s specific query samples. Notice that both 2D and 3D object modeling (shape, texture, compose etc.) are needed to handle the complex situation.

3. Research Technologies

Given below is a list of the research technologies that are needed to translate this vision into reality for all applications: visual primitives, color normalization, machine learning, multimodal interfaces, multimedia queries, dialogue/query design, database issues, model acquisition, image synthesis, AI techniques, networking issues, etc..

One goal of our research is to develop an architecture for Electronic Commerce wherein all these technologies can be developed such that for any given application within the domain of E-Commerce one can pick and choose the modules that are needed for that application.

We believe that Electronic Retailing (E-tailing) is a very promising area with much potential for growth and that it covers a spectrum of rich research topics. We are focusing our efforts on user interfaces, interactive techniques for retrieval and machine learning for personalization and customization. Efforts are underway to transfer this software on to a laptop for a truly portable multimedia E-tailing system.