A Wireless Client for Accessing Multimodal Digital Video Library Systems

Sam K. S. Sze
Computer Science and Engineering Department
The Chinese University of Hong Kong
Shatin, Hong Kong
samsze@cse.cuhk.edu.hk

Michael R. Lyu
Computer Science and Engineering Department
The Chinese University of Hong Kong
Shatin, Hong Kong
lyu@cse.cuhk.edu.hk

Henry K. P. Choi
Computer Science and Engineering Department
The Chinese University of Hong Kong
Shatin, Hong Kong
kpchoi@cse.cuhk.edu.hk

Edward H. H. Yau
Computer Science and Engineering Department
The Chinese University of Hong Kong
Shatin, Hong Kong
edyau@cse.cuhk.edu.hk

ABSTRACT

This paper presents the technologies we developed for transmitting video contents over heterogeneous wireless platforms. The video delivery and presentation schemes are encapsulated into a user-oriented, content-aware, and device-adjustable client system for accessing a multimodal, multilingual digital video library. The mobile access system, iVIEW client, provides a friendly user interface that meets the challenge of rich multimodal information presentation on wireless hand-held devices. An XML schema is employed to organize the multimodal metadata for better data interoperability. Furthermore, we investigated a context awareness mechanism complementary to the XML schema to facilitate scalable degradation under restricted resources in a wireless application environment.

Keywords

Browsers on mobile devices, Multi-modal interfaces and applications, Usability and experience, Mobile data interoperability

1. INTRODUCTION

Videos represent rich media contents. Evolution of digital video library (DVL) enables people to search and access rich video contents. The techniques involved in composing videos into vast DVL for content-based retrieval are provided in the literature [1-3].

In facing the challenges of implementing a client software system for rich video information delivery and presentation, specifically in a wireless environment, we encounter two major challenges. First, we require an intelligent user interface. The user interface should be able to support:

- **Flexible content selection: Enable users to select content to be presented in different scale. That is, from a coarse grain, allow the users to view the whole scenario; and from a fine-grain, provide a full-screen to grasp a specific media.**
- **Result set refinement: A large digital video library usually returns lots of results per query. User-friendly interfaces and schemes for data organization should be designed to support various user manipulations on the returned results.**
- **Integration of useful hand manipulation techniques into the overall user operation habits.**

The second challenge is: in a presentation session that synchronizes multiple media under a fluctuating wireless environment, we need a control scheme that manages CPU and memory resources together with the bandwidth utilizations.

2. iVIEW CLIENT ARCHITECTURE

The iVIEW client [4-5] is a component-based subsystem composed of a set of infrastructure components and presentation components. The infrastructure components provide services for client-server message communication and time synchronization among different presentation components by message passing. The presentation components accept messages passed from the infrastructure components and generate the required presentation result. This component-based approach makes the system scalable to support a potential expansion of additional modal dimensions. Figure 1 shows the architecture of the client system. Infrastructure components are marked in shaded color.

The client-server communication message is coded in XML through HTTP. The XML is embedded into an HTTP POST message. Using HTTP enjoys the advantage that the service is seldom blocked by firewalls and web servers are widely adopted in organizations [6]. Once a search result is attained, the multimodal description in XML is obtained from the server. The client parses the XML using Document Object Model (DOM). The infrastructure obtains the media time through playing of the video. The recorded media time is then matched with the media description to seek the event that a presentation component needs to perform at a particular time.

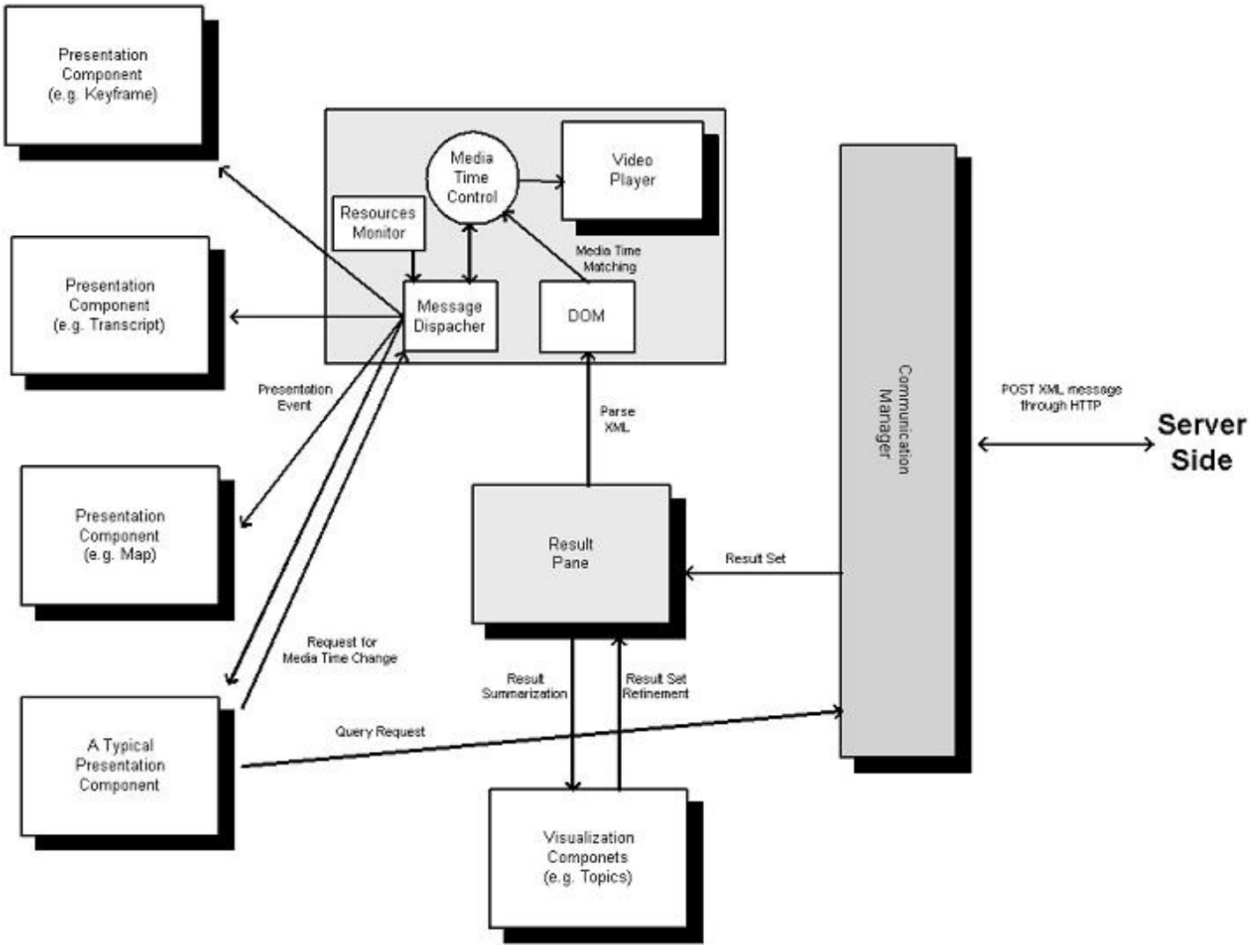

Figure 1. iVIEW Client System Architecture

3. USER INTERFACE

We define modality as a domain or type of information that can be extracted from the video. Examples are the text generated by speech recognition or optical character recognition (OCR), and the key-frames generated by shot breaks detection [7].

The iVIEW wireless client employs a set of mini-windows in its user interface. Each mini-window works as a user control or presentation of a modality. Users can arrange the best order for viewing manually according to the significance of a modality. There are pull-down menu options providing the open and close functions of a particular mini-window (see Figure 2).

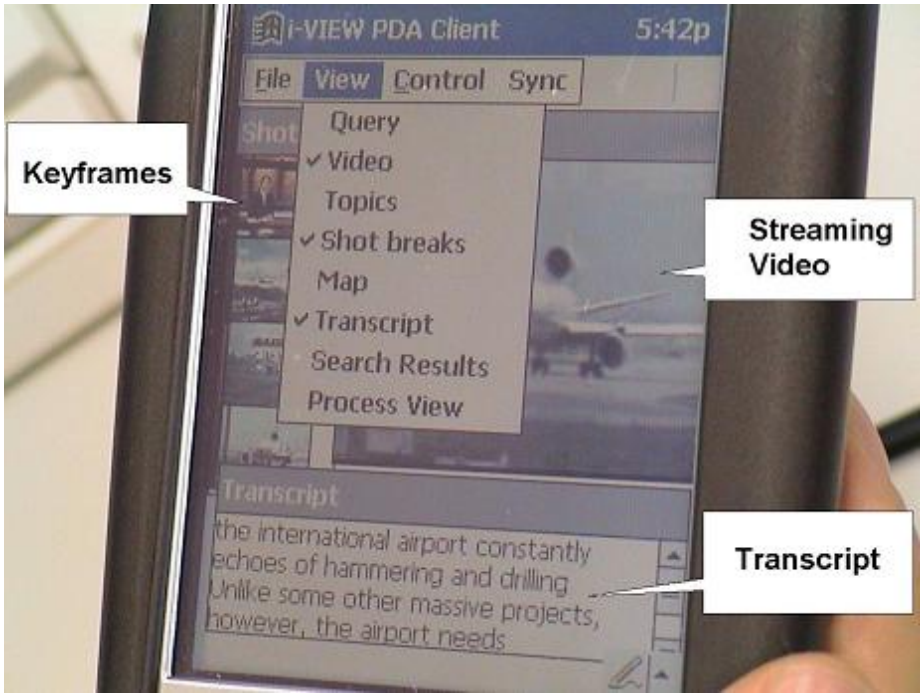

Figure 2. Client Interface

4. SCALABILITY MECHANISM WITH XML

Figure 3 shows an example on the multimodal presentation description XML. This XML format can facilitate scalable multimodal presentations.

```
<?xml version="1.0" encoding="utf-16" ?>
 <sequence path="/iview/video/">
     <time start="0">
         <script> GOVERNMENT HAS RESTORED FULL
              </script>
         <frame file="frame141_00.jpg" />
     </time>
     <time start="2">
         <script> DIPLOMATIC RELATIONS WITH LIBYA
         </script>
     </time>
   </sequence>
```

Figure 3. Multimodal Presentation Description Schema

In this XML schema, a particular video context type is assigned to each video. Moreover, the context type of each modality is specified. Therefore, the client system can be aware of the overall context and the individual modal context being presented.

Different modalities involve different level of system resource consumptions. As different modal information is complementary information for each other, the reduction of a modality may cause degradation in the overall content. Meanwhile, portion of the core content can still be kept depending on the significance of the modality being removed.

The iVIEW wireless client system embeds a performance monitor. That keeps track of system resources, including CPU load, memory usage and bandwidth consumption. Figure 4 shows the visible interface of the performance monitor.

The performance monitor works in coupling with the dispatcher. If the system utilization saturates, the dispatcher will be signaled to make a decision and stop dispatching a set of active modalities of least significance. Mathematically, we can describe the relation as:

$$\max \sum_{i=1}^{n} S_k(m_i)x_i \quad | \quad \sum_{i=1}^{n} R(m_i)x_i \le R_{available} \quad \text{where } x \in \{0, 1\},\ i=1 \ldots n$$

m_i is a particular modality
$S_k(m_i)$ is the significance of a modal presentation in video type k
$R(m_i)$ is the resources utilization of a modal presentation
$R_{available}$ is the total available resources

Figure 4. Resources Monitor Interface

A typical resources measurement result under GSM HSCSD (43.2 kbps data rate) is shown in Figure 5. In this scenario, the network consumption is saturated at the 25th second. The system automatically steps down the resources consumption by suspending the video streaming min-window.

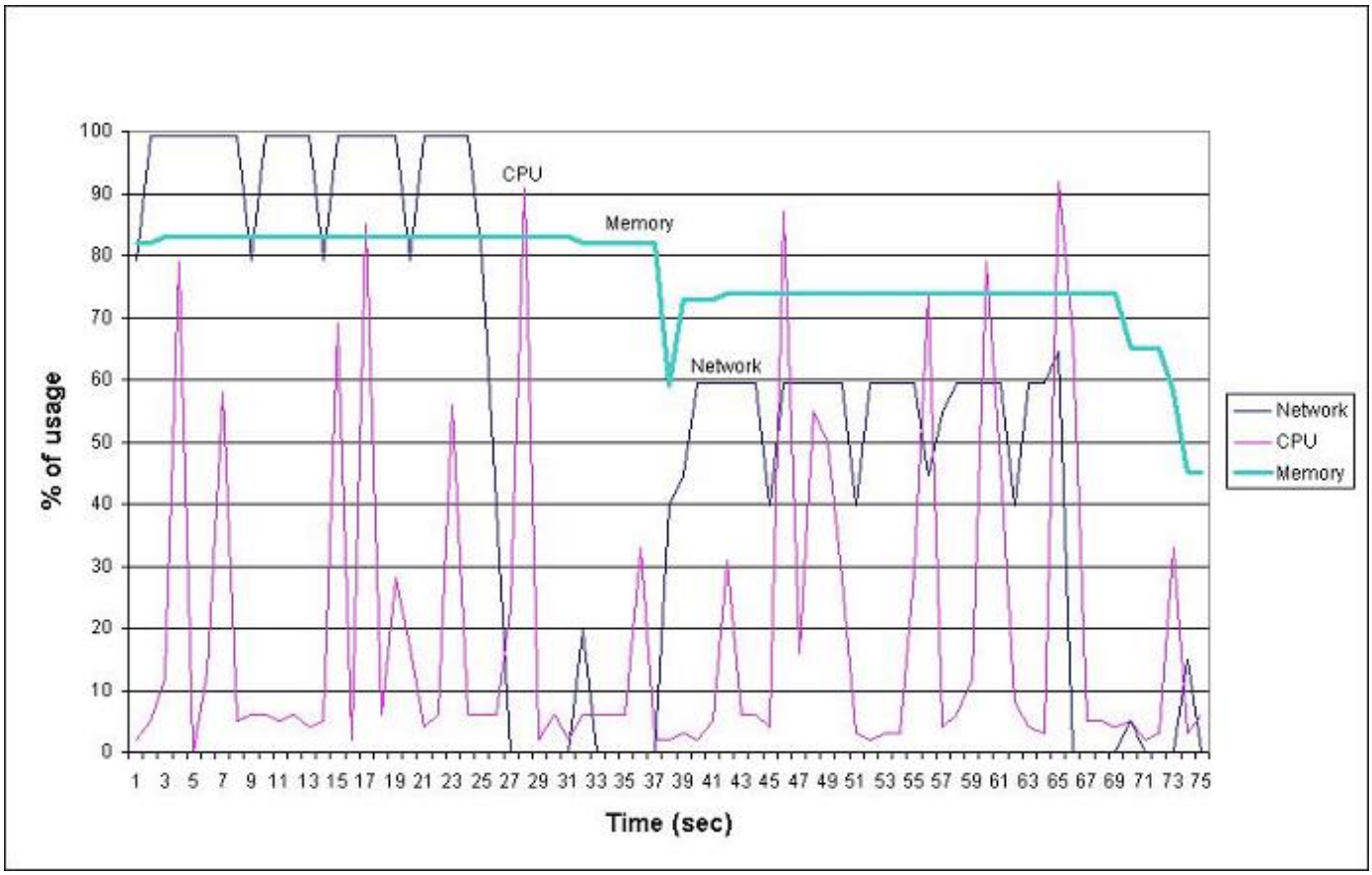

Figure 5. Transition Scenario at Video Suspension

5. CONCLUSION

We implemented a client system for accessing DVL that enables presentation of multimodal video information. We defined a multimodal presentation XML schema with the corresponding context-aware mechanism that optimizes the presentation content under limited resources and fluctuating environment.

6. ACKNOWLEDGEMENTS

The work described in this paper was fully supported by a grant from the Hong Kong Innovation and Technology Fund, under the project ID ITS/29/00, and a grant from the Research Grants Council of the Hong Kong Special Adminstrative Region, China, under the Project No. CUHK4360/02E.

7. REFERENCES

1. **M. Christel, H.D. Wactlar, S. Stevens, R. Reddy, M. Mauldin, and T. Kanade, "Techniques for the Creation and Exploration of Digital Video Libraries,"** ***Multimedia Tools and Applications (Volume 2)*****, Borko Furht, editor. Boston, MA: Kluwer Academic Publishers, 1996.**
2. **M. Christel, A. Hauptmann, and M. Witbrock, "Artificial Intelligence Technqiues in the Interface to a Digital Video Library",** ***Proceedings of the CHI-97 Computer-Human Interface Conference*****, New Orleans, LA, March 1997.**
3. **H.D. Wactlar, "Informedia - Search and Summarization in the Video Medium,"** ***Imagina 2000 Conference*****, Monaco, January 31 - February 2, 2000.**
4. **M. R. Lyu, H. H. Yau, and K.S. Sze, "A Multilingual, Multimodal Digital Video Library System,"** ***Proc. Joint Conference on Digital Libraries*****, Portland, July 14-18 2002.**
5. **M. R. Lyu, K. S. Sze, and H. H. Yau, "iVIEW: An Intelligent Video over InternEt and Wireless Access System,"** ***Proc. 11th International World Wide Web Conference (WWW2002)*****, Practice and Experience Track, Hawaii, May 7-11 2002.**
6. **W. H. Cheung, M. R. Lyu, and K.W. Ng, "Integrating Digital Libraries by CORBA, XML and Servlet,"** ***Proceedings First ACM/IEEE-CS Joint Conference on Digital Libraries*****, Roanoke, Virginia, June 24-28 2001, pp.472.**
7. **M. Christel, A. Warmack, A. Hauptmann, and S. Crosby, "Adjustable Filmstrips and Skims as Abstractions for a Digital Video Library,"** ***IEEE Advances in Digital Libraries Conference 1999*****, Baltimore, MD. pp. 98-104, May 19-21, 1999.**