Advanced media-oriented systems research: ubiquitous capture, interpretation, and access

1   Executive Summary

Using two large-scale research applications – a distributed education repository, and perceptual computational spaces for multimedia-based collaboration – as drivers, we propose to carry out extensive systems research and integration to support ubiquitous access, capture, and interpretation of a variety of multimedia streams. The PIs on this proposal are actively involved in research relating to the two applications, and the various aspects of systems research pertaining to these kinds of applications including operating systems and architecture issues in parallel and distributed systems and storage systems.

As we look to the future it is becoming clear that there will be an explosion in the need for interactive collaborative services that enable end users to manipulate large numbers of multimedia streams from a variety of devices that serve as access and capture points. There are several technology factors that back this assertion. First, as high bandwidth network connectivity becomes cheap, it will become feasible to communicate high quality voice and video in real-time both at home, the workplace, and public spaces. Second, the increasing use of multimedia input/output devices for interacting with the computer and the decreasing cost of mass storage enables the easy archival and sharing of personal, family, and professional information (such as photo albums, family videos, and clinical records). Third, computing devices are becoming ubiquitous in our everyday life and serve as a source of information capture and display that could be used in a variety of constructive situations such as educational, social, and professional interactions.

This situation is likely to spur the growth of several future computing scenarios such as:

- networked intelligent kiosks for public access to information and entertainment,
- advanced education techniques (distance learning, distributed classrooms, distributed student discussion groups, virtual field trips, and training in simulators),
- information rich environments such as an aware home that know its inhabitants and their activities,
- customized production (creation, editing, and composition of video, audio, media-mail, games, etc.) and distribution of home entertainment, and network television programs,
- enterprise-wide information exchange (video conferencing and on-demand presentation composition), and
- electronic commerce.

There are several service requirements posed by such emerging applications: capturing many simultaneous media streams, analyzing and interpreting such streams (perhaps on the fly and/or perhaps off-line) to recognize the user intent, inter-relating a stream or the data generated as a result of the analysis (both semantically and temporally) to other data (including other concurrent independent streams), and providing ubiquitous access to the streams themselves or to other information related to these streams.

One can enumerate a few application level research issues that arise in the context of such futuristic application scenarios. First, techniques and algorithms are needed to enable the perceptual processing of interactive streams coming from a variety of media sources such as cameras, microphones, and other sensors. Second, algorithmic work is required to interpret and inter-relate such streams. Third, techniques have to be invented for accessing the captured content with inexact information.

There are a number of research challenges on the system side as well. An important point to note is that such applications have some significant differences from scientific applications that have been the focus of high performance computing research for some time now. A fundamental difference is in the level of dynamism in these emerging applications due to their interactive nature. Thus, the computational requirements of such an application are directly related to the extent and sophistication of the interaction in the application at any given time. Further, since timeliness is important in interactive applications, programming such applications to meet the inherent synchronization and communication demands is non-trivial. Therefore, there is a need for inventing data abstractions and associated runtime mechanisms to enable the development of such applications, especially on parallel and distributed platforms. Another point to note is that there may be considerable diversity in the computational capabilities of the devices that act as capture and access points (e.g. laptops to immersive desks), and their connectivity to other devices as well as to the repositories that

serve as content stores. Not all devices and users may require the same level of service all the time. The implication of the last two points is that the computational, network, and storage resources can be shared among a number of such access and capture points. However, these resources need to be managed to ensure adequate levels of service for such applications. It may be necessary to provide continuous availability in many emerging application scenarios (for e.g. electronic commerce), so that partial failures of the resources result only in reduced levels of service.

In this proposal, we use two driver applications to give a concrete context to the application- and system-level research to be carried out. The first is a university-wide education repository called *Intelligence Quotient* (IQ), a logical extension of *Classroom 2000*, a home-grown system for capturing in-class interactions (such as audio and video recordings, class notes, and prepared slides) for later use by students and instructors. IQ is intended to evolve Classroom 2000 in interesting new dimensions, essentially allowing students and instructors to interact without necessitating their physical co-location either in space and/or time. It will have support for instructor-less classrooms, virtual classrooms, and for linking course material to other on-line resources (such as previous offerings of the same course, and other concurrent related course offerings). The second application is a more exploratory one called *computational spaces*, and may be envisioned as networked intelligent *perceptual kiosks* that enable distributed multimedia collaboration. Such kiosks may range from simple laptops with wireless connectivity, to immersive desks and caves with gigabit connectivity, to specialized video walls with built in cameras and microphones. A kiosk may be located in public places such as campuses, stores, museums, and airports, and serves as an access and capture point to interact with multiple people and backend systems (such as Web servers and education repositories) for education, entertainment, and information.

Both these applications are fertile grounds for spurring a number of application- and system-level research explorations. The requested research infrastructure will serve as an experimental platform for validating, and quantifying the performance implications of the ideas we generate through these explorations. At the application-level we will explore techniques and algorithms for perceptual and aware computing. The basic idea is to deploy sensors (such as cameras, and microphones) to capture the data from the environment, and analyze the captured data (using a variety of computer vision and speech recognition techniques) to glean lower level information (e.g. human presence or mood of the person). We will also develop tools and techniques for higher level processing based on inference that can be drawn from inter-relating concurrent media streams to gauge the intent of the user from these captured streams. Further, we will also explore database techniques for supporting fuzzy queries in these kinds of applications.

At the system level, we will develop a quality of service (QoS) architecture for managing the distributed resources consisting of the computational and storage clusters, and the diverse access and capture points in the application. The intent is to identify key application parameters that can influence resource management decisions to suit the dynamic needs of the application. Techniques will be developed to dynamically adapt the use of the resources (such as computation, network, and storage) to match the instantaneous application needs. For example, the level of consistency provided to shared state in an application may be influenced by such adaptive resource management techniques. We will also explore caching and replication of shared ephemeral state in the applications to increase both performance and availability. To facilitate cross-correlating information streams at both the semantic and temporal levels we will develop APIs that automatically and semi-automatically create and manage the metadata. The metadata is expected to facilitate the application-level need for satisfying information access needs based on fuzzy queries. We will also explore the storage architecture for organizing both the metadata and the content created by such stream-oriented applications. In particular, we will investigate prefetching techniques and replication at the storage level to provide adequate response for access to continuous media data. A number of middleware and run-time issues to aid in the development of complex cluster-based applications are also to be investigated. In particular, we will explore the data abstractions and programming idioms that enable the development of collaborative multimedia applications on parallel and distributed platforms. We will investigate parallelism management within a cluster to take advantage of the inherent data and task parallelism in dynamic applications. We will also investigate hardware enhancements that allow computations to be migrated to the network interfaces thus reducing the performance penalty for stream-oriented applications on standard CPU boxes due to memory hierarchy effects.

We will use the requested research infrastructure to build testbeds of interactive spaces, including classrooms and distributed networked kiosks, equipped with cameras, microphones, and displays. We are already in the process of building another testbed called "aware home", with some experimental funding from the *Broadband Telecom Center* (BTC) of Georgia Tech. This home is being equipped with a variety of capture devices to allow multi-way interactions among the occupants of the home. These testbeds will provide us with the ability to explore automatic interpretation and resynthesis of people's actions in environments of different types and would give us an unprecedented opportunity to evaluate the impact on real users. As part of the proposed work, we will carry out a systematic evaluation of the design alternatives for the various subsystems and their performance impact on the overall system and the applications.

The broader research goal of this proposal is to develop tools and techniques for constructing large-scale network-based collaborative multimedia applications that allow the manipulation of both stored content as well as dynamically generated content. Specifically, the expected contributions of the proposed effort are summarized below:

- tools for capture, integration, access, and interpretation to enable multimedia collaboration,
- tools and techniques for adaptive resource management based on a QoS architecture,
- a distributed clustered storage architecture spanning multiple levels,
- programming abstractions (and efficient cluster implementations thereof) for developing parallel and distributed interactive multimedia applications,
- novel caching and replication protocols for distributed shared object state of collaborative applications, and
- a detailed evaluation of the system to quantify the effects of the design alternatives.

We have current research efforts already underway towards achieving some of these goals. The proposed research will help to focus these efforts toward the larger vision presented above.

The requested infrastructure falls into the following major categories:

- computational clusters and high-bandwidth low-latency networking,
- mass storage (both secondary and tertiary),
- immersive desks, non-linear editing stations, stationary intelligent kiosks, and mobile capture and access stations,
- video wall as a display and interaction device for a community of users.

The research infrastructure identified in this proposal is intended to directly facilitate the proposed research. In addition, it is expected that the equipment will enable several other research projects that are not featured in this proposal. For example, Professor Richard Fujimoto's research on *High-Level Architecture* (HLA) for distributed simulation and distributed virtual environments, Professor Jessica Hodgins's research on interactive animation of human-body motion, and Professors Ellen Zegura's and Mostafa Ammar's joint projects on scalable video multicast and replicated server selection will benefit from the research infrastructure being requested in this proposal.