From MultiMedia to UniversalMedia

Masahito Hirakawa
Department of Mathematics and Computer Science, Interdisciplinary Faculty of Science and Engineering, Shimane University, Nishikawatsu, Matsue 690-8504, Japan

Abstract: Multimedia is referred to ideally as the use of computers to present text, graphics, video, animation, and sound in an integrated way. But, in reality, “integration” is read as “synchronization” among those media. The term multimedia reminds us of the existence of multiple media. In this paper, we present a notion of UniversalMedia. It is not just text, or graphics, or video, or animation, or sound. Rather it is an all-in-one active object. Some trials the author has carried out as a step toward realization of such active media are presented.

Keywords: multimedia, UniversalMedia, human-computer interaction, situation, context-based retrieval, invisible interface, idea generation.

Reference to this article should me made as follows: Hirakawa, M. (2005) ‘From MultiMedia to UniversalMedia’, Int. J. of Computational Science and Engineering, Vol. x, No. x, pp. xx-xx.

Biographical notes: Masahito Hirakawa graduated from Hiroshima Institute of Technology, Japan in 1979, and received his M.E. and D.E. degrees from Hiroshima University, Japan in 1981 and 1984, respectively. He is currently a Full Professor at Shimane University. His research interests include multimodal interfaces, augmented reality, information visualization, video retrieval and editing, and multimedia computing.

1 INTRODUCTION

Multimedia initiatives began more than 15 years ago. Nowadays this term is in common use. Now then, what is multimedia? In [1], multimedia is defined by “the use of computers to present text, graphics, video, animation, and sound in an integrated way.” Integration of multiple media - that’s fine. However, I would say that “integrated” is recognized in most cases as “synchronized” in multimedia community.

For example, consider the Synchronized Multimedia Integration Language (SMIL) [2]. It is an XML-based language, and used for rich-media presentations which integrate streaming audio and video with image, text or any other media type. It enables authors to specify what and when should be presented. That is, SMIL enables them to control the precise time that a sentence is spoken and make it coincide with the display of a given image appearing on the screen.

In addition, we are faced with a serious issue that there are many different hardware and formats for each medium. With regard to this, Golshani brought up a criticism as in the following [3]:

> Why should consumers have to purchase a different device for every type of multimedia file they want to play? As long as a processing unit has adequate memory, a reader (perhaps optical), a visual display unit, and an audio output device, shouldn’t the consumer be able to enjoy a movie, listen to a song, and view pictures? When will we be freed from vendor-dependent formats and technologies?

We are having trouble with such technology-centered approach.

Meanwhile, graphic/web/multimedia designers have experienced for years how the information should be organized and presented. This could contribute to the enrichment of multimedia capability. But they are interested in utilizing media which are available. I would say that we, as engineers, have responsibility of preparing well-formed media to attain further success in media research.

Around 40 years ago, a commercial was telecasted on Japanese TV at which one gentleman sang at the top of his voice – bigger (chocolate) is nicer. And a proverb says that says seeing is believing. Is the picture superior to the text and, again, is the video superior to the picture? Is the video the ultimate media in the sense that it includes motion picture, audio, and text?

The number of media being adopted is, however, not a main concern. What is more important is how natural and fit the media are. The term *multi*media may mislead us into useless research. We thus propose in this paper using UniversalMedia as a replacement of multimedia.

The same wording “universal media” is used as a Web3D consortium working group solution that increases the realism of on-line Web3D worlds (VRML, Java 3D and other on-line 3D technologies) [4]. It aims at allowing content authors to create media-rich worlds that can be instantly loaded over even the slowest dial-up modem Internet connections by providing a small, cross-platform library of locally resident media elements and a uniform resource name mechanism. To tell our idea from this Web3D consortium working group solution, we capitalize the first character of the words and express it by UniversalMedia.

In the rest of the paper, what we intend by UniversalMedia is discussed, followed by some works the author has carried out as a step toward such goal. Specifically, in Section 2, a basic idea of UniversalMedia is presented. Our experimental works will be explained extensively in Sections 3 through 6. Finally, in Section 7, a concluding remark is given.

UNIVERSALMEDIA

When we talk about multimedia, one may refer to data entities such as text, image, audio, and video. Meanwhile, multimedia systems take advantage of human senses to facilitate our communication with the computer. Interaction between human and computer can be improved by the use of multiple sensing channels such as gesture, voice, gaze, and facial expression. In this case, the word modality is preferably used rather than media. A multimodal interface exploits different sensations toward realization of advanced human-computer interface which is natural and expressive to people. Multimodality expresses techniques from user's viewpoint, while multimedia does from system's viewpoint. They are one and indivisible.

Jain who is the first editor-in-chief of the IEEE MultiMedia magazine quoted the fable "Six Blind Men and an Elephant" and mentioned multimedia as follows [5]:

> Each blind man perceived something completely different - the nose as a snake, the leg as a tree trunk, the tail as a rope - and non comprehended the whole. Likewise, a narrow perspective leads most people to consider as multimedia only limited aspects of it. Multimedia is not just JPEG, or networking, or authoring tools, or virtual reality, or information systems. Multimedia is computers using multiple media to let people deal with information naturally.

Go and Carroll also referred to the same story in [6] as a metaphor for understanding different views of scenario-based system design.

Meanwhile, when people talk about multimedia (as data entities), they might have a common understanding that more is better - videos are better than pictures, and pictures are better than texts. I would say this is not correct. Videos are not the ultimate media. It is argued in [7] that sometimes a few words are worth more than any number of pictures.

Is it really essential for us to separate media into text, picture, video, sound, and so forth? Let me give you an interesting example.

Mobile phones are no longer a tool used only for talking. They are used mostly for sending text messages to others. Interestingly, some - mostly young girls - in Japan recognize Hiragana, the Japanese cursive syllabary, as images and form a sentence by a combination of completely irrelevant characters. Those characters or expressions are called Gal-Moji (characters of young girls) or Heta-Moji (awkward characters).

Figure 1 shows an example of Gal-Moji. Left-hand side is a Hiragana expression saying "How are you?", and right-hand side is its corresponding Gal-Moji expression.

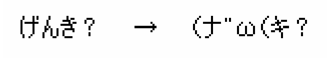

Figure 1 *Example of Gal-Moji*

Those who use Gal-Moji are not peculiar people any longer. In fact, a Karaoke company in Japan has started a service of attaching Gal-Moji to songs.

Gal-Moji users don't care about the type of media. What they are interested in is how much the expressiveness of media is. This shows a necessity of investigating a new type of media that is *all-in-one*. UniversalMedia we present in this paper is such media. Let us modify the Jain's message and say that "UniversalMedia is not just texts, or pictures, or pictures, or videos, or sounds. UniversalMedia is computers using all-in-one media to let people deal with information naturally."

Specifically, UniversalMedia should support the following features.

- **Multiple media expressiveness**
  This is of course a must. However it is noted again that the information could be in text, picture, video, or any other forms, but its meaning may not be captured properly just by looking at the corresponding bit sequence, as seen in Gal-Moji. In other words, it is needed to provide a media conversion facility.
- **Interactiveness**
  It is obvious that the user communicates with the computer continuously by changing his/her query to get a reasonable solution or placing commands to complete a task. Sophisticated interactive capabilities are essential to sophisticated UniversalMedia.
- **Invisible/harmonized computing**
  As noted, UniversalMedia is an all-in-one medium. The word, all-in-one, however doesn't imply just a combination of individual media entities. Those entities must be harmonized so that people feel the medium is one coherent object in its expression and don't need to aware what entity they are manipulating. In other words, this idea is a media version of ubiquitous/invisible computing proposed by Weiser in [8]. Eventually, media - a vehicle for communicating thought or the information in general - should vanish into the background.

- **Context/situation-sensitivity**
  Context- or situation-sensitive computing is one of the active topics in the computer community. The meaning of a UniversalMedia object varies depending on the context to which the object is situated. This implies that UniversalMedia are personalized. Personalization is the process of tailoring the media information to individual users' characteristics or preferences by the use of information either previously obtained (i.e., personal interaction history) or provided in real-time about the user. This helps the user to get a solution satisfying his/her needs more effectively and efficiently, interact with the computer faster and easier and, consequently, have better satisfaction.
- **Knowledgeable**
  UniversalMedia are exchanged among people to convey one's message to others. To ensure a deep understanding of the message (or partners through the message), like the semantic web [9], UniversalMedia should be knowledgeable about themselves and self-explanatory.
- **Lifelong capturing**
  UniversalMedia should be continuous temporally - in other words, the system should capture the lifelong users' activities. Making records of human life itself is of great importance in UniversalMedia. It is presented in [10] that "digitally logging every moment and element of their lives often results in the unanticipated capture of valuable moments. Reviewing recorded, everyday personal interactions, family images, or other information items, such as notes, at any time might lend itself to a human experience with greater intensity and enjoyment." Furthermore, medium is changed - maintained and evolved. It is requested for UniversalMedia to support automated techniques to assist the maintaining, understanding and restructuring of very large and complex media data.

In the following sections, some of the trials we have done with relation to UniversalMedia will be described.

SITUATION-SENSITIVE, LIFELONG INFORMATION MANAGEMENT

Situated Computing

When the desktop metaphor was presented three decades ago, the computer was considered to be a tool for specific tasks such as writing a document, drawing an image and calculating budgets, even though those applications could be run in parallel. This design was reasonable when the user's position was fixed inside the office and the computers had very limited capabilities. With the technological advancement both in hardware and software, it has become possible to develop very advanced tiny mobile/wearable computers. Those computers can be connected to powerful mobile peripheral devices such as a video camera and a GPS receiver, and they also can possess the online wireless connection to the network. However, the interaction with computers are still designed based on the desktop metaphor.

The main objective in situated computing [11] is to enhance the human-computer interaction by utilising the user's real world situation. Here the system organization and data management play an important roll in the success of situated computing as well as the design of user interaction. The situation metaphor, which is proposed as the foundation for situated computing, goes beyond its semantic level in achieving the requirements of situated computing. It is not only an interactive metaphor for visual and non-visual interaction but also a design metaphor which influences on the functional and algorithmic design of applications.

The development methodology consists of three layers. The first level gives the theoretical design adopted in the declaration of situation metaphor. The SIFF (Situated Information Filing and Filtering) framework providing the software abstraction as the core foundation for all services/applications is on the second level. At the third level, tools and applications are organized based on the SIFF framework. We have implemented several example applications which include augmented album, pattern browser, and situation-dependent chat room. In the following we explain those applications extensively.

Augment Album

Multimedia database is one of the highly discussed areas in the multimedia community. In this research area, most of the studies have focused towards the content-based retrieval. The complex unstructured nature of multimedia data makes it very difficult to understand their content without direct human participation. In [12] we proposed the use of contextual information to enrich multimedia data retrieval capabilities, which is the complement to content-based approach.

In order to explain what the context-based retrieval is, let us consider the following two queries for a personal video/image collection: "Find pictures in which a dog is running" and "Find pictures of Christmas party last year at my uncle's house". The difference is that the latter one is much more context-oriented and the content-based approach doesn't work well. Frankly speaking it is impossible to find such pictures in the existing content-retrieval technologies (consider the difficulty in finding uncle's face or house, for example).

Advantage of the context-based approach could also be explained differently as in the following. Filing of data is usually considered a separate independent activity from finding of data, but a proper filing is the key to easy finding. If filing is done improperly, a collection of data becomes almost meaningless even though it has essentially lots of useful information. The inadequacy of keyword-oriented indexing for multimedia database retrieval shows this fact. The context-based retrieval for personal video/image database is considered a proposal of connecting finding with filing, and thus allows the user to retrieve the database effectively. Ideally, we say that queries can be formed according to the way the user remembers about images/video clips in the database.

Though the content of a picture itself represents a situation or situations, our approach was not focused on content-based analysis of the picture. We assume that contextual information is gathered in other ways. In our trial, three contextual parameters are considered: geographical location, time and the corresponding events.

Three components - Map Component, Time Frame Component, and Events Component are provided for query specification. Figure 2 shows a snapshot of the context-based video retrieval interface.

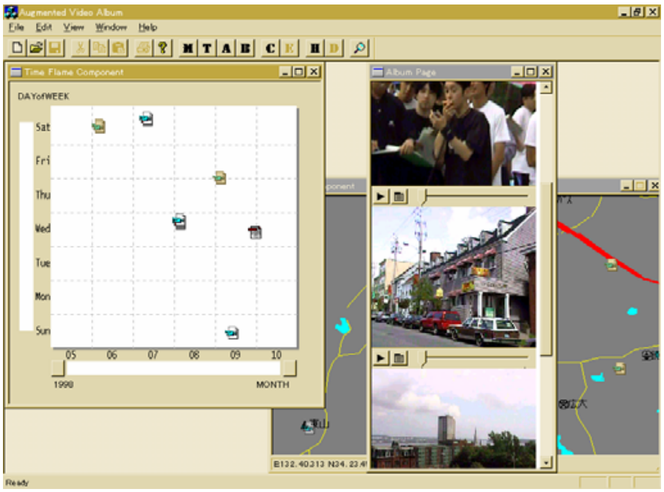

Figure 2 *Context-based video retrieval interface*

Pattern Browser

Pattern Browser is an interactive visualization tool and aims at helping a user to find valuable spatiotemporal patterns in lifelong data [13].

Pattern Browser has mainly two windows allowing the user to specify conditions for the visualization and also to observe patterns from different angles. One is a map window and the other a spiral window.

Map window displays a geographical map in a three dimensional visual space. When the user double clicks a certain location on the map, a spiral appears with the boundary indicating the geographical area it covers, as shown in Fig. 3. Here spirals are used to represent the timeline on which temporal events are marked. This spiral can be moved and its boundary can also be adjusted by dragging it. In this way, the user specifies any number of spirals, as he/she likes. Of course, zoom-in and -out operations are provided so as to be able to change the scale of the map.

A simple linear timeline which has been used commonly doesn't support viewing periodic aspects of the data, though the periodicity is one of the important factors in recognizing patterns. This is why we adopted the spiral expression for the timeline.

A spiral can take two forms depending on the dimensions in which it is visualized. The 2D form of a spiral, known as planar spiral, shows a continuous path of a point in a plane moving around a central point. The 3D form of a spiral is a helix, which is a coil formed by a wire around a uniform tube. Although these two spirals are conceptually described as distinct objects, they can be presented as two different views of a single object in the visualization process. The top view of a helix is presented as its planar spiral in the visualization.

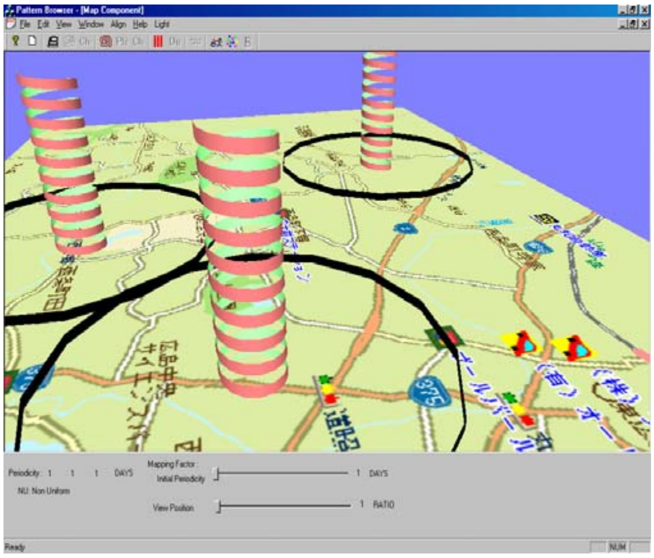

Figure 3 *Spirals on Map window*

Multiple spirals with colored square icons which represent events can be displayed in one window, as shown in Fig. 4. Planar spiral views and linear views are also given in the figure.

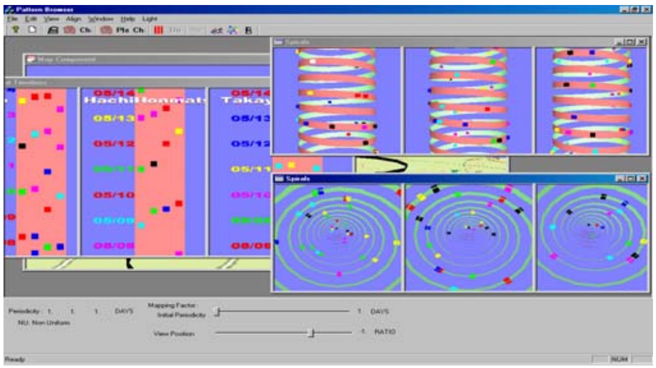

Figure 4 *Display of spirals*

Initially the radius of spirals is proportionate to a single day. This can be modified by changing the visual radius of the spiral or changing the time scaling factor. Furthermore the helix spiral view can be changed to planar view or linear view. The user can navigate through this space by changing the viewpoint, rotating spirals, or adjusting the visual radius of spirals.

Furthermore, target events belonging to a certain category can be aligned to help the user find a temporal pattern, resulting in a non-uniform spiral as shown in Fig. 5. In the figure, icons having yellow color (e.g., sale of a certain product) are aligned on a single line. If the radius is short, it means that the selected event occurs frequently, and vice versa. For example, the left-hand side of the spirals indicates that the sale of the product increases and then decreases repeatedly.

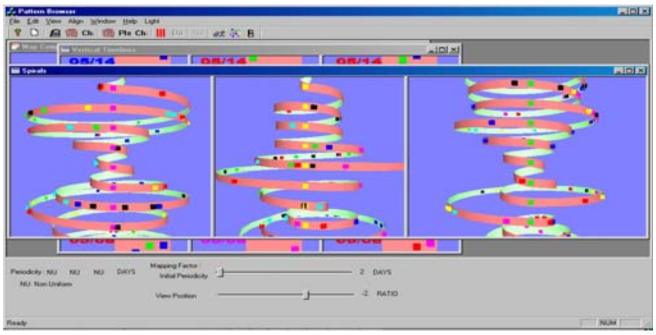

Figure 5 *Irregular spirals*

Chat Room

Situation-dependent Chat Room is another type of applications, which demonstrates how the situated computing is useful in social communication environments [11]. The system allows people in a particular social context to collaborate each other through situated computing. Amusement park such as Disneyland is an example of social environments. People having same objective/willingness gather in a specific area (the place for Mickey Mouse Dance, the place for Play House, etc.) and would like to communicate with others in the same area. But, at the same time, they are naturally shy and afraid of exposing their personal identities to strangers. The situation-dependent chat room provides facilities to establish such sophisticated, situation-dependent communication while protecting each one's individual identity. This application is applicable to other application scenarios at, for example, railroad stations, airports, stadiums and museums.

Figure 6 shows two chat windows and a map pad. As seen in the chat window, when a user enters into a certain area, a chat server allocates a character icon for his/her identification. The user can draw figures and mark locations using the map pad and then share it with other users within the same area.

A message is usually sent to all members in the area but if necessary the sender can restrict it to a specific person. Such private message is sent after clicking corresponding character icon of the target user and it is distinguished by changing the background color as shown in the sixth message in the figure.

In addition, the system also provides a facility called a steady link that allows two members to identify each other whenever they meet again in another area. When one user asks for the steady link, the server requests to a possible user whether he/she accepts it. For a steady couple, when they share a same situation later again, the system assists them to start a cheerful conversation.

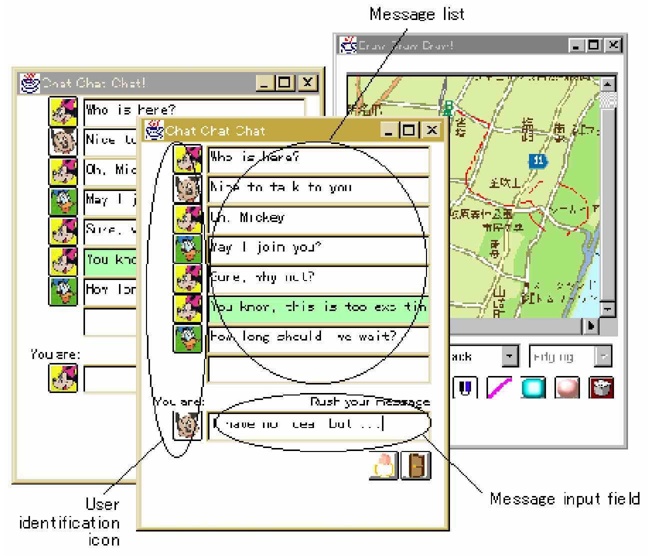

Figure 6 *Situation-dependent chat room interface*

VIDEO EDITING & RETRIEVAL

Video Editing with Change of Camera Motion

Development of consumer digital equipments, e.g., video camcorders and mobile phones with camera (camera phones), allows us to take videos casually. Those equipments are small enough to bring, and thus taking a video is becoming a common activity for us in our daily life. It has been reported that camera phone sales are expected to experience a compound annual growth rate, CAGR, of 55% to reach 656 million units in 2008.

Here those who take such videos are mostly non-professionals – people who have less knowledge of taking good videos. They just press a recording button and take a video without any scenario in advance. They leave what they take in a video to chance. As the result, video materials are not good enough in their image quality and structure to watch repeatedly after they have been taken. Video editing becomes more and more important and crucial for the success in video utilization.

There are various commercial products, such as Adobe Premiere and Apple Final Cut, which help the user to edit video materials he/she has recorded. Here, roughly speaking, facilities provided in such software are to cut a video into segments and then arrange those video segments with certain transitions and titles. In other words, editing of structural aspects (i.e., the order of video segments) is possible, which is called *intra-segment editing*. Yet other powerful tools/facilities allowing the user to edit the contents of each video segment are requested. We call this type of video editing *inter-segment editing*. Researches on inter-segment editing have been active in recent years. However, still further studies are needed.

We have presented a new idea of inter-segment editing [14]. The user can edit a video shot in a scene by specifying desirable camera motions - specifically, pan, tilt, and zoom -

as a postproduction condition. For example, while an original shot is taken without any camera motion, it is possible to generate a shot featuring a continuous pan with an additional zoom-in in a certain part of the shot in order to call viewer's attention to the target zoom-in object.

Image mosaicing which is a well-known technique in computer vision and computer graphics for creation of a comprehensive overview of a video shot content is used as a means of interacting with the user. The video content is visualized in a form of mosaic image, as shown in Fig. 7. The user can specify his/her preference in video editing over the given mosaic image.

It is noted here that the specification of camera motion is also given in a mosaic form. A camera frame denoted by a rectangle box in Fig. 7 is a primitive component used to specify a preferable camera work, and is defined by dragging a mouse on the mosaic image. Semantically it specifies static parameters on the position of camera center and the zoom factor. Dynamic features, that is, the movement of a video camera, are specified in terms of a sequence of camera frames, forming a visual sentence as in the figure.

Figure 8 shows a sequence of snapshots (a part) taken from a source video (upper row) and a resultant video (lower row), for the given specification of visual sentence in Fig. 7.

Figure 7 *User-interface for video editing*

Figure 8 *Source and resultant videos*

Content-based Video Retrieval based on Grammar of Film

In content-based video retrieval, most of the attention have been directed at the motion of an object, its color, shape, texture, and spatio-temporal relationship among objects. We proposed taking simple yet meaningful visual features such as shot length, image dynamics, and the pattern of similar shots in videos, which are not observable explicitly by the viewer [15].

Here the basic idea was to use a concept so-called "grammar of the film". It includes rules or techniques that film directors/editors have experienced and captured in long years, aiming at emphasizing a certain meaning of a scene as well as giving better understanding of scenic situations to the viewer. We explain briefly three types of rules that we have implemented.

- **shot combination by length**

The length of a shot implies the passing of time. When quickness or heavy action needs to be emphasized, a number of short shots are connected. On the other hand, slowness can be emphasized by a sequence of long shots. Moreover, people feel tensed if the length of shots is gradually shortened. In summary, by analyzing the shot length, we can extract monotonous atmosphere, excitement/shock, increased tension, dynamic atmosphere, and a release of tension embedded in videos.

- **visual dynamics of shots**

Visual dynamics of shots do not refer to extraction of certain semantic objects, but focus on the overall appearance of images. There is a little difference between frames in a static shot, whereas there is much difference between frames in a dynamic shot. The calm/slow passing, active/busy, and lively/energetic scenes can be retrieved by evaluating visual dynamics of shots.

- **shot combination by similarity**

We focus on the rough appearance of shots in considering shot similarity. If color range and color distribution are roughly the same for two shots, they are regarded as similar shots. Here there are two types of combination of similar shots. A pattern of A-B-A'-B', where A' (or B') denotes a similar shot of A (or B) is used in the case of a conversation in a static scene. On the other hand, a chase scene is represented with a pattern A-A'-B-B' as well as A-B-A'-B'.

Figure 9 shows a snapshot of the retrieval interface. When the user selects particular semantic content in a given list, a set of editing/scene features is presented. He/she can choose the level of strictness of the query condition, which corresponds to changing the combination of editing features or their thresholds referred to in query evaluation. The result of the retrieval is shown at the lower part of the interface window as in the figure.

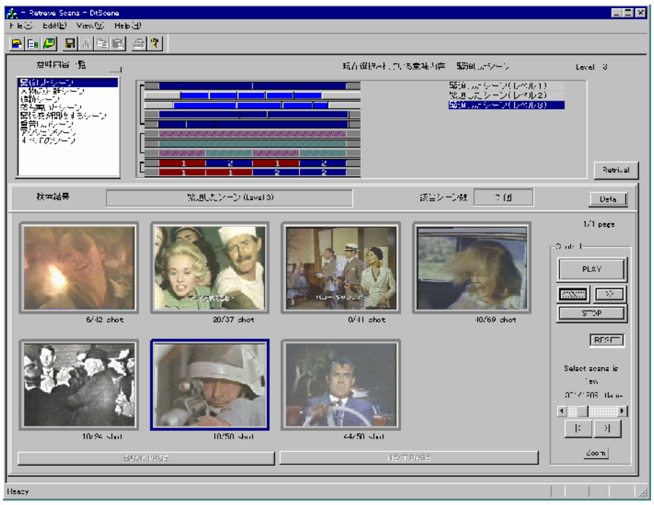

Figure 9 *Interface for retrieving videos based on grammar of film*

TRANSPARENT INTERFACE MEDIA

Media research covers human-computer interaction. One interesting approach from UniversalMedia point of view would be augmented reality, or mixed reality in a broader sense, where computer-generated digital information is mixed with real world objects and they are perceptually assimilated. In other words, we can feel that objects in the real world are augmented in their properties and functions with the help of computers. In addition computers are invisible to us. We don't need to be aware of the computer and learn how to operate it. What to do for us is just behaves as usual as they do in the real world - for example, placing a book on a desk, opening it, and pointing at an object on the page.

There are several different approaches to integrate the digital information with real world objects. Utilization of a head-mounted display (HMD) is the most common approach. Though HMDs are powerful and advantageous as a tool to immerse the user in a smartly integrated environment, the user is forced to wear an HMD. This is bothersome for the user and impediment to his/her work.

As another approach, we assumed a projector-based one where digital information is overlaid onto real world objects through a video projector [16, 17]. Unlike other existing projector-based systems, we proposed to use a transparent, film-type screen which is placed between the user and the real world objects. The user can see those objects through the display. When the user selects an object by pointing at it with his/her finger, the associated computer-generated information is presented on the display.

Interestingly, the display is almost transparent and doesn't have any screen frame which separates the virtual world from the real world. The real and virtual worlds can be integrated seamlessly, resulting in a possibility of truly smart media space.

We have demonstrated the system applicability to several application domains which include zoo/museum and library [16]. Figures 10 and 11 show snapshots of the system.

In Fig. 10, the user sees and points at an object through the display. The associated information is presented on the display. Position of the information on the display changes in accordance with the movement of the user. The display between the user and objects in the real world blocks his/her manipulation of those objects with hands. However, in such applications as museum and zoo, visitors are not allowed to touch exhibits and animals for protection, security, and safety. This system configuration doesn't lose generality.

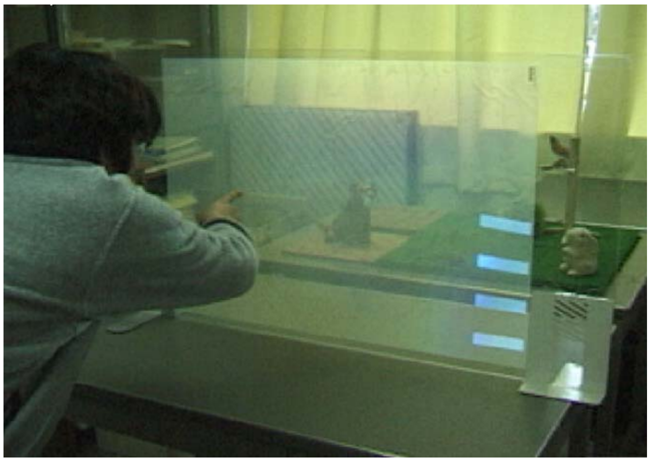

Figure 10 *Transparent interface system*

The system can present a relationship among objects as well, as in Fig. 11. When the user points at a tiger, a food chain relationship can be presented with an arrow, explaining that rabbits, mice, and birds are eaten by tiger.

Figure 11 *Display of a relationship among objects*

Furthermore, additional facilities allowing multiple users to exchange their ideas and/or keep pace each other via object(s) in the real world have been implemented [17]. Real world objects serve as a mediator among users, as well as a media to be augmented in its value by the computer. A snapshot of the system is given in Fig. 12.

A real world object (moving duck toys or LEGO car in our implementation) can work as a career of a message. The message is kept within the object and transferred to someone who is interested in the object. In the current implementation, several possible messages are provided in advance and the user selects one of them. The message can be recalled by choosing the corresponding object later. Incorporation of a voice input/output facility enhances the usefulness of the application.

Figure 12 *Interaction with object-based message exchange system*

PROVIDING A CLUE FOR IDEA GENERATION

In our daily life, we are asked occasionally to generate ideas or remember something important, ranging from serious ones (e.g., finding out a solution to a research problem and making a plan for wedding) to everyday stuff (e.g., making out a menu for dinner, selecting a birthday present, remembering an appointment with someone, and doing a given assignment). However that is a tough work. Spending a lot of time/energy does not necessarily bring a good result. Rather we sometimes get an idea while we are walking on a street, having a lunch, going to sleep, or doing something irrelevant to the work, triggered by chance by a clue.

Imagine, for example, hot water overflows when you soak in a bathtub. You may think "Did I put on a bit of weight?" On the other hand, if you were Archimedes, you might notice that a body immersed in a fluid is buoyed up by a force equal to the weight of the displaced fluid.

Providing a certain clue could trigger us to take an action which may not come to mind without the clue. This feature is very interesting and effective for us. We have developed a tool of providing a clue for assistance in idea generation, remembrance, and recalling [18].

A basic assumption is the following. A message could become a clue for a user if the semantic/cognitive distance between the message and (internal) ideas in his/her memory is shorter than a certain threshold. An idea which is associated with the message comes out as a visible one. It is expected that a message takes a higher abstraction level so that the message works as widely as possible for idea generation. However, if the level of abstraction exceeds a tolerable limit, the message can no longer be a clue since it is impossible for the user to associate the message with any stuff in his/her memory. A message should thus be carefully designed.

In our study, we proposed using pictorial symbols as an aid of idea generation or recall, not like the case (textual messages) in fortune-telling which is a traditional way of idea generation. Though texts allow to convey concrete meaning which is recognizable by almost all users, our aim in the study is to develop a software tool for helping the user to generate ideas or recall something important. Polymorphism, or a capability of expressing a variety of meaning, provided in pictorial symbols may sometimes cause a trouble in graphical user-interface, but works effectively in the context of our study.

Actually we adopted pictorial symbols (icons) provided in the LoCoS visual language, which was originally designed as a language of communication among people in the world. Pictorial symbols provided in LoCoS are easy-to-learn. Interestingly, those symbols take rather abstracted expression. This can help the user enjoy using his/her imagination.

We implemented three types of recall support system - a slot machine version, a roulette version, and an one's-own-way version. In the first two versions, each of the LoCoS symbols used is classified in advance into one of the three roles; that is, a subject, a verb, and an object. One symbol is chosen arbitrarily from each of the three role sets, resulting in a sentence. The user may catch its meaning as what the sentence represents under the original LoCoS semantics. It is noted that, however, the sentence generated in the system may not semantically be correct.

Meanwhile, though each LoCoS symbol has its own meaning, it is not requested for the user to completely understand the meaning as it is, since our goal is to help the user recall events, appointments, or some other stuff. It is encouraged for the user to have a variety of interpretations on the sentence. Any interpretation should be left for the user.

According to this design idea, we implemented the one's-own-way version. Figure 13 shows an interface of the system. LoCoS symbols, or visual elements, appear in a two-dimensional area of the interface. One symbol or a group of symbols which are adjacent each other on the screen could serve as a clue to the user.

Each of them moves independently of others. Here, the initial setting of symbol selection, position, moving direction, and rotation angle is determined randomly. The user is allowed to choose the moving speed and the number of symbols to be displayed on the screen arbitrarily by pressing a button or sliding a lever. In addition, as you can see in Fig. 13, symbols are transparent. When two or more symbols are overlapped, they are merged graphically, resulting in a new symbol. As the result, we have a chance of getting an infinite number of symbols.

Figure 13 *Idea generation system*

CONCLUDING REMARK

Media research is crucial for the success in next-generation computing. Considering that the word multimedia is misleading, we presented in this paper a concept of UniversalMedia to make the point clearer. Simply, UniversalMedia is not just text, picture, video, or sound. It is all-in-one media to let people deal with information naturally. In addition, UniversalMedia is not passive in the sense that it is capable of sensing others and changing its role with the correspondence to others' properties. Moreover, it keeps a lifelong history of target occurrences and events, resulting in achieving media evolution.

We also presented some trials toward realization of such media. However they are considered just in the initial stage. Further studies must follow.

In summary, efforts should be devoted to enriching the human's creativity and life.

ACKNOWLEDGEMENT

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research, 16500060, 2004.

REFERENCES

1 http://www.webopedia.com
2 http://www.w3.org/AudioVideo/
3 Golshani, F. (2005) 'Pushing forward with interoperability', *IEEE MultiMedia*, Vol. 12, No. 1, c3.
4 Walsh, A.E. (2000) 'Universal media: Media-rich content for bandwidth-starved devices', *ACM SIGGRAPH Computer Graphics*, Vol. 34, Issue 2, pp. 37-41.
5 Jain, R. (1994) 'Multimedia computing', *IEEE MultiMedia*, Vol. 1, No.1, pp. 3-4.
6 Go, K. and Carroll, J.M. (2004) 'The blind men and the elephant: Views of scenario-based system design', *ACM interactions*, Vol. XI, No. 6, pp. 44-53.
7 Johnson, J. (2000) 'Textual bloopers', *ACM interactions*, Vol. VII.5, pp. 28-48.
8 Weiser, M. (1991) 'The computer for the 21st century', *Scientific American*, Vol. 265, No. 3, pp. 94-104.
9 Berners-Lee, T., Hendler, J., and Lassila, O. (2001) 'The semantic web', *Scientific American*, Vol. 284, No. 5, pp. 34-43.
10 Nack, F. (2005) 'You must remember this', *IEEE MultiMedia*, Vol. 12, No. 1, pp. 4-7.
11 Hirakawa, M. and Hewagamage, K.P. (2001) 'Situated computing: A paradigm for the mobile user interaction with multimedia sources', *Annals of Software Engineering*, Vol. 12, No. 1, pp. 213-239.
12 Hewagamage, K.P. and Hirakawa, M. (2000) 'Augmented album: Situation-dependent system for a personal digital video/image collection', *Proceedings of IEEE International Conference on Multimedia and Expo*, New York, USA, pp. 323-326.
13 Hewagamage, K.P. and Hirakawa, M. (2001) 'An interactive visual language for spatiotemporal patterns', *Journal of Visual Languages and Computing*, Vol.12, No.3, pp. 325-349.
14 Hirakawa, M. and Nakashita, K. (2004) 'Video editing with change of camera motion', *Proceedings of International Conference on Distributed Multimedia Systems*, San Francisco, USA, pp. 196-200.
15 Yoshitaka, A. Ishii, T., Hirakawa, M., and Ichikawa, T. (1997) 'Content-based retrieval of video data by the grammar of the film', *Proceedings of IEEE Symposium on Visual Languages*, Capri, Italy, pp. 310-317.
16 Hirakawa, M. Kojima, Y. and Yoshitaka, A. (2003) 'Transparent interface: A seamless media space integrating the real and virtual worlds', *Proceedings of IEEE Symposium on Human Centric Computing Languages and Environments*, Auckland, New Zealand, pp. 120-122.
17 Hirakawa, M. and Koike, S. (2004) 'A collaborative augmented reality system using transparent display',

Proceedings of International Workshop on Multimedia and Web Design, Miami, USA, pp. 410-416.

18 Tanabe, C. and Hirakawa, M. (2005) 'A LoCoS-based tool capable of providing a clue for idea generation', *Proceedings of International Conference on Active Media Technology*, Takamatsu, Japan, pp. 51-56.