4 Toward Creation of a New Media Environment

4-1 Research on Interactive Communication Media and Contents

Rieko KADOBAYASHI, Hisashi MIYAMORI, Tatsuya YAMAZAKI, Ichiroh KANAYA, Tadahiko KUMAMOTO, and Myint Myint Sein

The Interactive Communication Media and Contents Group, which was established on last April, is aiming to develop a digital museum. In the digital museum digital contents collected via ultra high speed network from all parts of the world will be dynamically and adaptively displayed to a visitor based on the visitor's wants, feelings, and sense. In this article we describe the research subjects we are tackling.

Keywords

digital museum, virtual heritage, scene retrieval, quality of service, human-computer interaction

1 Introduction

With the rapid development in recent years of digital technologies in a variety of fields, such as broadcasting, publishing, medical care, education, entertainment, and communications, demand is growing for the establishment of high-level multimedia information-processing platforms that will integrate multimedia content. Given the popularity of the Internet in particular, digital content will come to be used more widely than ever, via networks. However, today’s digital content-processing technologies cannot meet the requirements of diversified forms of presentation and use, nor can they respond to the needs of service providers and users. Creators of digital content want to work as efficiently and accurately as possible, but it is currently very costly, for example, to build a three-dimensional digital model of an object. In addition, content providers cannot implement precise billing methods when trying to collect their fees for particular content services, due to the absence of appropriate billing frameworks. This suggests that users are not paying appropriate prices for services.

If such problems are solved, another problem will arise, as digital content becomes structured to be delivered in appropriate forms over the Internet, and stored in large-scale distributed databases. Namely, the new challenge will be in retrieving and providing content that best matches user needs from a massive amount of information. Information in conventional forms, such as broadcasts and newspapers and exhibitions in museums and art galleries, has always been delivered to an indeterminate number of people. As a result, these media have been ill-equipped to fulfill particular, individual needs. One of the features of digital content is that users and providers can easily reproduce it, more so than with conventional media content. Services are therefore expected to respond to individual needs, taking advantage of this feature of digital content[1].

Given these circumstances, we have conducted research into fundamental software and verification systems for the media, aiming to create, store, edit, retrieve, and distribute digital content and to offer services adaptable to individual needs within the next-generation high-speed Internet framework. Our specific research themes are as follows.

(1) Information retrieval and summarizing techniques based on content analysis

One goal of our research is to automatically extract objects (information about people, objects, events, and so on) included in media content and to effect automated indexing, using multi-modal information. We are developing technologies for searching and summarizing indexed content. Then, it becomes possible to view a large number of video images, for example, with a high level of efficiency in a short period of time.

(2) Media quality-control techniques

We are developing technologies for media conversion, filtering, and editing, as well as scalable media presentations corresponding to user needs and communication conditions. These needs and conditions include not only networks, but also end-user systems (such as information terminals and information appliances).

(3) Techniques for individual adaptation of service

We are developing agents that will change media quality and the retrieval display interface, and assist users in creating a viewing plan in response to individual contexts that will vary according to user skills, feelings, interests, and conditions of use.

(4) Digital archive construction techniques

We are developing technologies for constructing digital archives that will meet the needs of the public as well as the professional research sector. In particular, we investigate efficient ways of creating precise three-dimensional models of objects and frameworks for distributing digital data on the Internet. We will construct an augmented reality environment that allows us to handle objects and information, from the present and the past, with no limitations, connecting museums and collections in the real world with virtual exhibitions on networks, and linking digital content with real-time video images (for example, live broadcasts of an archaeological excavation site). In this process, fundamental techniques are integrated and evaluated.

Following are overviews of each research theme.

2 Information retrieval and summarization

As the Internet evolves, a large amount of a wide variety of multimedia information (in the basic forms of video, audio, and spoken language) is being distributed and stored in many locations. As users do not have unlimited amounts of time to view such content, a technique should be developed that allows us to efficiently retrieve, from the massive amount of available information, only the information necessary for use. Search technology involves the automatic extraction of a variety of multi-modal information (about behavior, conditions, people, objects, and the like) from large amounts of media information, and provides the prompt presentation of the specific elements the user needs. The relevant fundamental techniques include automated indexing (marking) of each type of media information and a description and exploration of the standardized context representation format. In addition to these techniques, there is "information summarization," which involves the presentation of extensive media information in brief summary form. Related fundamental techniques include those for describing context-based content (explicit, implicit, or implications behind the context) expressed by the information, and techniques for summarizing their descriptions in a brief format, for ease of comprehension.

Take video scenes, for example. If such retrieval and summarizing techniques are applied to a 90-minute soccer game recorded with a digital TV, it becomes possible to search for user-specified shots, such as shooting and goal scenes, from out of all of the video data, to condense the game to a user-defined length (five minutes, for example), and to view only highlighted scenes.

We have proposed a system[2][3] that enables such content retrieval by focusing on the actions of people in the video. While conventional features in video (such as color, figure, texture and camera motion) have advantage that can be applied to general-purpose video images, the level of feature expression is limited; it is very difficult to describe video content on the same level as that of human perception. Our approach is to use the actions of the people in the video as an explicit index, in order to implement data searches in a way that is similar to the way in which humans comprehend video images. Fig. 1 shows the major movements of the players (with thin black lines) and those of the ball (thick red lines) during an approximately 20-second soccer scene. These are the results of extracting each object in the soccer game scenes and reproducing the movements of each object in the field through automated extraction of camera motion parameters. Fig. 2 shows the action indices for a soccer game. Each index is described, with each transition point of action being regarded as a description boundary and with an identified action ID as the minimum unit. In the figure, (A) and (B) indicate teams and "Obj. X" is the ball. Fig. 3 is a list of search keywords that can be entered by the user during a content search of the soccer video; "Goal" has been chosen as a search key in this example. "Individual action" in the figure is searched with an action index designated for each player. On the other hand, "Offensive play" and "Defensive play" are events created by more than one player, and they are searched based on distance, deployment, and order of interactions, between players. Fig. 4 demonstrates the query results of "Goal." The top window, including goal, is displayed, and

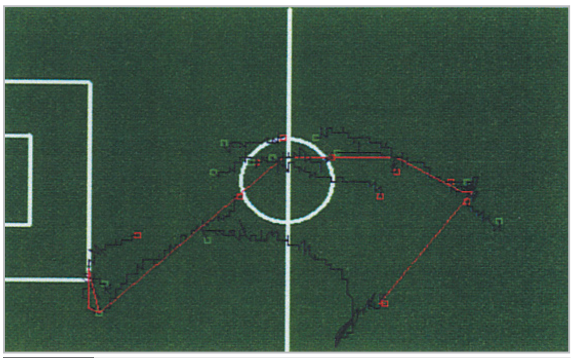


Fig.1 Extraction of objects in soccer game

Thin black line: movements of major players
Thick red line: movements of the ball

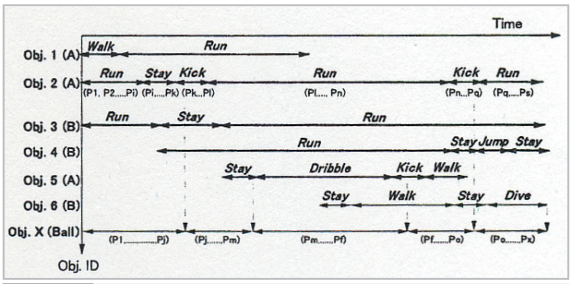


Fig.2 Action index in soccer game

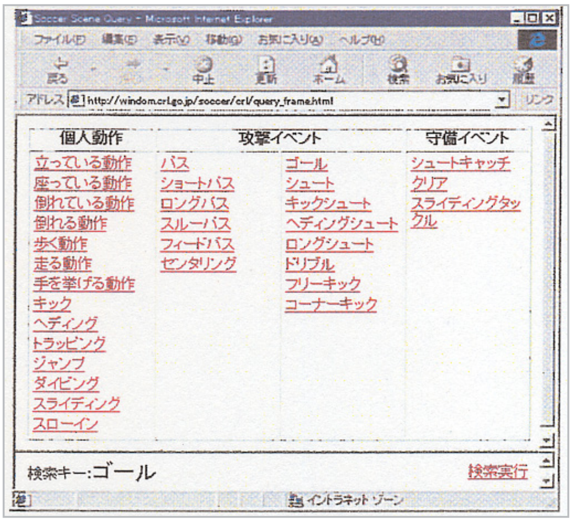


Fig.3 Retrieval Conditions input window

Fig.4 Query results for a specific scene

four goal scenes have been extracted from the prepared video sequence. If you click the top window, the corresponding video scene is replayed.

Based on the above approach, we plan to extend our research into the following items:

(Object extraction from video data)

It is important to have a technique to divide video data into object-to-object pieces for retrieval and to edit each object as well as to control of quality of service. We are developing techniques for object extraction in accordance with user intentions and system requirements.

(Indexing using multi-modal information)

We are developing techniques for generating indices based not only on video information but also on human voice and other audio information. For example, when the audience witnesses exciting events such as a goal, a touchdown, or home run, they cheer. If information on audience cheering is utilized, it is expected that it will be possible to stably extract the indices necessary to search for “Goal” scenes and the like.

(Presentation techniques to help users comprehend content)

This is a technique for helping users to comprehend large amounts of data in a short time. To be more specific, we are developing techniques for “context representation formats,” to describe copious content in a hierarchical manner and to present brief summary descriptions. Included are presentation techniques for scene replay (using a combination of real image and Computer Graphics), and multi-view systems - providing views from positions and angles that cannot be provided with conventional means.

The above research is expected to lead to the following applications.

- ■**Educational application 1**: It will be possible to immediately show students specific scenes retrieved from a video database that hold recorded examples of sports plays and art works. In addition, it will be possible to compare the retrieved examples with student performance, on the spot. This technology will also contribute to flexible search functions on context-based levels, and improvement education and development of human resources based on diverse display functions.
- ■**Educational application 2:** Text materials can be provided that discuss events of great historical and cultural importance (e.g., the typical films and music of each era). The content will be presented in summary, with Q&A and related information automatically displayed on-screen. The system can dynamically change the display levels of the relevant information, detecting misunderstandings on the part of the user, if any, and checking the extent of comprehension. This technology is expected to contribute greatly to educational methodology.
- ■**Monitoring applications:** For example, a number of cameras and sensors may be deployed to monitor rare species, detecting changes from predicted behavior automatically. Or three-dimensional information/textural information may be introduced and moving models created, so that it will be possible to record life forms in a high-resolution digital format. Construction of databases and other means of content-sharing will become possible on a worldwide basis.
- ■**Application to mobile content services:** This is a function allowing for delivery and replaying of packages of selected scenes, on displays of a limited size. A summary page may be created, for example, and displayed on the screen of a mobile terminal. The whole content of the program is recorded on a digital TV at home, for example.

3 Individual adaptation of service and control of quality of service

If digital archives are distributed via networks and widely used by many users, users can share knowledge on a day-to-day basis; the importance of digital archives in society is therefore expected to further augment[4]. When providing digital content via networks, it will become increasingly important to provide information not in a unidirectional way, with a fixed format given by the content provider, but rather in ways that will meet all users' individual needs. In other words, technologies will be required that can retrieve digital content and offer presentations selectively, based on an understanding of individual desires, tastes, feelings and senses. Also required is technology that will edit and convert digital content and services, in accordance with individual user's communication environments. One of the goals of our research project is to develop fundamental technologies and to integrate them into a system that will provide innovative services, through which users can enjoy a variety of digital archives. These will be deployed to create novel communication service considering individual intentions and situations over heterogeneous and complex networks. Also, as pointed out in reference[5], such research will greatly contribute to the evolution of a "network society" and will explore how humans interact with systems and with others, and how such interaction is linked to the development of innovative learning.

In order to adapt integrated service to each individual, the following fundamental technologies will be of importance. First, a technique is required to correctly determine what kind of information each user wants based on a minimum amount of user-input information (voice and interactions with the system, for example). This is principally related to the interaction between humans and systems, as explored in the field of HCI (Human Computer Interaction). This approach requires the analysis of thinking and feeling patterns, an analysis that is deeply related to Kansei Engineering[6]. Further required are techniques for searching for content from among various resources (for example, digital archives) on networks, based on an interpretation of user needs and on technology for dynamically displaying the retrieved results in an organized format suited to each individual.

In addition to techniques for fulfilling individual needs, of importance in the distribution of content on networks will be the delivery of services suited to the communications conditions of each user. The communications conditions include not only the network infrastructure but also the user terminals(or end systems). In this case, IP-based communications on the Internet will be the target of study.

The Internet is a distributed network system consisting of a multitude of autonomous and independent domains correlated with each other. Thus the network quality changes dynamically, depending on which way the data travels in general. In this case, the network quality includes transmission speed, transit delay, and error rate. Despite recent remarkable improvements in broadband technology, there remains a gap in provided quality between core networks and access networks. It is thus necessary to compensate this gap by realizing seamless communications, enabling the provision of higher levels of service.

Digital bits transmitted in the form of packets on networks become meaningful information when received by user terminals. User terminals may be in the form of hardware ranging from a desktop PC (equipped with excellent processing capability and abundant hardware resources) to a mobile terminal (with limited hardware resources, designed for mobile connection to the Internet). In order to provide content in a format suitable to each user, the application quality should be defined beforehand, in relation to the application that handles and provides information to the user. Therefore, it will be essential to have a technique that controls application quality based on the hardware resource of each terminal.

It is also important to control end-to-end(i.e., from information provider to information receiver) quality in a consistent way in order to raise the resource utilization efficiency of the whole information communications system. Thus it will be essential to develop technology for quality control that correlates network quality with application quality. Part of such fundamental technology includes research on QoS (quality of service) map-

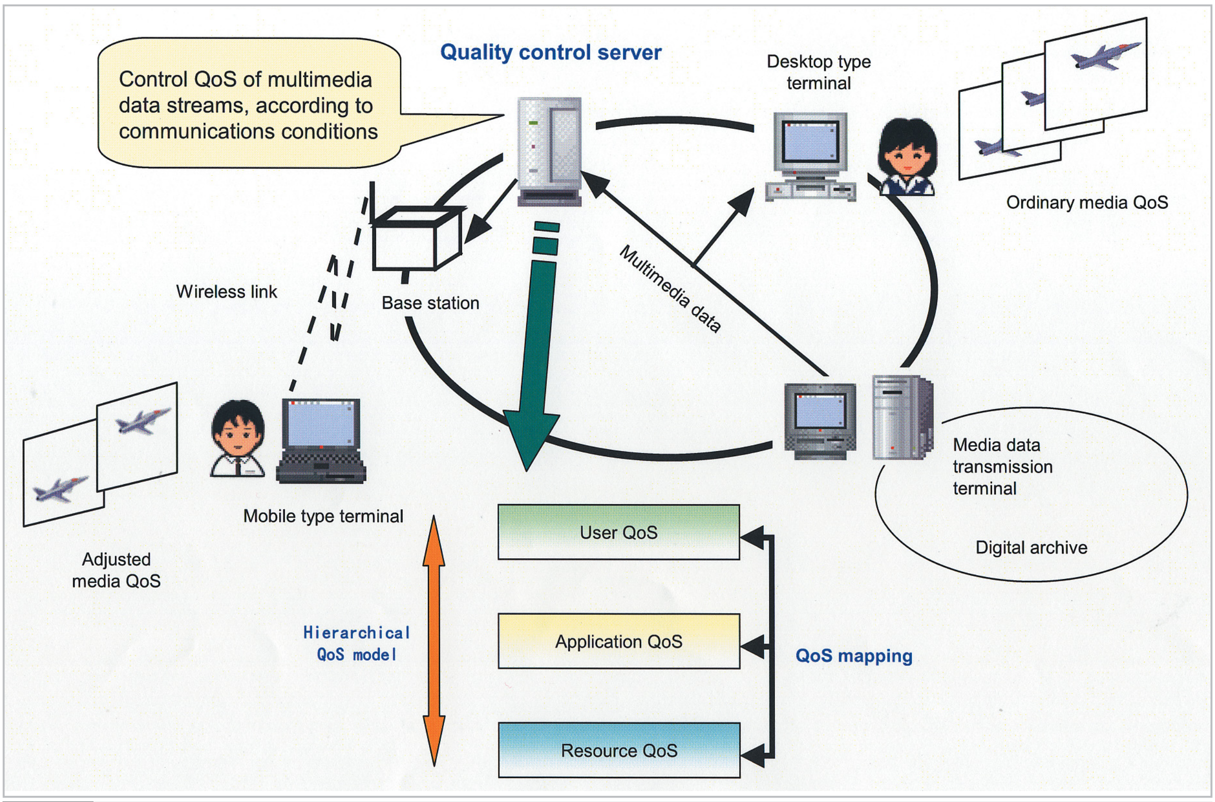


Fig.5 *Quality control for digital content delivery*

ping[7]. One of the primary challenges in QoS mapping is to clarify the relationship between differing QoS parameters in layers and in media and to reflect this clarification in QoS control. The user QoS, which is ranked as the highest layer, is based on subjective estimation, according to user perception. A study considering whether the application quality realized in the user QoS is the best for a given user will be closely related to the skill of each user in adapting to technology and will contribute to the realization of user-friendly and detailed information-communication services.

4 Three-dimensional geometry measurement for "virtual heritage"

4.1 What is "virtual heritage?"

The term "archaeological information" has been used since around 1995. This is the name of a new academic field that was born in the cross-disciplinary area between information science and archaeology. In particular, research for vitalizing items of archaeological heritage such as remains and relics (and the "virtualized" products thereof) is referred to as "virtual heritage".

Then what is "virtualization?" This is the extraction from an entity of information sufficient for virtually reproducing the entity. In general, three-dimensional geometry, surface optical characteristics, and texture, for example, are modeled when constructing a virtual heritage of real artifacts.

There are the following benefits in constructing virtual heritage from real artifacts.

(Eternal conservation) Although artifacts such as relics and other such legacies are subject to degradation with time, their geometries and surface characteristics (i.e. patterns or texture) will never be lost once they have been virtualized.

(Universal disclosure) We have to visit excavation sites and museums to see real relics and legacies. Virtualized relics and legacies, however, can be distributed worldwide via computer networks such as the Internet.

(Academic use) Archaeologists will be able to, for example, easily compare the shape and size of corridor-style stone chambers. Unless the chambers have been virtualized, archaeologists must imagine the three-dimensional shape of each chamber from two-dimensional map.

There are a number of benefits to virtualizing real relics and other legacies, as described above. On the other hand, Computer Graphics products we often come across, like computer renderings of castle towers that do not exist, are not part of a virtual heritage. This is because such renderings are simply imaginary, while a virtual heritage has been digitized from artifacts in existence.

4.2 Construction of virtual heritage

As mentioned above, the virtual heritage is what has been digitized from real relics and other legacies in existence. "Virtualization" is the operation of modeling objects in existence in a way suited to computerization, and then storing the extracted data in computers. The process also involves presenting the model to users in a visually convincing virtual form. The most important characteristics of relics and legacies are their geometries, then their surface patterns and the surface texture. Also important is associated information, such as excavation conditions. When such information is digitized, a virtual heritage can be constructed. For this purpose, we have to acquire precise data concerning the three-dimensional geometries of relics and legacies, with a high degree of efficiency. Non-contact geometry measurement is often required, where mechanical contact with relics and legacies must be avoided in order to protect them.

In light of such conditions, we have been developing image measurement method with high degrees of efficiency and accuracy. Specifically, a multi-sensor fusion method is under development that measures geometry data with several sensors corresponding to several measurement ranges (long, intermediate, or short distances from the object) and integrates the obtained data in computers.

5 Conclusions

This paper has described the summary of research themes our research group is currently addressing. In order to create excellent digital content with ease and to provide many people with opportunities for access to content such as our "virtual heritage" via networks, we have to develop fundamental, wide-ranging technologies and systems that will meet end-user needs. These goals may look contradictory at first glance, but we believe that these goals will balance out, if we continue to pay attention to the ways in which information technology can enrich our lives.

References

1 Rieko Kadobayashi, Yuya Iwakiri, and Kenji Mase: "Personalizing the Museum Exhibitions: Arrangement Issues," to appear in Veljko Milutinovic and Frederic Patricelli eds. Electronic Business and Education: Recent Advances in the Internet Infrastructure", Kluwer Academic Publishers.

2 H. Miyamori, T. Echigo, and S. Iisaku, "Proposal of Query by Short-time Action Descriptions in a Scene", IAPR Workshop on Machine Vision Applications, 3-18, pp.111-114, 1998.

3 M. Kurokawa, H. Miyamori, S. Iisaku, et. al., "Representation and Retrieval of Video Scene by Using Object Actions and Their Spatio-Temporal Relationships", IEEE Signal Processing Society 1999 International Conference on Image Processing (ICIP-99), Oct.1999.

4 Ken Sakamura, "Digital Museum -Construction of a New Type of Museum that Fully Utilizes Computers-", IPSJ (Information Processing Society of Japan) Magazine, Vol. 39, No. 5, pp. 385-392, May 1998 (in Japanese) .

5 Toyoaki Nishida, "Towards Interactive Media for Community Knowledge Creation", IPSJ (Information Processing Society of Japan) Magazine, Vol. 41, No. 5, pp. 542-546, May 2000 (in Japanese) .

6 Hirosi Harashima (Ed.), "Kansei Information Processing", Ohmsha, Ltd., Tokyo, 1994 (in Japanese) .

7 T. Yamazaki and J. Matsuda, "On QoS Mapping in Adaptive QoS Management for Distributed Multimedia Applications", Proceedings of ITC-CSCC'99, 2, pp.1342-1345, Jul. 1999.

***Rieko KADOBAYASHI**, Ph.D*

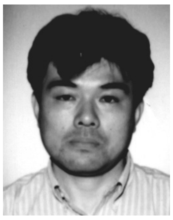
Senior Researcher, Interactive Communication Media and Contents Group, Keihanna Human Info-Communications Research Center

Digital museum, Virtual Heritage

***Hisashi MIYAMORI**, Ph.D*

Researcher, Interactive Communication Media and Contents Group, Keihanna Human Info-Communications Research Center

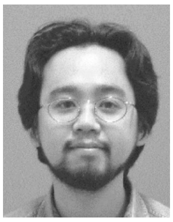
Image understanding, Pattern recognition, Visual communication



Tatsuya YAMAZAKI

Senior Researcher, Interactive Communication Media and Contents Group, Keihanna Human Info-Communications Research Center

Multimedia communication



***Ichiroh KANAYA**, Ph.D*

Research Fellow, Interactive Communication Media and Contents Group, Keihanna Human Info-Communications Research Center

Image processing, Virtual heritage

***Tadahiko KUMAMOTO**, Ph.D*

Senior Researcher, Interactive Communication Media and Contents Group, Keihanna Human Info-Communications Research Center

Human-computer interaction

***Myint Myint Sein**, Ph.D*

Research Fellow, Interactive Communication Media and Contents Group, Keihanna Human Info-Communications Research Center

3D Shape Reconstruction