Desktop program production
- DTPP -

Kazumasa Enami, Kazuo Fukui, and Nobuyuki Yagi

NHK Science and Technical Research Laboratories

Abstract: In order to conform to the needs of effective program production in multimedia era, we are studying Desk Top Program Production system. With the DTPP, users can easily produce multimedia program icluding video, sound, and ancillary data, and freely handle video images synthesizing video components retrieved from video database. This paper describes the new program production system, DTPP and its key technologies such as cooperative program production via multimedia network, indexing and utilization of attribute information of images, and image segmentation and spatio-temporal editing.

1. Introduction

With the progress of technologies in the fields of digital audiovisual processing, mautimedia computer, semiconductor, and human interface, broadcast services are expected to undergo significant changes. The applications of the new services includes multi-channel broadcasting, digital HDTV broadcasting, multimdia broadcasting, electronic newspapers, single frequency network, and mobile TV reception.

Thus activities are intense in the development of media for new services. However,whether these services will be successful or not depends extremely on the contents. The development of contents and its tools to supply programs to these media, which should be taking place ahead of that of the media, is lagging behind.

Program production is a considarable creative task that requires the concerted efforts of all people involved, including scenario writers, producers, artists, and technical staff. Is it easy to improve the environment of program production by technology just as writers have switched from "paper and pen" to word processors to aid their creative work? Unlike writers and art painters, program producers need very expensive and complex video/audio machines. Further, the work is rather labor-intensive by a lot of staff. In other words, it takes a lot of time and manpower to produce TV programs. What we need, therefore, are machines that handle routine part of the work and tools that let the producers concentrate in their creative work.

Our research group has been working on a way to provide a new program production environment for more efficient production of high-quality programs. We call this new work environment DTPP (Desktop Program Production) [1]. In the field of publishing, the latest technology is DTP (Desktop Publishing) which allows a user to edit and lay out words, photographs, figures, and charts on a computer display at will. The DTPP is a multimedia version of the DTP. Whereas the DTP in publishing needs only to handle words and static images, the DTPP has to handle moving images and audio, which requires technologies far more complex and wide-ranging.

In this paper, we first introduce the system configuration of DTPP that consists of multimedia workbenches, media server, computing server, and then describe the main functions of DTPP; handling attributes data linked with audiovisual data and spato-temporal edting with video components.

2. DTPP system

DTPP we have proposed is a total system supporting the entire program production procedures from planning to broadcasting. Fig. 1 shows the concept of DTPP and examples of support at the each process of program production

The DTPP system consists of multimedia workbench, a media server to store multimedia data including video/audio materials, and their attributes information; a computing server to process the multimedia data, and a multimedia network to connect these equipment.

Ideally, the network should be able to accommodate several non-compressed video signals including HDTV signals with a data rate of 1.2 Gbit/s. The cost and scale of the high-speed network system will be extremely high. With the script-driven process, which is described later, currently available transmission media based on today's technology can be applied to reasonably realize such a network. Under the script-driven process, the DTPP system handles compressed video data for monitoring via the network and uncompressed video data for broadcasting.

(1) Multimedia workbench

As a desktop production environment, each of the production staff is provided with a computer terminal capable of handling multimedia, that is the multimedia workbench. These workbenches are installed not only at a broadcasting station, but at local stations and other geographically distant places, providing a common work space for the staff at some distance from each other. The user can operate these

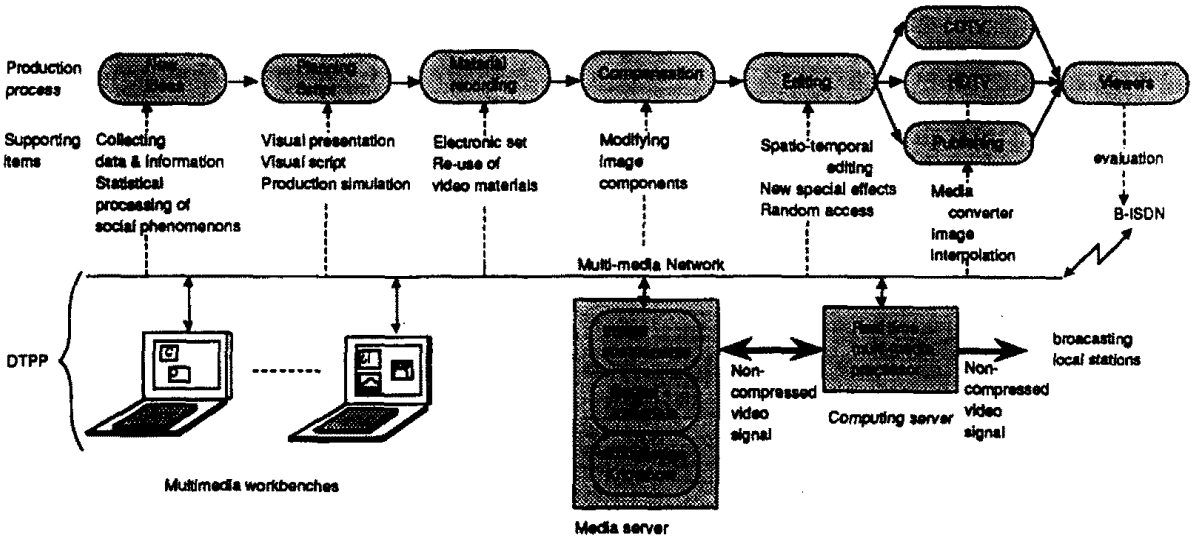

Fig.1 New program production environment -DTPP-

multimedia workbenches just as ordinary computers, retrieving information on the Internet at a planning stage or producing plans and scripts by utilizing the word processing function. Further, the user can perform many other different tasks, including: non-linear editing[2] in which users determine editing points or special effects while randomly accessing and browsing video/audio materials stored in the multimedia server; visual communication with other terminals; sharing of production information; and controlling the computing server that processes video/audio.

(2) Media server
In addition to video/audio, the media server stores attributes data and hyper-link information that describes relations among these multimedia data. The multimedia data stored in the server are managed with the time axis using time codes.

Physically, several media servers are distributed along the network, but they are synchronized; pictures within different media servers can be read and composited simultaneously. Further, they can be accessed by several users at the same time; random accessing is also possible. As the system is required to possess huge memory capacity, it is expected to have a hierarchical architecture using semiconductor memory, hard disk, and tape media as shown in Fig.2 [3].

(3) Computing server
The computing server is capable of real-time and programmable processing of video/audio signals. With just a single type of hardware, it performs not noly such different functions of current program production systems as video switching, video effecting and image quality correcting, but also new processing items required by producers and designers. The server with this much flexibility and the ability of real-time processing of video signals and other high-speed data would naturally have a parallel processor architecture such as Picot [4].

The computing server processes multimedia data read from the multimedia server to create special effects and quality correction. The processing functions can be flexibly changed according to the script written by the user just like a computer which can change the contents of data processing depending on software. The server can be used for HDTV and other television systems as well as for regular television broadcasting. The server sends out the results of its processing to transmitters and local stations for broadcasting.

(4) Script-driven processing[5]
The video/audio materials are edited while they are monitored on the DTPP terminal, but the results are not recorded and sent out as directly processed video/audio signals. Instead, they are recorded as the procedures of processing and editing (we call them the "script"); the materials themselves are left intact. From the multimedia workbench, only this script is sent via the network to the media server and the computing server. The script is actually used for real-time processing in the computing server when the program is broadcast.

With this script-driven processing, we can construct a system with a relatively low-speed network and immediately produce several programs using the same material simply by changing the script. Further, the system of this kind retains high picture quality as the video materials are composited and edited only during outputting.

(5) CSCW (Computer Supported Cooperative Work)

A TV program is produced by producers, casters,

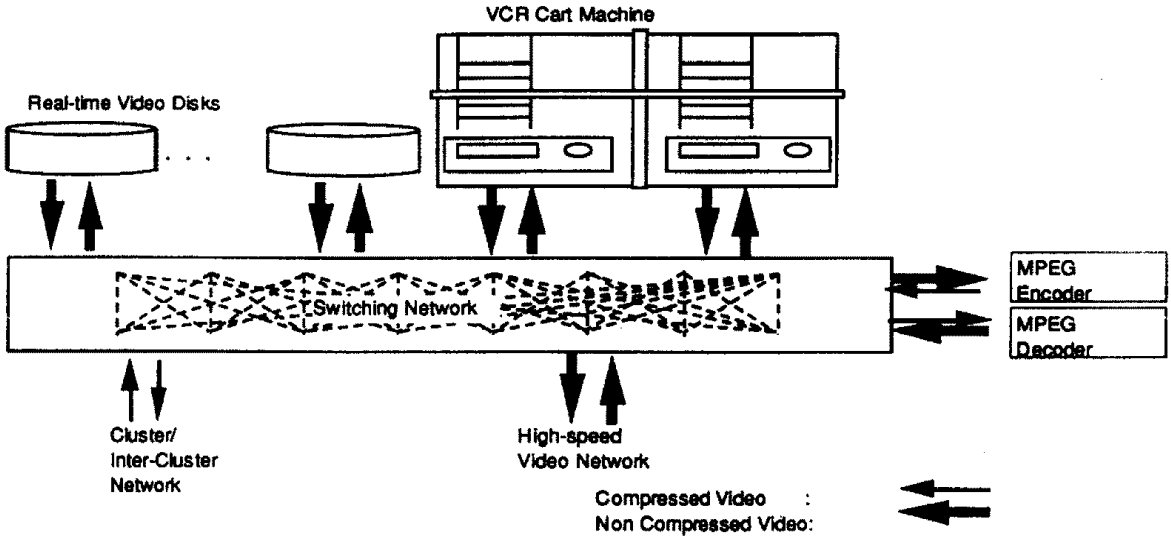

Fig.2 Cofiguration of the Multimedia Server -hierarchical memory system

cameramen, video/audio technical staff, CG producers, and studio setting artists doing their own jobs along the predetermined procedures. These people check each other's work progress as they proceed with their own tasks by following the schedule.

Some project in Australia has reported on an experiment of CSCW during the pre-production stage[6]. The experiment tested visual communication between project staff who were geographically separated to discuss casting, location points, methods of video effects, contents of CG, and so on. The project reported that the CSCW system could successfully reduce both time and cost of production.

The DTPP aims to improve the efficiency of program production by promoting CSCW via the network not only for pre-production but also for proceeding procedures. In the new environment of cooperative work, all the staff members write, check or change, with the multimedia workbenches, the schedule control sheet, messages, scripts, story board (instruction sheet), and video/audio program itself as the final product.

The CSCW in the DTPP features Story-board processor. The story-board controlled by this processor facilitates gradual but steady program production by replacing the imagery written on the story board with the actual video/audio. Suppose that, for some reason, the order of the scenes on the story board has been changed. The system will then automatically switch the video in the program that has been produced.

3. Attributes data and Video Indexing

Some researchers have reported on image processing methods to extract features from images for video material browsing and retrieval. A system that can automatically detect cut points from video sequences, for instance, takes a start scene from each cut and use it as the representative image[7], or structuralize the video by interpreting it from a viewpoint of time and space and then select and display a representative image [8]-[10].

On the other hand, the DTPP preserves edit information produced during off-line editing of video/ audio materials and control codes for special effects managed along with the time code. They precisely indicate the video cut points in the program without using image processing methods. Further, the ability to detect the camera works (panning, tilting, dollying) during material shooting provides an important clue as to video movement. These physical data are important attributes data.

In the process of program production, a variety of words information is produced including planning sheet, a script, a story board, and memos. The program will be produced according to these documents. When viewed from the finished program side, we can say that the planning sheet clearly describes the contents of the program and the intention of the producer, while the script and the story board indicate how the images in the program is constructed. Memos contain a variety of information on the site of shooting, weather, description on the scene and shooting objects, comments of interviews, and so on. Because of this, they express the contents and meanings of the scenes composited by video/ audio in a straightforward manner. In the DTPP these various descrptions are also preserved as attributes data.

As shown in Fig. 3, the amount of attributes data increases as program production progresses. By

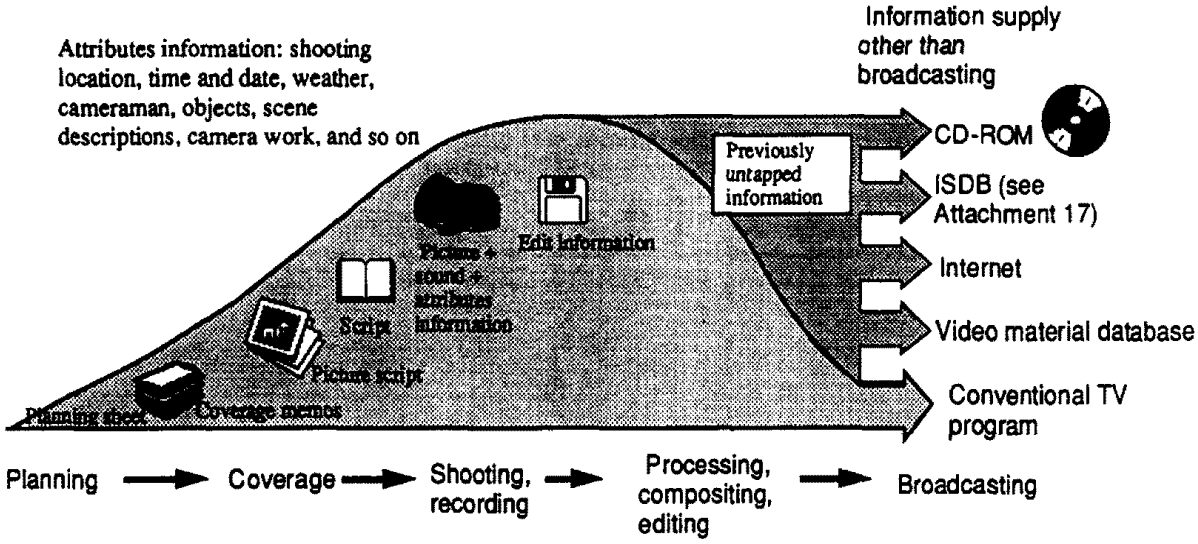

Fig.3 Attributes data generated during program production

processing them, we can easily produce various kinds of multimedia such as the "home information refrigerator" [11] that is a proposed multimedia broadcasting utilizing video index, package media, and video databases. This is a substantial improvement from regular TV program production made of only video/audio condensed within the limited framework of time.

A number of ways have been put forward to get the attributes data, but manual handling is still the main mode of operation at the moment. Take NHK's "Eizo no seiki (Century of Video)," for instance. Tens of thousands of cuts were prepared in this program— the attributes data were gotten and inputted into personal computers by part-time workers for video indexing. These data have greatly contributed to labor-saving in material retrieval. With the integrated work environment of the DTPP, producers will simply process a variety of information on the DTPP workbenches, generating attributes information naturally linked to video/audio without any additional burden.

We are studying on ways to indexing the attributes data on video materials with natural language, and not with keywords. The aim is to develop a system that allows producers with little technical background to input attributes information without difficulty, and that coverts script descriptions written in a natural language into codes that the computer understands[12].

With attributes information, we are examining the method of scene descriptions and structuring [13]. Fig.4 is an example of scene description. The top of the figure expresses images, and the contents of the script are shown below it. In the bottom, time is represented by the horizontal axis, while the objects in the scene, their actions, and camera works (these are called descriptive components) are by the vertical axis. The time axis of each descriptive component shows their existence in the scene. The image is structured using the existing lengths of these descriptive components. By structured scenes, we can automatically select the image that best represent the scene or summarize the program.

4. Video Components and Spatio-temporal Editing

The primary task of program production is to visualize ideas that a producer has in mind. For this purpose, actors perform in a set created in studio and/or scenes are taken on location. In actual program production, however, physical limitations restrict the freedom of video expression. CG can then be utilized to create an image that does not exist in real world, but these images produced by the currently available CG technology still look far from natural. We need a means for free and natural video expression.

Therefore, we attempt to apply CG technologies to the pictures actually shot. Actual pictures are segmented into video components and the components are modified and placed at arbitrary position compositing with other video componets in a virtual 3-D space imbued with the concept of time [14] as following procedures.

(1) Video components: Objects such as a desk, a flower and a human from pictures actually shot are segmented, and managed as video components. At

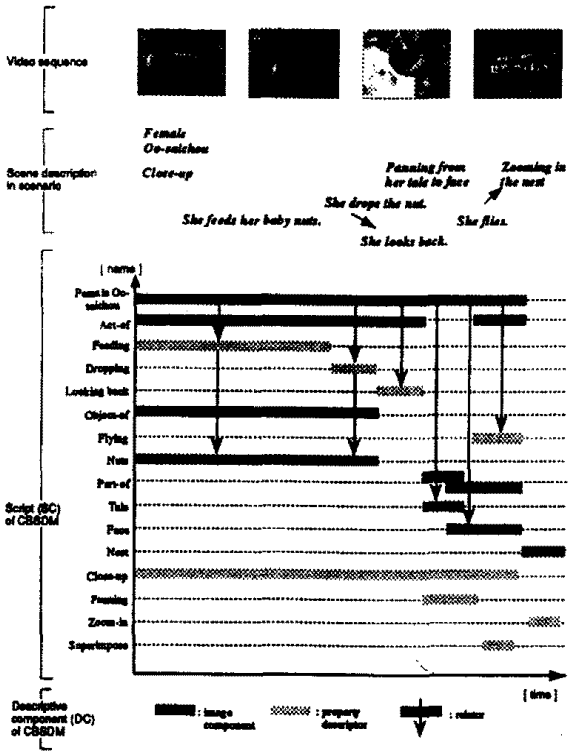

Fig.4 Video sequence and scene description

this time, the 3-D shapes of the objects, lighting conditions, the space where the objects occupy in the picture, and camera work are also taken so that the picture shot from another view points or with different lighting conditions can be produced recombining the video components. These are then stored in the media server together with the video components conditions as attributes data.

Video segmentation can be achieved by shooting an object in front of the background with chromakey blue or by semi-automatically extracting method [15]. In order to produce a picture from any viewing points, we need to acquire these video component images three dimensionally, a task executed by shooting the object with several cameras or taking pictures of the object placed on a turntable.

As a by-product of this research on video components, we have developed a system for appreciating art objects such as pots, sculptures and others with three-dimensions[16]. To construct video component database, we took pictures of art objects from many different angles. A viewer operates the specially designed manipulator to observe an object read from the database on an HDTV monitor from any viewing point. This creates an impression as if the viewer were seeing the object held in his/her hands as shown in Fig.5.

(2) Arrangement in 3-D space: CG places a defined 3-D model freely inside the virtual 3-D space. With the actual picture, the video components obtained above are also arranged in a virtual space. Ideally, this space will be composed of a video memory with a depth-wise address in forming a 3-D space. The problem is that it requires very large memory capacity. So, we constructed a pseudo 3-D space by a video memory having several planes. We have developed an HDTV synthesizer using a 4-plane memory [17]. A picture created by this memory looks something like a stage setting, a 3-D space with a roughly quantized depth axis.

(3) Projection to 2-D picture: With a viewing point fixed, the CG image is then projected onto a screen for conversion into a two-dimensional image. In real shooting, the video components written in the memory are processed in such ways as scrolling, expanding/shrinking, spatial filtering, cover/uncover process depending on the distance from the viewing point and relative positions among video components. Changes can also be made in shading and shadowing depending on the lighting conditions of the composite pictures. Here, we can create natural pictures by utilizing the attributes data of these video components.

With these technologies described in (2) and (3) above, producers can freely manipulate images in the space adding to current video editing process along time axis. We call this new way " spatio-temporal editing." We have developed a virtual camera and a virtual studio using the video components and spatio-temporal editing techniques [18].

(4) Scene description: All the images of CG is created by programming. Similally, we can composite and edit arbitrary video pictures by scripts with scene descriptions. This implies a possibility of precisely

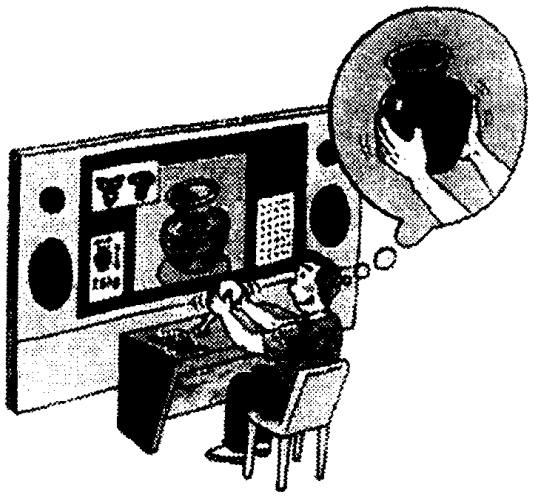

Fig.5 Image of 3D art appreciating system

describing the image or the scene in real pictures.

5. Conclusion

In this paper, we have described the concept of DTPP as a new program production environment. This is a tool that significantly expands the freedom of video/ audio expression by providing an environment for more easy and efficient operation as shown in followings;

(1) Desktop work environment eliminates the need to push the cart loaded with a heap of VTR tapes and go to the edit room and the sub-control room for editing and special effects.

(2) In publishing, DTP lets the user produce a manuscript and make detailed corrections while watching the whole layout, a process significantly more efficient than the conventional method of producing the manuscript first and then editing. Similarly, DTPP provides for more efficient program production with a good overall view. For instance, we can perform program simulations during planning, or produce the frame of complete program first and then shoot necessary scenes to be fitted into it. Moreover, DTPP clarifies work sharing among the staff, enabling parallel processing of the work instead of conventional sequential processing.

(3) Video components and spatio-temporal editting are possible to produce a TV program in a virtual studio of desktop environment. Conventionally, it has been difficult to reuse video materials taken by other producers because of the differences in production intentions. This problem can be overcome with this.

(4) Instead of actually processing the video materials, we can send out the script which describes the method of processing, the order of presentation, and other relevant information. With this, we can easily change the contents and proceed with program production without relying on HDTV, regular television and other media.

(5) The producers can share with each other the video materials, scripts and other information and know-how on program production that have consciously or unconsciously been accumulated in the database. These are used to construction an expart system for program production.

(6) Attributes data can be used not only for many multimedia systems, but it is useful in improving the production efficiency and creating new visual effects. For instance, attributes data based on the story board can be used to perform rough editing by automatically eliminating unnecessary cuts from among the vast amount of video materials.

To realize the DTPP, many research subjects such as networking, human interface, media handling, and high speed processor will have to be solved. We have briefly introduced some of our activities we have been researching to realize the DTPP. For details, we must ask our readers to refer to following references.

References:

(1) K.Enami: "Desktop Program Production - DTPP -", ITE Technical Report, VAI92.32, 1992 (in Japanese)
(2) Obari, Asari: "Non-linear Editing, Transmission System", ITE Technical Report, BPO95.63, 1995 (in Japanese)
(3)K.Enami : "Desk Top Program Production (DTPP) - A Scenario of Studio Digitalization-", 18th International Television Sumposium (June 1993) Montreux
(4)N.Yagi, K.Fukui,K.Enami,et al. :"A Progarammable Video Signal Multi-processor for HDTV Signals", Int. Symp. on Circuit and Systems, 40.9 1993
(5) Ito et al.: "Desktop Video Production by Script Drive", IEICE Spring Conf., 1994, pp262 (in Japanese)
(6)E.Gidney, A.Chandler, G.McFarlane: "CSCW for Film and TV Preproduction", IEEE Multimedia, Summer 1994, pp.16-26
(7)H.Ueda, T.Miyake, and S Yoshizawa, "Impact:An Interactive Natural-Motion Picture Dedicated Multimedia Authoring System", Proc.CHI 91, ACM Press, New York,1991,pp.343-350
(8)Y.Tonomura, A.Akutsu,Y.Taniguchi, and G Suzuki: "Structured Video Computing", IEEE Multimedia Fall 1994, pp.34-43
(9)H.Zhang, A.kankanhalli, and S.Smoliar: "Automatic Partitioning of Full-Motion Video", ACM Multimedia Systems ACM Press, New York, Vol1, 1993, pp.10-28
(10)R.Weiss, A.Duda, and D.K. Gifford: "Composition and Search with a Video Algebra", IEEE Multimedia, Spring 1995, pp.12-25
(11) A.Yanagimachi: "Advanced Broadcasting Service by ISDB and Multimedia", NHK R&D, July 1995, pp.2-13 (in Japanese)
(12)Y.B.Kim, M.Shibata, M.Hayashi:" An Integration of Natural Language and Vision Processing towads an Agient based Future TV system", AAAI94 Workshop on Integration of Natural Language and Vision Processing, pp99-106 (1994)
(13) M.Shibata: "Video Content Model and Application to Video Structuring", IEICE D-II, Vol.J780-II, No.5, pp754-764 (in Japanese)
(14) H.Mitsumine, S.Inoue: "Method to Acquire 3D Video Component Data for DTPP", TV Technical Report, VAI95-10, 1995 (in Japanese)
(15)S.Inoue: "An Object Extraction Method for Image Synthesis", Proc. SPIE Visual Communication and Image Processing, Vol.1606, pp43-54 (1991)
(16)H.Mitsumine, H.Noguchi, K.Enami, et al.: "Solid Object Fine Art Appreciative System", Technical Report of IEICE IE95-78 pp.17-24 (1995-11) (in Japanese)
(17)S.Inoue and M.Shibata: "Spatiotemporal Editing for HDTV Program Production", USENIX Summer 91, pp.95-104
(18) M.Hayashi, Y.Yamauchi, K.Fukui, K.Enami: "Virtual Studio", 8th Human Interface Symposium, pp449-452, 1992 (in Japanese)