T

Teleconferencing

SHERVIN SHIRMOHAMMADI[1], JAUVANE C. DE OLIVEIRA[2]
[1]University of Ottawa, Ottawa, ON, Canada
[2]National Laboratory for Scientific Computation, Petrópolis, Brazil

Synonyms

▶ Audio and video conferencing

Definition

Teleconferencing is an aggregation of audio conferencing, video conferencing, and data conferencing, and includes multiple participants in a live real-time session.

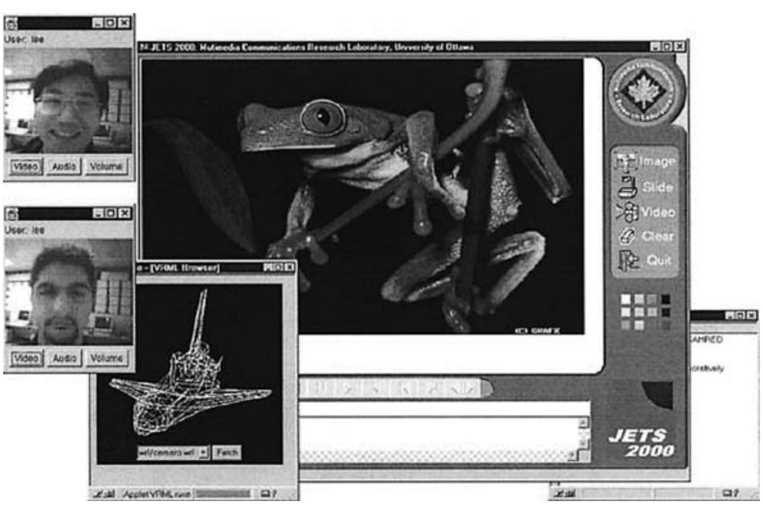
Introduction

Teleconferencing (a.k.a. Multimedia Conference Services [1]) consists of a live real-time session between multiple participants with the ability to hear and see each other as well as share data and applications. Alternatively, teleconferencing can be thought of as an aggregation of *audio conferencing, video conferencing,* and *data conferencing* (or application sharing). Although a subject of interest for many years, teleconferencing has recently grabbed a lot of attention due to current economic and social trends, which emphasize the need for rich media communication between geographically-distributed people. Economic incentives include cutting traveling costs, as well as reducing security risks, and increasing worker availability, whereas social incentives are caused by the higher expectations from technology and experience of ordinary users of today.

Figure 1 shows a sample teleconferencing session in the JETS 2000 system [2]. The participants in this session can see and hear each other, in addition to being able to do whiteboarding and application sharing. It should be noted that teleconferencing is, by nature, a *live* and therefore *real-time* multimedia application. Offline communication paradigms such as blogs, chat boards, and email are not considered to be part of teleconferencing.

Services and Requirements

A typical multimedia teleconferencing system should provide the following services:



Teleconferencing. Figure 1. A teleconferencing session running in the JETS 2000 system.

1. Audio conferencing
2. Video conferencing
3. Data conferencing
4. Control and signaling

Since these services are provided in a live environment, communication lag and deficiencies such as delay, jitter, packet loss, and lack of sufficient bandwidth adversely affects the execution of the teleconferencing session. This is particularly significant in networks that don't guarantee quality of service, such as the Internet.

Audio and video are medium which are continuous by nature. As such, they both suffer from network lag. However, it is a well-known fact that, from a human perception point of view, audio is affected more adversely than video in the presence of network lag. For example, if a given video frame is delayed, one can simply repeat the previous frame until the new one arrives. This causes some "unnaturalness" in the video, but it is acceptable for all practical purposes if the repeat duration is not too long. For audio streaming; however, if audio samples are delayed, they are either replaced by undesirable silence, which become especially irritating when happening in the middle of a word, or they are replaced by the last samples available until new ones arrive, which causes noise. In both cases, the flow of audio not only becomes unnatural, but also the conversation becomes incomprehensible. Under less than desirable network conditions, participants in teleconferencing do experience such problems. In order to mitigate this undesirable phenomenon, it is common to buffer a given amount of audio and video so that there is something available for playback in the event that delays occur. Naturally this buffering introduces further delay and needs to be limited, especially in the context of a live session. To accommodate the transmission of audio and video over the network, protocols such as the RTP (Real-time Transport Protocol) [3] are used.

Teleconferencing Standards

In any field, standards are needed for compatibility: allowing products and services offered by different vendors to interoperate. Teleconferencing is no exception. In fact, the need for standards in teleconferencing is most crucial due to the large number of network operators, product vendors, and service providers.

The International Telecommunications Union – Telecommunications Standardization Sector (ITU-T) [4] has created a large set of recommendations that deal with teleconferencing, encompassing audio, video, data, and signaling requirements. The ITU-T F.702 recommendation describes what is known as Multimedia Conference Services [1]. It defines the terminology used in multimedia conferencing, as well as a description of the service with a functional model, configuration, and roles of participants, terminal aspects, applications, and additional services. There is a set of recommendations which are typically umbrella standards, each containing a number of other recommendations specifically for audio, video, data, and signaling. The most widely-used of such umbrella standards for teleconferencing on the Internet is the ITU-T H.323 [5].

ITU-T H.323 – Packet-Based Multimedia Communications Systems

The H.323 standard provides a foundation for real-time audio, video and/or data communications across packet-based networks, such as the Internet. Support for audio is mandatory, while data and video are optional [5]. By complying with the H.323 recommendation, multimedia products and applications from multiple vendors can interoperate, allowing users to communicate with ensured compatibility. H.323 is specifically designed for multimedia communications services over Packet Based Networks (PBN) which may not provide a guaranteed Quality of Service (QoS), such as the Internet.

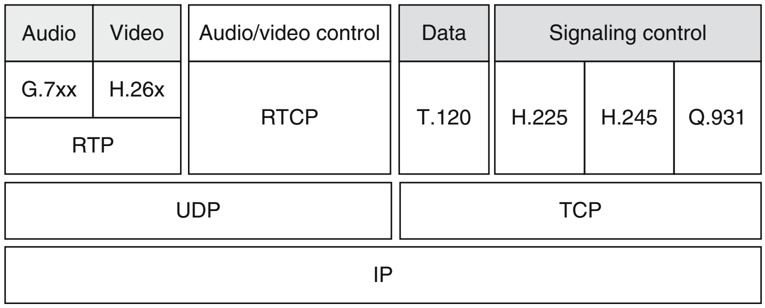
The standard is quite broad in its scope. It encompasses stand-alone devices, embedded devices, point-to-point, and multipoint conferences. It also covers issues such as multimedia management, bandwidth management, and interoperation with various terminal types.

As an umbrella standard, H.323 specifies other standards for audio, video, data, and signaling. Figure 2 shows an example of an H.323 recommendation compliant system, illustrating audio, video, data, and signaling, and their relation with ITU-T standards and the TCP/IP protocol stack.

The specific components are briefly discussed next.

Call Signaling and Control

Call signaling and control is concerned mostly with setting up, maintaining, and taking down connections between teleconferencing parties. In addition, the recommendations define a *Gateway*: an optional element that provides many services, mainly used for translating between H.323 conferencing endpoints and other terminal types. This includes translation



Teleconferencing. Figure 2. An example configuration of ITU-T H.323.

between transmission formats (H.225.0 to H.221) and between communications procedures (H.245 to H.242 [6]). The Gateway also translates between audio and video codecs and performs call setup and clearing on both the packet-based and the switched-circuit network side. Some of the relevant standards for signaling and control are:

1. H.225.0 – Call signaling protocols and media stream packetization for packet-based multimedia communication systems [7]
2. H.235 – Security and encryption for H-series (H.323 and other H.245-based) multimedia terminals [8]
3. H.245 – Control protocol for multimedia communication [9]
4. Q.931 – ISDN user-network interface layer 3 specification for basic call control [10]
5. H.450.1 – Generic functional protocol for the support of supplementary services in H.323 [11]

Audio

Audio signals transmitted over the network are digitized and, most of the time, compressed. H.323 supports a number of compression algorithms but only one is mandatory: H.323 terminals must support the G.711 recommendation [12] for speech coding, which is basically uncompressed 8-bit PCM signal at 8 KHz in either A-Law or μ-Law format, leading to bitrates of 56 or 64 Kbps. Support for other ITU-T audio recommendations and compressions is optional, and is implementation specific depending on the required speech quality, bit rate, computational power, and delay. Provisions for asymmetric operation of audio codecs have also been made; i.e., it is possible to send audio using one codec but receive audio using another codec. If the G.723.1 [13] audio compression standard is provided, the terminal must be able to encode and decode at both the 5.3 Kbps and the 6.3 Kbps modes. If a terminal is audio only, it should also support the ITU-T G.729 recommendation [14]. Note that if a terminals is known to be on a low-bandwidth network (<64 Kbps), it does not need to disclose capability to receive G.711 audio since it won't practically be able to do so. The relevant H.323 audio codecs are:

1. G.711 – Pulse Code Modulation (PCM) of voice frequencies [12]
2. G.722 – 7 kHz audio-coding within 64 Kbps [15]
3. G.723.1 – Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 Kbps [13]
4. G.728 – Coding of speech at 16 Kbps using low-delay code excited linear prediction [16]
5. G.729 – Coding of speech at 8 Kbps using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP) [14]

Video

Video support in H.323 is optional. However, if a terminal is to support video, it must at the very least support the H.261 [17] codec at the QCIF frame format. Support for other H.261 modes or the H.263 [18] codec is optional. During the initial setup, a specific video data rate is selected during the capability exchange. This rate should not be violated throughout the duration of the session. The H.261 standard uses communication channels that are multiples of 64 Kbps, known as p × 64, where p∈{1, 2, 3, ... 30]. From a video encoding perspective, there are no Bidirectional or "*B*" frames in H.261. Instead, it uses Intra or "*I*" frames which are fully and independently encoded, and Predicted or "*P*" frames which code the difference between the frame and its previous frame by using motion estimation.

Compared to H.261, H.263 uses 1/2 pixel motion-estimation for better picture quality, and a Huffman

Teleconferencing. Table 1. ITU-T standard frame formats for H.261 and H.263

Picture format	Frame size	H.261	H.263
SQCIF	128 × 96	not specified*	required
QCIF	176 × 144	Required	required
CIF	352 × 288	Optional	optional
4CIF	704 × 576	N/A	optional
16CIF	1408 × 1152	N/A	optional

coding table that is optimized specifically for low bit rate transmissions. H.263 defines more picture modes than H.261, as seen in Table 1. In addition, H.263 introduces the PB frames, which consist of a P frame interpolated with a Bi-directional or "*B*" frame: a frame that depends not only on a previous frame but also on a forthcoming frame. Similar to a P frame, a B frame uses motion estimation to reduce the amount of information to carry. See the related article on "Video Compression and Coding Techniques" and "Motion Estimation" for further details.

Data Conferencing

The ITU-T has a set of standards for data conferencing and application sharing. The ITU-T T.120 recommendation [19] summarizes the relationships amongst a set of protocols for data conferencing, providing real-time communication between two or more entities in a conference. Applications specified as part of the T.120 family include application sharing, electronic whiteboarding, file exchange, and chat.

Data conferencing is an optional capability in multimedia conferences. When supported, data conferencing enables collaboration through applications such as whiteboards, application sharing, and file transfer. The list below summarizes the related recommendations in the T.120 family set forth herein:

1. T.120 – Data protocols for multimedia conferencing [19]
2. T.121 – Generic application template [20]: It defines the generic application template (GAT), specifying guidelines for building application protocols and facilities that manage the control of the resources used by the application. T.121 is a mandatory standard for products that support T.120.
3. T.122 – Multipoint communication service – Service definition [21]: It defines the multipoint services allowing one or more participants to send data. The actual mechanism for transporting the data is defined by T.125, T.122 and T.125 together constitute the T.120 multipoint communication services (MCS).
4. T.123 – Network-specific data protocol stacks for multimedia conferencing [22]: It defines the sequencing and transporting of the data and its flow control across the network.
5. T.124 – Generic Conference Control [23]: It defines the generic conference control (GCC) for initiating and maintaining multipoint data conferences. The lists of conference participants, their applications, and the latest conference information is kept here.
6. T.125 – Multipoint communication service protocol specification [24]: It defines how data is actually transmitted. One can think of T.125 as the implementation of T.122 services, among other things.
7. T.126 – Multipoint still image and annotation protocol [25]: It defines how the whiteboard application sends and receives data. Both compressed and uncompressed form for viewing and updating are supported.
8. T.127 – Multipoint binary file transfer protocol [26]: It defines how files are transferred, sometimes simultaneously, among users. Similar to T.126, it supports both compressed and uncompressed forms.
9. T.128 – Multipoint application sharing [27]: It defines the program sharing protocol; i.e., how participants can share programs that are running locally.

Other Standards

H.320 *Narrow-band visual telephone systems and terminal equipment* (March 2004) [28] supports videoconferencing over ISDN. This protocol has a long and successful history. Sometimes considered a "legacy" protocol, private industry still relies heavily on this protocol, and it provides an important bridge to the PSTN.

H.321 *Adaptation of H.320 visual telephone terminals to B-ISDN environments* (Feb 98) [29] provides support for videoconferencing over ATM. A number of successful systems have been built upon this technology, though the general scarce deployment of ATM limits the overall reach of these systems.

H.322 *Visual telephone systems and terminal equipment for local area networks which provide a guaranteed quality of service* (Mar 96) [30] used over LANs that guarantee bandwidth, such as ISO-Ethernet.

H.324 *Terminal for low bit-rate multimedia communication* (Mar 2002) [31] provides support for low bandwidth videoconferencing over PSTN. This protocol enjoys success particularly in Asian cellular markets.

Tools

There are many teleconferencing tools currently available. Some of the most popular commercial tools are Microsoft's NetMeeting (http://www.microsoft.com/netmeeting/), CU-SeeMe (http://www.cuseeme.com/), ICUII (http://www.icuii.com/), and Isabel developed at the Universidad Politécnica de Madrid (http://isabel.dit.upm.es/). In terms of open-source solutions, OpenH323 (http://www.openh323.org/) is a widely used platform that implements the H.323 standard.

Recording a Teleconference Session

The ability to record a teleconferencing session is an important requirement for many applications. In many cases, it is necessary to play back the events that took place in a session. For example, when a participant misses a teleconferencing meeting, he/she can play back exactly what happened in the meeting, if the session was recorded. This includes conversations and discussions among participants, as well as applications and documents that were shared. Ideally, this recording should be done in a transparent way; i.e., user applications need not be modified for this recording to take place. One approach to achieve this would be for the "recorder" module to join the session as a regular client and observe what is taking place and record it. This is also referred to as non-intrusive recording. The J-VCR system is an example of a non-intrusive teleconferencing recording tool [32].

Cross-References

▶ Audio Conferencing
▶ Data Conferencing
▶ Recording Teleconference Sessions
▶ Video Conferencing

References

1. International Telecommunication Union, Telecommunication Standardization Sector F.702 Recommendation – Multimedia Conference Services, July 1996.
2. S. Shirmohammadi, J.C. Oliveira, and N.D. Georganas, "Applet-Based Telecollaboration: A Network-Centric Approach," IEEE Multimedia, Vol. 5, No. 2, April–June 1998, pp. 64–73.
3. H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications," IETF RFC 1889, January 1996.
4. International Telecommunication Union, Telecommunication Standardization Sector, http://www.itu.int/ITU-T.
5. International Telecommunication Union, Telecommunication Standardization Sector H.323 Recommendation – Packet-based multimedia communications systems, July 2003.
6. International Telecommunication Union, Telecommunication Standardization Sector H.242 Recommendation – System for establishing communication between audiovisual terminals using digital channels up to 2 Mbit/s, March 2004.
7. International Telecommunication Union, Telecommunication Standardization Sector H.225.0 Recommendation – Call signaling protocols and media stream packetization for packet-based multimedia communication systems, July 2003.
8. International Telecommunication Union, Telecommunication Standardization Sector H.235 Recommendation – Security and encryption for H-series (H.323 and other H.245-based) multimedia terminals, August 2003.
9. International Telecommunication Union, Telecommunication Standardization Sector H.245 Recommendation – Control protocol for multimedia communication, January 2005.
10. International Telecommunication Union, Telecommunication Standardization Sector Q.931 Recommendation – ISDN user-network interface layer 3 specification for basic call control, May 1998.
11. International Telecommunication Union, Telecommunication Standardization Sector H.450.1 Recommendation – Generic functional protocol for the support of supplementary services in H.323, February 1998.
12. International Telecommunication Union, Telecommunication Standardization Sector G.711 Recommendation – Pulse code modulation (PCM) of voice frequencies, November 1988.
13. International Telecommunication Union, Telecommunication Standardization Sector G.723.1 Recommendation – Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 Kbit/s, March 1996.
14. International Telecommunication Union, Telecommunication Standardization Sector G.729 Recommendation – Coding of speech at 8 Kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP), March 1996.
15. International Telecommunication Union, Telecommunication Standardization Sector G.722 Recommendation – 7 KHz audio-coding within 64 Kbit/s, November 1988.
16. International Telecommunication Union, Telecommunication Standardization Sector G.728 Recommendation – Coding of speech at 16 Kbit/s using low-delay code excited linear prediction, September 1992.
17. International Telecommunication Union, Telecommunication Standardization Sector H.261 Recommendation – Video codec for audiovisual services at p $\times$ 64 Kbit/s, March 1993.
18. International Telecommunication Union, Telecommunication Standardization Sector H.263 Recommendation – Video coding for low bit rate communication, January 2005.
19. International Telecommunication Union, Telecommunication Standardization Sector T.120 Recommendation – Data protocols for multimedia conferencing, July 1996.

20. International Telecommunication Union, Telecommunication Standardization Sector T.121 Recommendation – July 1996.
21. International Telecommunication Union, Telecommunication Standardization Sector T.122 Recommendation – Multipoint communication service-Service definition, February 1998.
22. International Telecommunication Union, Telecommunication Standardization Sector T.123 Recommendation – Network-specific data protocol stacks for multimedia conferencing, May 1999.
23. International Telecommunication Union, Telecommunication Standardization Sector T.124 Recommendation – Generic Conference Control, February 1998.
24. International Telecommunication Union, Telecommunication Standardization Sector T.125 Recommendation – Multipoint communication service protocol specification, February 1998.
25. International Telecommunication Union, Telecommunication Standardization Sector T.126 Recommendation – Multipoint still image and annotation protocol, July 1997.
26. International Telecommunication Union, Telecommunication Standardization Sector T.127 Recommendation – Multipoint binary file transfer protocol, August 1995.
27. International Telecommunication Union, Telecommunication Standardization Sector T.128 Recommendation – Multipoint application sharing, February 1998.
28. International Telecommunication Union, Telecommunication Standardization Sector H.320 Recommendation – Narrow-band visual telephone systems and terminal equipment, March 2004.
29. International Telecommunication Union, Telecommunication Standardization Sector H.321 Recommendation – Adaptation of H.320 visual telephone terminals to B-ISDN environments, February 1998.
30. International Telecommunication Union, Telecommunication Standardization Sector H.322 Recommendation – Visual telephone systems and terminal equipment for local area networks which provide a guaranteed quality of service, March 1996.
31. International Telecommunication Union, Telecommunication Standardization Sector H.324 Recommendation – Terminal for low bit-rate multimedia communication, March 2002.
32. S. Shirmohammadi, L. Ding, and N.D. Georganas, "An Approach for Recording Multimedia Collaborative Sessions: Design and Implementation," Journal of Multimedia Tools and Applications, Vol. 19, No. 2, 2003, pp. 135–154.

Tele-Haptics

Xiaojun Shen, Shervin Shirmohammadi
University of Ottawa, Ottawa, ON, Canada

Synonyms

► Transmission of computer generated tactile sensations

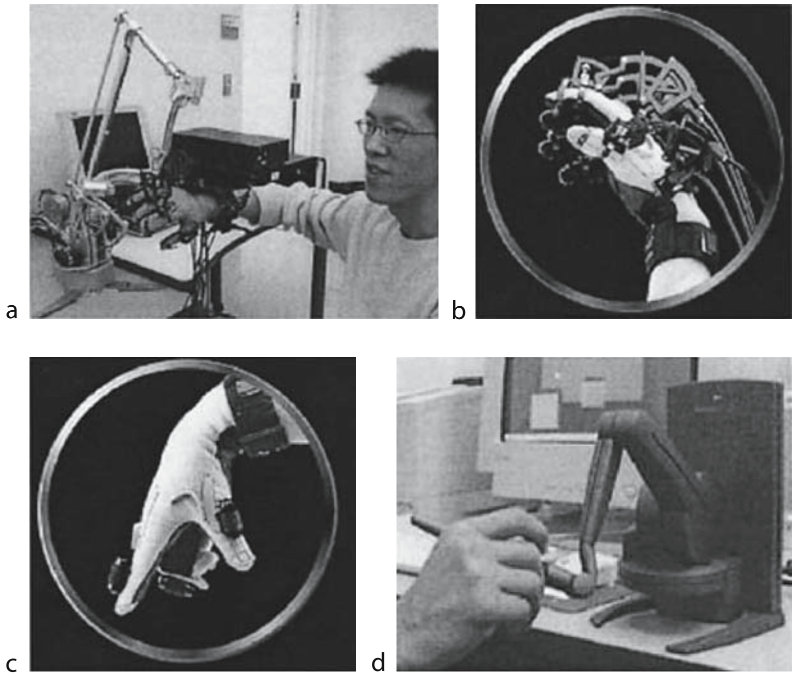
Definition

Tele-haptics refers to transmission of computer generated tactile sensations over the network.

Introduction

Multimedia and Information technology is reaching limits in terms of what can be done in multimedia applications with only sight and sound. The next critical step is to bring the sense of "touch" over network connections, which is known as *tele-haptics. Haptics*, a term which was derived from the Greek verb "haptesthai" meaning "to touch," introduces the sense of touch and force in human-computer interaction. Haptics enable the human operator to manipulate objects in the environment in a natural and effective way, enhance the sensation of "presence," and provide information such as stiffness and texture of objects, which cannot be described completely with visual or audio feedback only. The technology has already been explored in contexts as diverse as modeling and animation, geophysical analysis, dentistry training, virtual museums, assembly planning, mine design, surgical simulation, design evaluation, control of scientific instruments, and robotic simulation. But its true potential in these areas has not yet been achieved, and its application to all aspects of dexterous training, for example is still untapped. Haptic devices typically consist of microcontrollers that use input information from sensors, and control effectors to create human sensations as outputs. Sensors range from pressure, temperature and kinesthetic sensing devices, to biofeedback equipment. Haptic effectors, evoking precise perceivable sensations, range from small motors, fans, heating elements, or vibrators; to micro-voltage electrodes which gently stimulate areas of the skin (creating subtle, localized, "tingling" sensations). Some Haptic devices are shown in Fig. 1.

Compared with visual and audio displays, haptic displays are more difficult to build since the haptic system not only senses the physical world but also affects it. In contrast, we can not change the object while listening or just looking at it except in a magical world. In the haptic device world, tactile haptic devices provide sense of sensing the temperature and pressure, grasping, and feeling the surface textures in details of an object, while force haptic devices provide users with the sense of general shape, coarse texture and the position of an object. Apparently, tactile haptic device is more difficult to create than that of kinesthetic. Due to technical problems, currently there is no single haptic device that could combine these two sensations together although we cannot separate them in the real life.

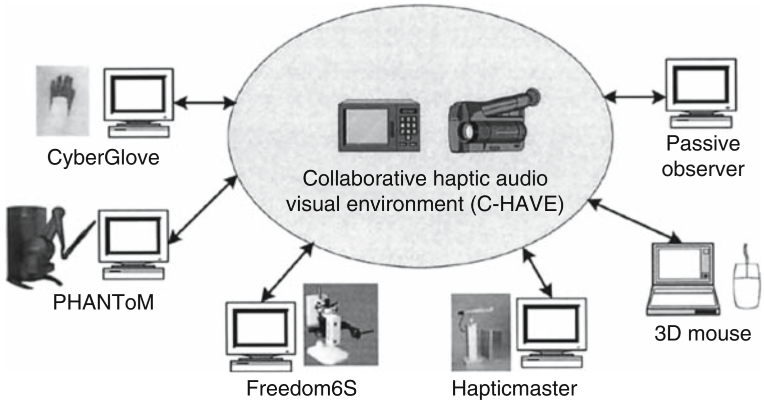


Tele-Haptics. Figure 1. Some haptic devices: (a) Immersion's CyberForce (b) Immersion's CyberGrasp (c) Immersion's CyberTouch (d) SensAble's PHANToM desktop.

Tele-haptics can be defined as the use of haptics in a network context; it is the science of transmitting computer generated tactile sensations over networks, between physically distant users. Applications for haptics are broad, and tele-haptics in no way constrains that potential. For every single-user desktop application of haptics technology, a tele-haptics equivalent introduces, at the very least, a richer collaborative experience. A desktop modeling package can extend to collaborative modeling or design review; a dexterous task trainer can extend to remote teaching and assessment; a surgical simulator can be coupled directly to a surgical robot.

Much of the academic research and development on tele-haptics in recent years has attempted to ease the construction of virtual environments providing increased immersion through multi-sensory feedback. These environments that support collaborative touch in virtual environments are termed *Collaborative Haptic Audio Visual Environments* (C-HAVE) (Fig. 2), where participants may have different kinds of haptic devices, such as the Sensable's PHANToM (http://www.sensable.com/), the MPB Freedom6S Hand Controller (http://www.mpb-technologies.ca/), FCS Robotics' HapticMASTER (http://www.fcs-cs.com/), or Immersion's CyberGrasp (http://www.immersion.com/); or they can just be passive observers. Some of the participants in the virtual world may only provide virtual objects as a service to the remaining users. Adding haptics to a conventional *Collaborative Virtual Environment* (CVE) creates additional demand for frequent position sampling, collision detection/response, and fast update. It is also reasonable to assume that in CVEs, there may be a heterogeneous assortment of haptics devices with which users interact with the system.

User scenarios for C-HAVE evolve from haptic VE applications. Collaboration may involve independent or dependent manipulation of virtual objects, or tele-mentoring. Independent manipulation of virtual objects allows multiple participants to haptically interact with separate objects. Each object is "owned" by a single participant. While other participants may feel the owner's manipulation of an object, in a similar fashion to a virtual audience watching virtual actors in a conventional CVE, where the owner does not receive haptic feedback from those other participants. This is termed *unilateral tele-haptics*. Dependent manipulation introduces *bilateral tele-haptics*, whereby multiple participants haptically interact with identical



Tele-Haptics. Figure 2. C-HAVE: Collaborative haptic audio visual environment.

or coupled objects. Here, each participant feels the other's manipulation of the environment, but that sensation is indirectly perceived through the environmental modification. *Tele-mentoring* allows direct coupling of haptic devices over a network. Tele-mentoring is an educational technique that involves real-time guidance of a less experienced trainee through a procedure in which he or she has limited experience. A typical example of this is the education of surgeons, whereby a trainee can feel an expert's guiding hand. Independent manipulation of virtual environments involves augmentation of a client station and may be viewed as a simple integration of conventional CVEs with conventional hapto-visual VEs. Both application families are typically implemented upon none or soft real time operating systems. By contrast, dependent manipulation and tele-mentoring impose more stringent requirements and, like tele-robotics, demand hard real time guarantees. Since Tele-Haptics are a networked extension of Haptics, let us start by having a more detailed look at Haptics and their applications.

Haptics Applications

Haptics is the study about the simulation of the touch modality and the related sensory feedback. Haptics applications are often termed "hapto-visual" or "hapto-visual-acoustic applications," in addition to the more generic "multi-sensory" applications.

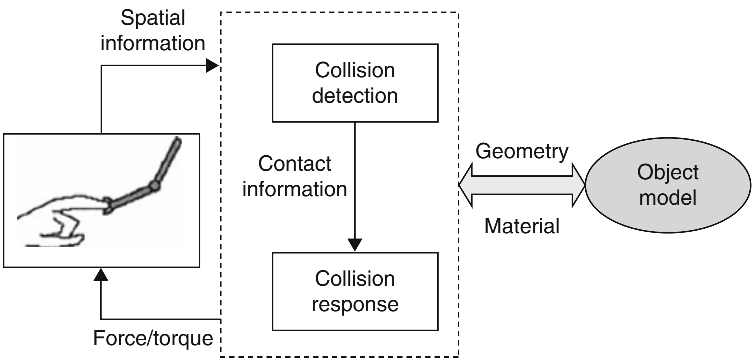
Generally there are two kinds of sensory information related with human touch: *kinesthetic* and *tactile*. Kinesthetic is the human's perception on the relative movements among body parts and is determined by the velocity of movement of the links, therefore the touched object can be reconstructed in the mind by knowing the position of the corresponding links. Tactile is the sense of touch that comes from some sensitive nerve sensors on the surface of skin such as information about pressure and temperature. Haptic perception of touch involves both kinds of sensations.

Haptic Rendering

Haptic rendering is the process of converting haptic interface spatial information (position/orientation) to net force and torque feedback (some haptic devices may only generate force feedback to the users). Two major tasks in haptic rendering paradigms are *collision detection* and *collision response* (Fig. 3). Collision detection is to detect collisions between the end point of the generic probe and the objects in the scene, while collision response is to respond to the detection of collision in terms of how the forces reflected to the user are computed.

Stable Haptic Interaction

A haptic device links a human operator with a virtual environment together where the operator feels the objects in the virtual scene with the sense of touch. There is physical energy flowing to and from the operator. Since in the haptic interaction, the physical haptic device generates force feedback, instability of system can damage the hardware or hurt the operator physically. For example, the Haptic MASTER from FCS Robotics can easily generate a few hundred Newton's force in less than one second which is enough to hurt a person seriously. Instability also destroys the illusion of a real object immediately. Furthermore, the instabilities degrade transparency in haptic simulation. A transparent haptic interface should be able to emulate



Tele-Haptics. Figure 3. Haptic rendering.

any environment, from free space to infinitely stiff obstacles.

Many researchers have studied the stability issues in haptic simulation in different ways. In a haptic simulation, there are at least three components: human operator, a haptic interface, and a computer model of the virtual environment. Figure 4 shows their relationship using the two port network model. The star superscript indicates the variable is discrete. Typically the haptic simulation can be classified into two categories: impedance control and admittance control. Impedance control means "measure position and display force" while admittance control means "measure force and display motion or position." All these three components affect the stability of the system.

From the perspective of control theory, Minsky et al. [1] derived an expression for the guaranteed stable haptic interaction using the impedance control theory and then modified it with the experimental results. The impedance here is used to describe a generalized relationship between force and motion of the haptic device. They noted a critical tradeoff between sampling rate, virtual wall stiffness, and device viscosity and analyzed the role of the human operator in stability concern. In the experiment, they tried to simulate a virtual wall with stiffness K and viscosity B. The derived equation for a guaranteed stability is given as follows:

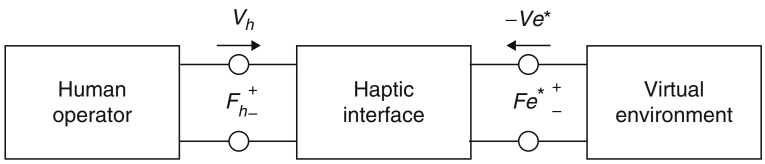
$$B + b > \frac{KT}{2}$$

where T is sampling time and b is the device viscosity. The results shows that the rate to sample the haptic spatial information and output the force to haptic device is the most crucial factor that affects the highest stiffness can be achieved while maintain a stable system. At the same time, the device viscosity plays a role in the stability issue: the higher the device viscosity, the higher stiffness can be achieved. So it is possible to make the system stable by adding extra viscosity, or by reducing the stiffness of the simulated hard surface. But the human will feel resistance and sluggishness even in free space if the viscosity is too high.

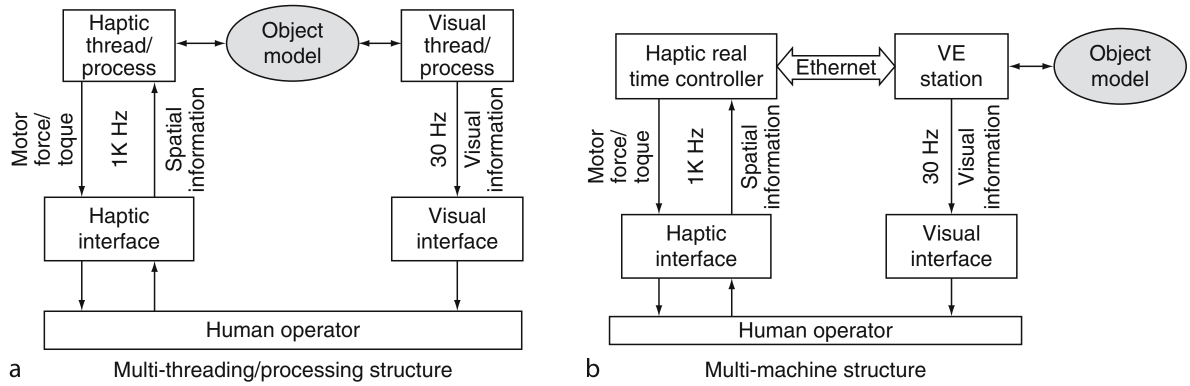
Architecture of Haptics Applications

Multiple tasks, such as haptic sensing/actuation and visual updates must be accomplished in a synchronized manner in haptic applications. It becomes commonplace to separate tasks into computational threads or processes, to accommodate different update rates, distribute computation load and optimize computation. Conventionally, multithreading and multiprocessing software architectures (Fig. 5(a)) are applied to develop effective multimodal VEs and the optimal usage of the CPU capabilities. However, an important but little discussed consequence of the conventional architectures is that it makes the operating system an inherent component of the applications, with operating system scheduling algorithms limiting the application's quality of service. The application may request a theoretical rate of force display but it is the scheduler that determines the actual rate. This scheduler is itself a complex algorithm, particularly when considered in terms of its interactions with the other services provided by the operating system.

A multi-machine solution for haptic application was addressed in [2,3] as shown in Fig. 5(b). The multi-machine architecture is comprised of three parts: haptic device, Haptic Real Time Controller (HRTC) and Virtual Environment (VE) graphics



Tele-Haptics. Figure 4. Haptic simulation diagram.



Tele-Haptics. Figure 5. Architectures for haptics applications.

station. HRTC communicates with its VE station through a local Ethernet connection. HRTC relies on hard real time operating systems (e.g., QNX Neutrino, VxWorks, or Windows CE) to guarantee the stability of the control loop. The separation of functionalities of haptic and graphic rendering makes this architecture easier to extend to existing applications. Unlike conventional multithreading or multiprocessing approaches for haptics, this multi-machine model solution applies a hard real-time operating system for haptic control, while applying a mainstream OS such as Win2K or WinXP for the application and graphics.

Tele-Haptics – Networked Haptics

The concept of networking haptics is referred to as "tele-haptics" and occurs when haptic devices are located at remote locations over a communications network. Sometimes referred to as *e-touch*, tele-haptics is based on the bilateral transmission of spatial information such that either end of the communication can both sense and/or impart forces. One of the stumbling blocks to enabling real-time control of haptic devices over a network is network lag: delay, jitter, and packet loss. Haptics can actually cause greater time delay issues due to instability. In other words, the mechanical devices used to impart the sense of touch could vibrate and become dangerous. For example, with Internet time delays as little as 100 ms, it is possible for on-line simulated training participants to be essentially involved in two totally different scenarios. In the case where a transcontinental latency exists, there will be a mismatch between the two participant's computers due to network time delays. If this were to be used for mission-critical simulations for pilots, for example, the effects render the training process useless.

There are, however, solutions to mitigating network lag issues [4]. These solutions, typically implemented over the transport layer (UDP), consist of communication techniques that quickly synchronize both ends in the presence of packet loss [5], or reduce the jitter by short-term buffering [4]. Another type of solution, which deals with the problem from a human-computer interaction point of view as opposed to a networking point of view, uses "decorators": visual cues that inform the user about the presence of network lag, allowing him/her to intuitively adjust to the situation accordingly [6]. Another concept that has been around for many years is "*dead reckoning*." This concept has proven somewhat useful in the past, but is very susceptible to noise and becomes ineffective as time delay becomes large. Given that telehaptics is very sensitive to time delay, more advanced techniques need to be adopted to enable telehaptic applications to operate in a real-time environment, such as the Handshake (http://www.handshakeVR.com) method of time delay compensation. This method utilizes advanced intelligent predictive modeling techniques for real time human-in-the-loop interaction. It effectively overcomes network instability issues and provides superior tracking performance as compared to dead reckoning techniques to allow a user to interact in real-time with his/her environment.

Generic Architecture for C-HAVE

Tele-haptic platforms rely on hard real time operating systems to guarantee the stability of the control loop and to minimize the delays in the network stack. As described below, the Haptic Real Time Controller compensates for network latency, so increased or unreliable latency in the network stack provided by the host operating system will either increase the complexity of the latency compensation algorithms or decrease the effective separation of tele-haptic collaborators.

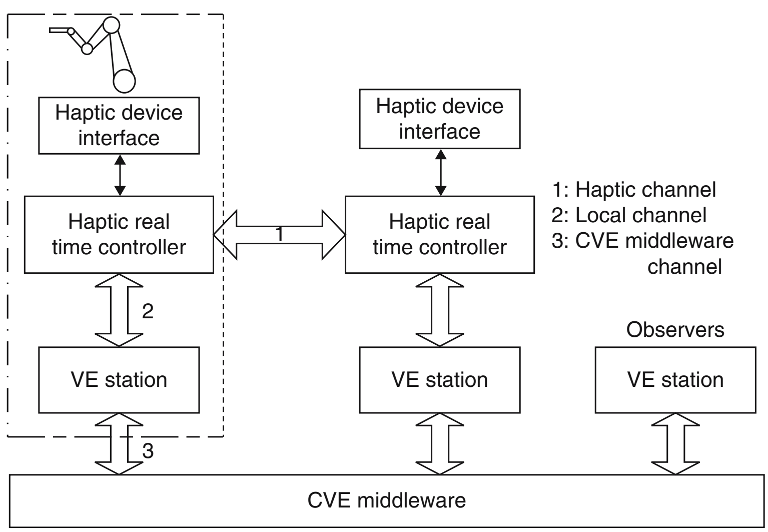
A node in a C-HAVE environment with a haptic device is comprised of three parts: haptic device, HRTC, and Virtual Environment (VE) graphics station, as illustrated in Fig. 6. The HRTC controls the haptic device through a device driver. A local HRTC is linked with remote HRTC(s) through a haptic channel (1 in Fig. 6) and communicates with its VE station through a local channel (2 in Fig. 6). The VE graphics stations are interconnected over CVE Middleware such as HLA/RTI (3 in Fig. 6).

Tele-Haptics Platform – Haptic Real Time Controller

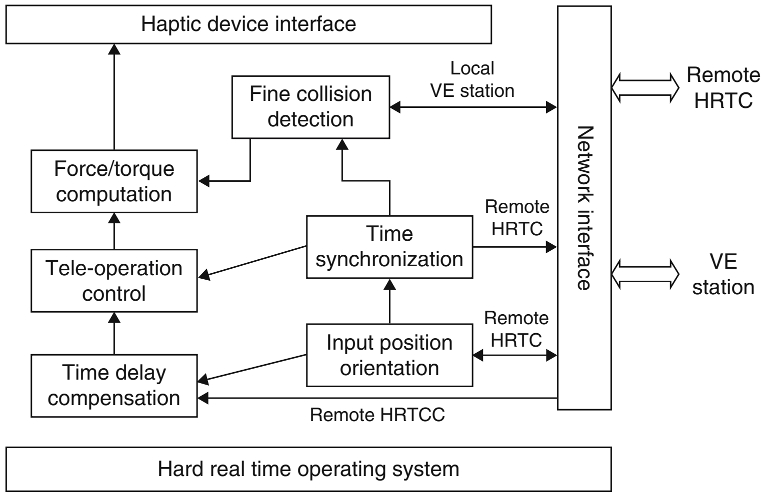
The Haptic Real Time Controller (HRTC), shown in Fig. 7, in a C-HAVE system provides a modular and flexible platform to allow real time control of a haptic device or application device at one node of a C-HAVE network by a haptic device or application hardware at another node of C-HAVE, which is referred to as client/server mode. It is not limited to a client-server configuration, but is modular enough to support multiple clients or other network configurations.

To enable this technology, it is required that the clocks on both nodes be synchronized closely, have precise sampling periods and have the ability to perform complex control computations so that desired performance can be achieved at the client end. An important component of the HRTC is the use of hardware/software solutions to accomplish this synchronization and precision in timing. In order to enable real time control and data exchange, the software of the HRTC runs on a modular and robust real time operating system.

The time-varying network delay degrades the performance of many network applications, and can cause instability in applications involving bilateral control of haptic devices. HRTC includes sophisticated algorithms to compensate for the time-varying network delays. In essence, if the compensation were perfect, the time delays would be transparent to the system. The HRTC allows the time delay compensation to vary between full delay compensation which gives higher performance but may introduce more noise and



Tele-Haptics. Figure 6. Generic architecture for C-HAVE.



Tele-Haptics. Figure 7. Tele-haptics platform – Haptic real time controller.

overshoot, to partial delay compensation, to no compensation which results in low performance and instability. The time delay compensation in the HRTC does not require a mathematic model of environment, and is robust to uncertainties due to the human-in-the-loop situation.

Cross-References

► E-touch
► Haptics
► Haptic Devices

References

1. M. Minsky, M. Ouh-Young, O. Steele, F.P. Brooks, and M. Behensky, "Feeling and Seeing Issues in Force Display," Computer Graphics, Vol. 24, No. 2, pp. 235–243, 1990.
2. X. Shen, F. Bogsanyi, L. Ni, and N.D. Georganas, "A Heterogeneous Scalable Architecture for Collaborative Haptics Environments," Proceedings of the IEEE Workshop on Haptic, Audio and Visual Environments and their Applications (IEEE HAVE' 03), Ottawa, Canada, September 2003, pp. 113–118.
3. X. Shen, J. Zhou, A. El Saddik, and N.D. Georganas, "Architecture and Evaluation of Tele-Haptic Environments," Proceedings of the IEEE International Symposium on Distributed Simulation and Real Time Applications (IEEE DS-RT' 04), Budapest, Hungary, October 2004, pp. 53–60.
4. S. Shirmohammadi, N.H. Woo, and S. Alavi, "Network Lag Mitigation Methods in Collaborative Distributed Simulations," Proceedings of the International Symposium on Collaborative Technologies and Systems (CTS' 05), St. Louis, MO, May 2005.
5. S. Shirmohammadi and N.D. Georganas, "An End-to-End Communication Architecture for Collaborative Virtual Environments," Computer Networks Journal, Vol. 35, No. 2–3, February 2001, pp. 351–367.
6. S. Shirmohammadi and N.H. Woo, "Evaluating Decorators for Haptic Collaboration over Internet," Proceedings of the IEEE Workshop on Haptic Audio Visual Environments and their Applications (IEEE HAVE' 04), Ottawa, Canada, October 2004, pp. 105–110.
7. M.A. Srinivasan, and C. Basdogan, "Haptics in Virtual Environments: Taxonomy, Research Status, and Challenges (PDF)," Computers and Graphics, (Special issue on Haptic Displays in Virtual Environments), Vol. 21, No. 4, pp. 393–404.

Tele-Operation

Definition

The objective of tele-operation is to allow the user to interact with and operate equipment and devices in a remote location without actually being in that location.

Tele-operation can be thought of as a special case of telepresence. The objective of tele-operation is to allow the user, referred to as the "operator," to interact with and operate equipment and devices in a remote location without actually being in that location. The term itself literally illustrates two layers of implications: the users are virtually "operating" at the remote site, and that the operation is "tele" i.e., distant. The implications lead to two essential components in such systems: *location* and *communication*. Locations indicate the locations in which the users are virtually operating. Standard locations vary from conventional VR to augmented reality (AR) and to real world environments.

The interactivity between host and remote locations is accomplished via networking and multimedia communications. It is generally accepted that there are three independent aspects of presence which contribute to achieving a sense of remote telepresence [1]:

1. The extent of the sensory information: It is necessary to provide the operator with the same level of sensory information that they would receive if they were actually present at the remote site.
2. Control of sensors: It must be possible to move the sensing devices around the environment as if the operator was at the remote site.
3. The ability to modify the remote site and actively interact with it.

Cross-References

▶ Telepresence
▶ Virtual and Augmented Reality

References

1. J. Pretlove and R. Asbery, "Get a head in virtual reality," IEE Colloquium on Control of Remotely Operated Systems: Teleassistance and Telepresence, 12 May 1995, pp. 8/1–8/5.

Telepresence

Xiaojun Shen, Shervin Shirmohammadi
University of Ottawa, Ottawa, ON, Canada

Synonyms

▶ Virtual presence

Definition

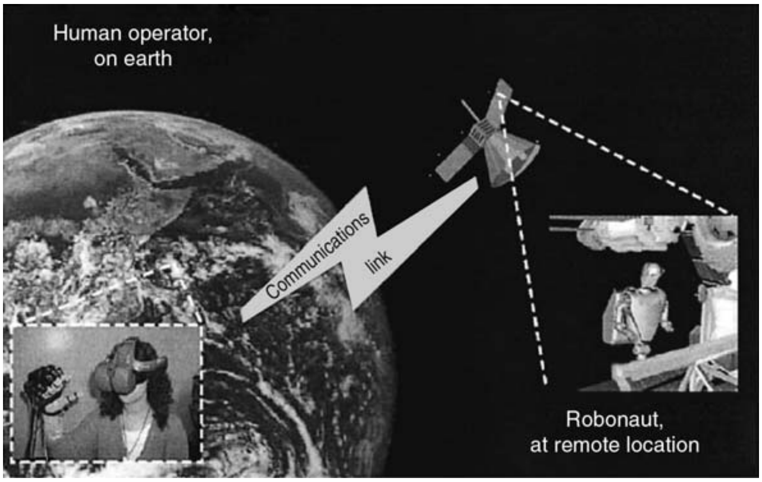
Telepresence, also called virtual presence, is a technique to create a sense of physical presence at a remote location using necessary multimedia such as sound, vision, and touch.

Introduction

The concept of *Telepresence* may come from science fiction since science fiction writers have described conceptual versions of virtual reality (VR) and telepresence for decades [1]. The term "telepresence" was coined by Marvin Minsky in 1980 in reference to teleoperation systems for manipulating of remote physical objects [2]. According to Witmer and Singer "(Tele) presence is defined as the subjective experience of being in one place or environment, even when one is physically situated in another" [3]. Telepresence, also called *virtual presence*, is basically a technique to create a sense of physical presence at a remote location using necessary multimedia such as sound, vision, and touch. This is a networked paradigm by nature, and multimedia communications is used for the transport of information between the user and the remote site. The sense of presence is achieved by generating sensory stimulus so that the operator has an illusion of being present at a location distant from the location of physical presence. Unlike conventional concept of VR that is defined as a new medium to experience presence in a virtual space, telepresence emphasizes the interaction between people and remote site and enables people who are immersed in telepresence systems to act and receive input as if they were at the remote site. Its examples cover applications from the fields of telemedicine to remote surveillance, entertainment, education and other applications in situations where humans are exposed to hazardous and hostile environments. Both telephone and video-conferencing allow for limited types of telepresence since people's voice or video can be there without the person being physically there, while some forms of VR may enable a greater range of interactions. Telepresence systems that incorporate successful force-feedback components can be extremely helpful for a variety of applications [4], for example, space operations such as Robonaut (http://vesuvius.jsc.nasa.gov/er_er/html/robonaut/robonaut.html), the robotic astronaut that is now in development. With the benefits of telepresence system, Robonaut can be deployed in difficult or hazardous situations with the replacement of an astronaut, while its operator, safely housed on a nearby spaceship or even on the earth, controls its actions in a fluid, intuitive fashion. Figure 1 demonstrates the concept of remotely controlling Robonaut's operations.

Enabling Technology for Telepresence

The term "tele-presence" literally illustrates two layers of implications: people are virtually *present* at the remote site, and that presence is "tele" i.e., distant. The implications lead to two essential components in telepresence systems: telepresent *location* and *communication*. Telepresent locations indicate the locations in which the users are virtually present.



Telepresence. Figure 1. Example of telepresence control of robonaut.

Standard locations vary from conventional VR to augmented reality (AR) and to real world environments. The interactivity between host and remote locations is accomplished via network link. It is generally accepted that there are three independent aspects of presence which contribute to achieving a sense of remote telepresence [5]:

1. The extent of the sensory information. It is necessary to provide the operator with the same level of sensory information that they would receive if they were actually present at the remote site.
2. Control of sensors. It must be possible to move the sensing devices around the environment as if the operator was at the remote site.
3. The ability to modify the remote site and actively interact with it.

To achieve the ultimate sense of presence at a remote site requires the full combination of all these aspects and is essentially what most telepresence systems are attempting to create. The telepresence aspect of the teleoperator is to provide sensory feedback to the operator in such a way that the feeling of sensory presence is conveyed. A high degree of sensory presence is required when the operational tasks are wide ranging, complex and uncertain, for instance, in hazardous or difficult environments.

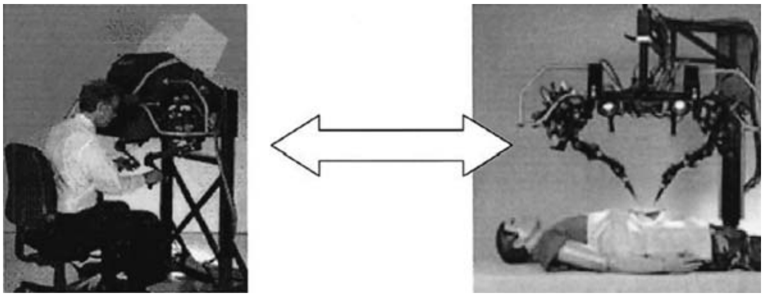
Telepresent Location Technology

Telepresence enables people physically located in their host location to behave and receive stimuli as though at a remote site. The technology used in host location is similar to what is applied in conventional VR and AR applications. The difference is that, in VR and AR applications, users are immersed in a computer synthesized environment (VR) or a combination of a real location and supplemental objects or information (AR), whereas, in telepresence, users are immersed in a remote real world. Multimedia sensory feedback is required to achieve this, which takes multiple modalities including haptic/tactile, sound, vision or even olfactory.

A visually coupled system utilizes a display system such as a Head-Mounted Display (HMD). These systems are worn over the eyes of the user and generally only give the impression that the user is immersed within the environment displayed in the images. HMDs enable a higher degree of user immersion compared with standard monitors or television displays. Haptic interfaces bring more and more realism to telepresence. Olfactory displays could also be used but practical commercial systems have yet to be developed.

Control is a key element of telepresence operations to create vivid awareness. By tracking the operators head position with a sensor, the remote camera system can be controlled in such a way that the head motion is replicated in the remote environment. The effect of this relationship between the operator and the remote mechanical system is to further enhance telepresence and to allow the operator to concentrate on the task itself instead of how to achieve it.

Most of the equipments deployed at remote locations focus on acquisition of the surroundings' information, modification of the remote site and interaction with it. Typically, these equipments include: pan/tilt monoscopic



Telepresence. Figure 2. Telepresence surgery application (courtesy of SRI International).

and stereoscopic camera platforms, other sensor platforms including microphones and touch/force feedback sensors, slave manipulators and grippers, and mobility providers such as wheeled or tracked vehicles. Mostly, the approach of capturing/compressing video/sound in real time is in demand due to the bandwidth limitation of telecommunication link.

Readers are referred to two other chapters in this book for further information: Virtual and Augmented Reality, and Tele-Haptics.

Telecommunications Link

Although any available telecommunication links may be used in a telepresence system, specific systems dictated different demands on network configurations, such as bandwidth requirements, and the sensitivity to network impairments (latency, jitter and packet loss). For instance, in a telepresence system enhanced with haptic interface, the standard implementation requires that haptic information (force/kinematic data) to be transferred over communication link between home and remote locations. Since network latency adds phase lag to the signal, this lag limits the effective bandwidth of closed haptic control loop and may result in an unstable haptic display. Such instability poses a safety threat because of the relatively large force-generating capability of the hardware. Significant research results have been achieved in teleoperation domain to deal with varied network delays [6–8].

Telepresence Applications

The flexibility of telepresence system allows a myriad of applications. The following is a selection of some of the current telepresence systems, both completed and under investigation. Each can be fully realized with the latest technology, and most can operate, albeit at a reduced service level, over the existing network.

Telemedicine

Telemedicine is the delivery of healthcare services, where distance is a critical factor. For example, in a battlefield scenario, the first hour after a soldier has been injured is critical. Telepresence could be used to transport the "presence" of the medical doctor, surgeon and consultant to hospital operating rooms or the scenes of battlefield to assist paramedics. This would facilitate the more effective use of an increasingly limited resource – the medical specialist. The armed forces have an obvious interest since the combination of telepresence, teleoperation, and telerobotics can potentially save the lives of battle casualties by allowing them prompt attention in mobile operating theatres by remote surgeons, as illustrated in Fig. 2.

Hazardous and Hostile Environments

Many other applications in situations where humans are exposed to hazardous situations are readily recognized as suitable candidates for telepresence. Mining, bomb disposal, military operations, rescue of victims from fire, toxic atmospheres, or even hostage situations, are some examples. Remote controlled robotic devices are being used to detect and remove bombs and land mines. By providing the operator the ability to feel what the robot is feeling, the detection/removal task can be done more efficiently and safely.

Entertainment

Telepresence systems could be incorporated into theme or nature parks to allow observers to travel through coral reefs, explore underground caves, or in amusement parks the elderly or infirm could experience the thrill of live roller coaster rides without the associated risks. Figure 3 shows an example of a telepresence musical performance with musicians at two remote

Telepresence. Figure 3. Connected performance between two difference spaces.

sites: Stockholm's KTH Royal Institute of Technology in Sweden and Stanford in the United States.

Remote Surveillance

Security forces throughout the world are increasingly resorting to video recording as a means of crime certification and prevention. Telepresence is one of the logical developments in this trend, with each security officer linked visually and orally at all times to headquarter. Sensor platforms operating autonomously in robotic mode could monitor sensitive areas. They could use software algorithms that would identify, for example, intruders. They would then proceed to track these intruders while sending an alert to a human operator requesting attention. The operator would be wearing a HMD helmet and become telepresent at the remote location and take over control of the sensor platform. This would then allow the operator to investigate more.

Teleconferencing

Teleconferencing literally means "conference at a distance," where multiple geographically dispersed users have a meeting of some sort across a telecommunications link. It is one of the uses that the integral display on the desk could be put to, producing a life size head-and-shoulder image of the remote user. The large size of the screen ensures that peripheral vision is substantially filled, thus creating the illusion of "being there." The high definition of the display allows it to be multitasked during teleconferencing as the main viewer and a computer monitor. By using an infra red light pen, the screen can also function as an electronic white board, allowing multiple participants to interact in the same media-space in real time. People sitting at desks thousands of miles apart can thus come together to realize a real time working environment that closely mimics reality.

Cross-References

► Tele-operation
► Virtual Presence

References

1. W. Gibson, "Neuromancer," Ace Books, New York, 1984.
2. M. Minsky, "Telepresence," Omni, New York, 1980, pp. 45–51.
3. B.G. Witmer and M.J. Singer, "Measuring Presence in Virtual Environments: A Presence Questionnaire," Presence: Teleoperators and Virtual Environments, Vol. 7, No. 3, 1998, pp. 225–240.
4. A. Fisch, C. Mavroidis, J. Melli-Huber, and Y. Bar-Cohen, "Haptic Devices for Virtual Reality, Telepresence, and Human-Assistive Robotics," in Y. Bar-Cohen and C. Breazeal (Eds.), Invited Chapter in "Biologically-Inspired Intelligent Robots," SPIE Press, 2003.
5. J. Pretlove and R. Asbery, "Get a head in virtual reality," IEE Colloquium on Control of Remotely Operated Systems: Teleassistance and Telepresence, 12 May 1995, pp. 8/1–8/5.
6. R. Anderson and M. Spong, "Bilateral Control of Teleoperators with Time Delay," IEEE Transactions on Automatic Control, Vol. 34, No. 5, 1989, pp. 494–501.
7. N. Hogan, "On the Stability of Manipulators Performing Contact Tasks," IEEE Journal of Robotics and Automation, Vol. 4, No. 6, 1988, pp. 677–686.
8. S. Shirmohammadi, N.H. Woo, and S. Alavi, "Network Lag Mitigation Methods in Collaborative Distributed Simulations," Proceedings of the International Symposium on Collaborative Technologies and Systems (CTS' 05), Saint Louis, CO, May 2005.

Three Dimensional Content-Based Search and Retrieval of CAD Objects

Athanasios Madelmis, Petros Daras, Dimitrios Tzovaras, Michael G. Strintzis
Aristotle University of Thessaloniki, Thessaloniki, Greece

Synonyms

▶ Three-dimensional object retrieval; ▶ Three-dimensional search engine; ▶ CAD objects; ▶ Similarity

Definition

A system capable of retrieving computer aided design (CAD) objects similar to a given CAD object.

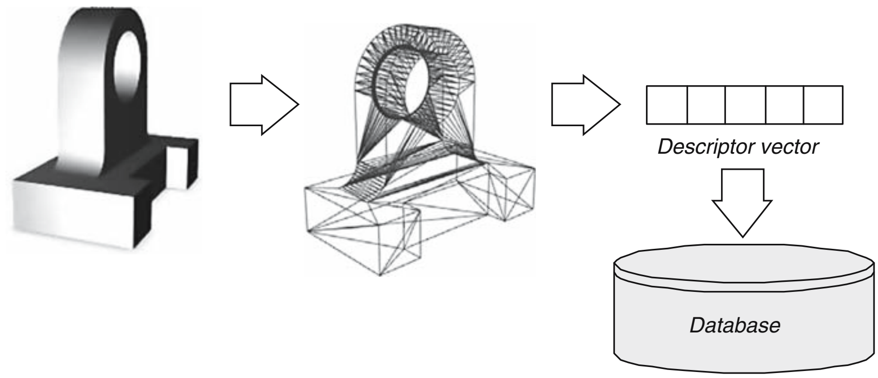
Introduction

During the last years, the development of modern, user friendly, and powerful CAD systems, led to the creation of a great number of mechanical parts used for the development of larger mechanical assemblies. In professional and collaborative CAD environments, very large databases with simple and complex 3D mechanical parts have already been created, yet the task of retrieving the desired 3D object is not a trivial task. This mainly happens due to the fact that the traditional text-based search and retrieval techniques cannot be successfully applied for 3D objects because they rely on annotations provided by humans, thus, the retrieval efficiency depends on the human annotator. Therefore, the annotation can be ambiguous, incomplete, or limited. Thus, methods able to retrieve 3D objects have been presented in the literature, which uses 3D objects as queries (query by example). Generally, the existing approaches focus on 3D shape retrieval; however, only few of them are evaluated in the special application of 3D CAD design.

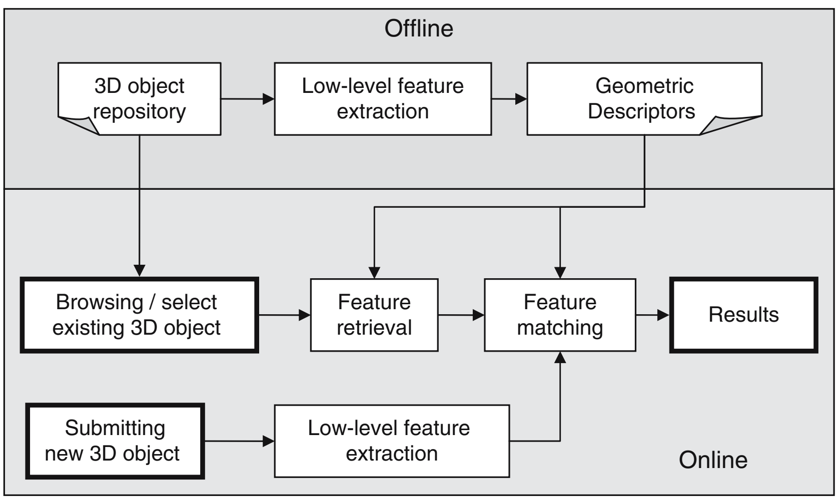
The major difference between 3D CAD parts and general purpose 3D objects derives from their usage and their representation. A 3D CAD part search engine can enhance the productivity of the designers and extend the product life of the parts. Although the majority of the parts are reusable and can be utilized in many assemblies, the limited exploration capabilities of the existing PDM software do not allow the designers to fully exploit the existing content. Thus, parts that already exist in the database are redesigned from scratch because the designer is unable to locate them, resulting in a longer production time. A 3D search and retrieval engine, allows the designer to query the system for the desired object using an example as a query. In addition, if the designer fails to find the exact part, the retrieved results could be used as queries for a new search, or, if the retrieval fails (in these cases, probably the desired object is not stored in the database), the designer may use one or more of the retrieved results to build the desired part. Although the latter is the worst case scenario, the modification usually requires less effort than the design of the part from scratch.

Additionally, the multiple and proprietary storage formats and the copyright of the most models have seriously discouraged the research on the field; however, the creation of the National Design Repository (NDR) [1] and the Engineering Shape Benchmark (ESB) [2] provided the research community with engineering specific datasets. NDR consists of six smaller datasets each one dedicated to a particular category (e.g., CAD primitives, industrial CAD, LEGO objects). ESB follows the multiple layer classification approach by classifying every mechanical part in three broad categories and 45 more specific categories. The majority of the methods that will be presented in the sequel are evaluated with respect to their retrieval accuracy using these datasets.

A typical 3D shape search and retrieval framework is presented in Fig. 2 and is adopted by the majority of the presented methods consisting of a database where the initial 3D objects are stored along with their shape features and an online query engine. Every 3D model is described by a shape descriptor, which provides a compact overall description/representation of the shape. A shape descriptor is commonly a vector (Fig. 1), a matrix, an attributed graph, or any combination of them. To efficiently search a large collection online, indexed data structures and searching algorithms are utilized to provide fast and accurate retrieval. The online query engine computes the query's shape descriptor, and performs the search and retrieval tasks by intuitively comparing the features of the query object with the features of the shapes stored in the database. The objects similar to the query 3D objects are retrieved by matching the shape descriptors in the database to the query's shape descriptor. The matching accuracy between two descriptors is measured using similarity metrics. Finally, the retrieved models are visualized to the end user (Fig. 3).



Three Dimensional Content-Based Search and Retrieval of CAD Objects. Figure 1. Descriptor extraction procedure.



Three Dimensional Content-Based Search and Retrieval of CAD Objects. Figure 2. Search and retrieval architecture.

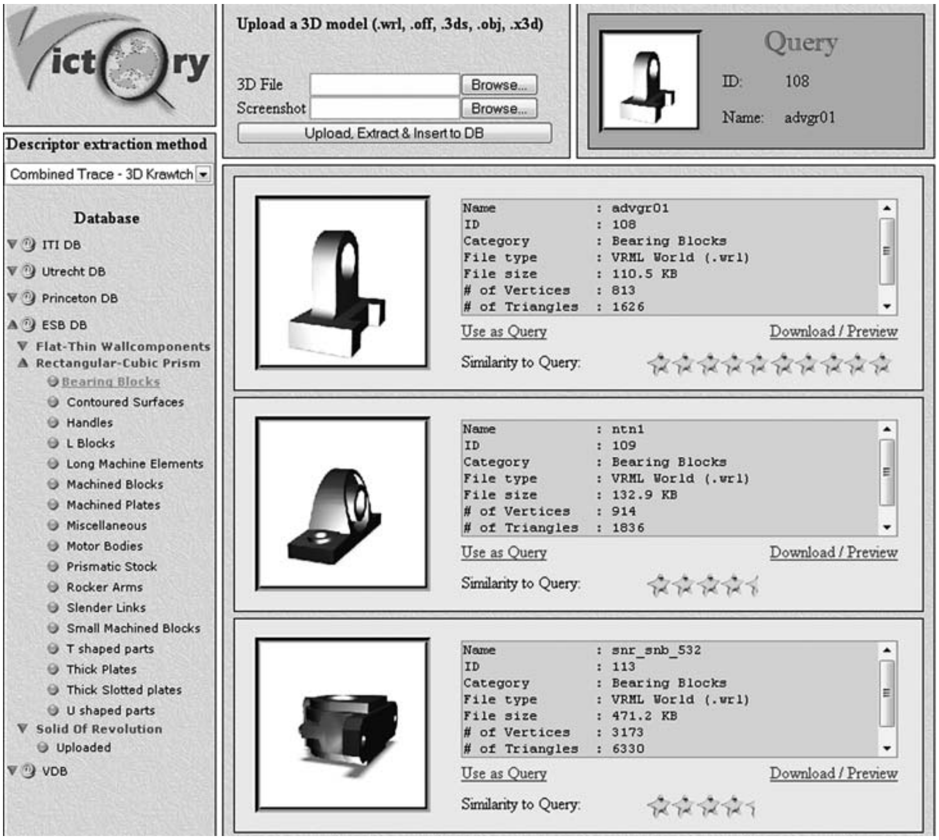
The user has usually the following options for the selection of a query 3D object:

- Browsing: The user selects a new query object by browsing the repository, or from the obtained results of a previous search.
- Query by example: The user provides a 3D model that he or she owns, or creates a 3D shape query from scratch using the online tools.

In the rest of this article, 3D search and retrieval approaches dedicated to the CAD parts are briefly reviewed. Additionally, methods which have been previously proposed by the authors and compared to well-known general purpose 3D search approaches for their effectiveness in CAD search, using the ESB database are also presented.

The approaches that have been presented in the literature, utilize various shape features and can be classified into four broad categories: The *global approaches*, which extract the shape features considering the whole object, the *local approaches*, which compute the descriptor of the shape by examining the 3D object locally or partially, the *graph-based approaches* that approximate the object with an attributed graph and the *view-based approaches*, which are based on, usually orthographic, views of the 3D object.

The global feature-based methods process the 3D object and produce a multidimensional descriptor vector, which captures shape information of the whole object. The vector's dimensionality is usually constant for all shapes, in order to utilize common distances as dissimilarity metrics. These approaches either compute



Three Dimensional Content-Based Search and Retrieval of CAD Objects. Figure 3. Search and retrieval of CAD objects (http://www.victory-eu.org/).

very fast a small number of simple features (e.g., aspect ratio, surface area) that can be used as an active filter before applying other methods to improve results and search times, or compute a complete shape descriptor, which captures the full shape information as a more intuitive descriptor (e.g., Fourier Transformation, geometric moments, etc.). The implementation of such approaches is usually fast and rather straightforward and can provide satisfactory retrieval accuracy. The local approaches compute local shape features for every point in the shape. These features are usually integrated to a single feature (e.g., using histograms or spherical harmonics). In contrast, the major advantage of the graph-based methods is that they take into account the topology of the object, while the local and global methods deal only with the geometry of the shape. Finally, the view-based approaches, focusing on providing multiview information redundancy by extracting image descriptors from various 2D views of the object. These descriptors are appropriately combined to be suitable for a 3D search and retrieval application.

Review of the Existing Approaches

Only few approaches concerning 3D search and retrieval of mechanical parts have been presented, until recently, due to the lack of a benchmark for evaluation of their retrieval accuracy. Due to the introduction of the NDR [1] and the ESB [2] along with the recent Shape and Retrieval Contest [3], where a track for engineering shapes retrieval was held, new perspectives on engineering shape retrieval have been arisen.

In [4], the development of a 3D Internet-based search engine is described, which allows the designers to locate existing parts or parts that have similar shape to a desired "newly designed" part. The search engine analyses the shape characteristics of the target model and performs a similarity matching through a filtering technique. The shape descriptor is *global* and is based on primitive geometric features: the volume and the surface of the object, their ratio, the bounding box's aspect ratio, the crinkliness, the compactness, and the number of holes. In [5], another *global* approach based on primitive geometric features is presented. Every 3D CAD object is appropriately transformed in a volumetric function and is normalized with respect to rotation using the classical principal component analysis (PCA) approach. The shape descriptor of every object consists of second-order moment invariants, spherical kernel invariants, and bounding box characteristics. The retrieval is performed by utilizing an appropriate hashing function.

In [6], the spherical trace transform (STT) is utilized for CAD object retrieval (Fig. 3). Initially, every object is translated and scaled so that all objects are expressed in the same coordinate system. To achieve the latter, a local coordinate system has been defined centered to the mass center of the object and scaled so that the object fits to the unit sphere. Then, the object's binary volumetric function is computed and the STT is performed as follows: a set of radius segments is defined. Every radius segment is formed by the intersection of a radius with the object. In addition, a set of spheres, concentric to the unit sphere is defined and for every sphere, a set of plane segments is defined. Each plane segment is formed by the intersection of a plane tangent to the sphere at point with the object. The radius segments and the points are uniformly distributed on the sphere's surface exploiting the icosahedric-based tessellation. Every segment is treated as a one-dimensional signal, and descriptors based on classical 1D discrete Fourier transform and an integration transform are computed. Every plane segment is treated as a 2D signal and descriptors based on the Krawtchouk and the Zernike moments are computed. Then, the spherical Fourier transform is applied on the extracted separately on every descriptor, so that the final descriptors are invariant under rotation and, thus, appropriate for 3D object matching. The descriptor vector comparison is based on the Minkowski distance.

The global approaches of generalized radon transform (GRT) [7] and the weighted 3D Krawtchouk moments [8] are combined with the STT [9] and are utilized for 3D CAD search and Retrieval (http://www.victory-eu.org). Two instances of GRT are used: the radial integration transform (RIT) and the SIT. In RIT, the object is partitioned into radius segments uniformly distributed in the surface of the object's bounding sphere. The integral of every radius segment forms the descriptor vector. The object is partitioned to spherical segments and the integral of every segment forms the SIT descriptor. The weighted 3D Krawtchouk moments are special category of geometric moments based on Krawtchouk polynomials and are able to capture sharp shape changes. Thus, their retrieval performance for this kind of objects is significantly increased.

In [10], a general approach based on spherical harmonics to transform rotation dependent shape descriptors into rotation independent ones is presented. The method is applicable to every shape descriptor, which is defined as either a collection of spherical functions or as a function on a voxel grid. In the latter case a collection of spherical functions is obtained from the function on the voxel grid by restricting the grid to concentric spheres. From the collection of spherical functions a rotation invariant descriptor is computed first by decomposing the spherical function into its spherical harmonics expansion, and summing the harmonics within each frequency, and computing the Euclidean norm for each frequency component. The resulting shape descriptor is a 2D histogram indexed by radius and frequency, which is invariant to rotations about the object's center of the mass. This approach offers an alternative for pose normalization, because it achieves rotation invariant shape descriptors.

In [11], a graph-based approach is proposed. A general indexing scheme applied for solid models of mechanical designs is performed by creating a mapping of solid model boundary representations and engineering attributes into a graph called "Model Signature Graph." Every graph is transformed to a feature vector, using a spectral-graph approach, and the metric distances, called EigenDistances among CAD models are computed, which are used for the creation of a spatial index of the model comparison space using a tree data structure. A set of heuristics has been constructed that exploit properties of the model comparison space to ensure that the tree indices remain balanced.

In [12], an application of the augmented multiresolutional Reeb graphs (aMRG) for shape retrieval of the CAD models is presented. The method is based on the well-known Reeb graphs topological descriptors. Using the multiresolution property of Reeb graphs and enhancing the resulting graphs with additional geometrical and topological information, the aMRG approach was proved to be very efficient in the retrieval of high-quality 3D CAD objects.

In [13, 14], a complete reconfigurable engineering shape search engine is presented, which combines global features with a graph-based descriptor. Every model is described using multiple features: a graph that captures the topology of the CAD part, generated after the extraction of the line skeleton of the part, and a global descriptor vector that consist of the traditional moment invariants and primitive geometric features (e.g., bounding box aspect ratio, total surface area). The retrieval of the similar objects is performed using an R-Tree-based index structure for the geometric features and a decision-tree-based approach for the skeletal graphs. The retrieval accuracy of the system is enhanced using "Relevance Feedback" approaches. In [15], an extension of the previous shape representation approach is given. The object is represented by a multi-scale hierarchical graph, which encodes at the same time the topology of the object along with primitive shape features.

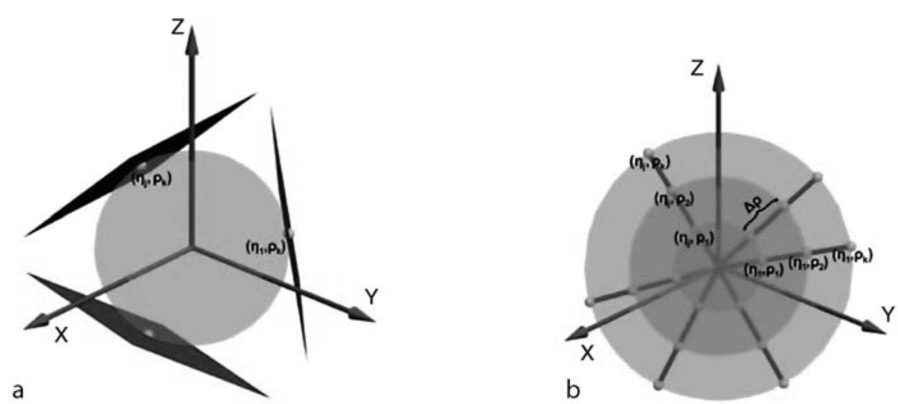
In [9], two *view-based* methods are proposed. The 3D object is initially normalized with respect to rotation and translation. The latter is achieved by applying PCA and scaling the object so that its enclosing sphere has a predefined radius. Then, views of the 3D objects are extracted. In all views the models silhouettes are extracted and compared in pairs. In the first method the similitude measure is obtained by integrating on the pairs of views the difference between the areas of the silhouettes' union and the silhouettes' intersection. In the second method the external contour of the silhouettes is examined. More specifically, their convexities and concavities at different scale levels are extracted and a multiscale representation is built. The pairs of contours are then compared by elastic matching using dynamic programming.

In [16], it is assumed that two models are similar, if they look similar from all viewing angles and the LFD is introduced. According to the lightfield descriptor (LFD) method, the similarity between two models is based on ten Light-Fields, silhouettes of the 3D shape obtained from ten viewing angles distributed evenly on the viewing sphere. Every light-field is considered as an image and descriptors are extracted based on Zernike moments and Discrete Fourier Transform. The dissimilarity of two shapes is found as the minimal dissimilarity obtained by rotating the viewing sphere of one LFD relative to the other LFD. The running time of the retrieval process is reduced by a clever multi-step approach supporting early rejection of non-relevant models.

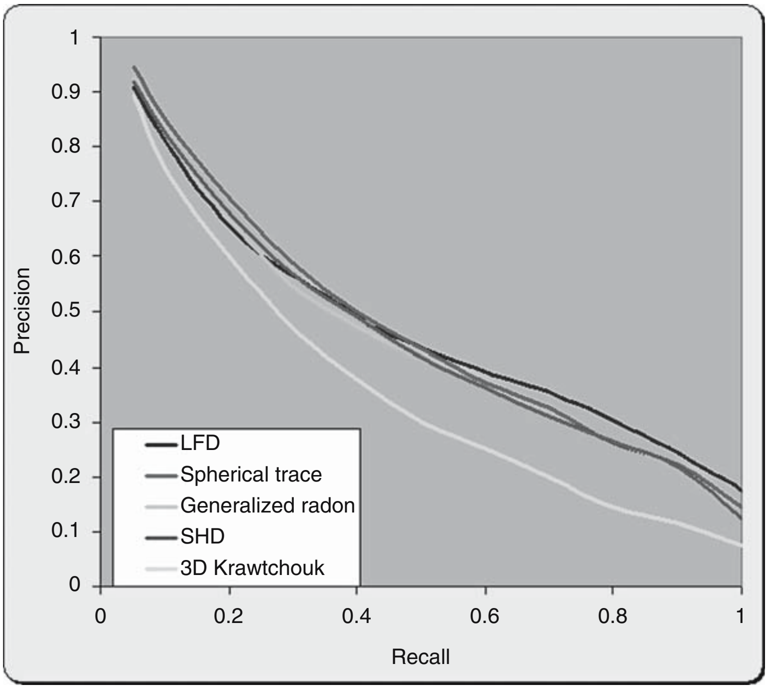
In [17], the distance measure, in addition to the shape feature is considered as a key factor in the shape based 3D model retrieval. Thus, a database-adaptive distance measure for 3D has been proposed. The method utilizes a feature dimensionality reduction approach based on an unsupervised learning of features to produce salient, lower dimensional feature from the original feature. This method also combines a multiresolution shape comparison approach with the database adaptive distance measure.

Comparative Retrieval Performance

The methods of GRT [7], STT [6], 3D Krawtchouk moments [8], LFD [16] and spherical harmonics descriptor (SHD) [10] have been evaluated for their retrieval performance using the ESB database. The evaluation has been performed using precision-recall



Three Dimensional Content-Based Search and Retrieval of CAD Objects. Figure 4. The spherical trace transform [6].



Three Dimensional Content-Based Search and Retrieval of CAD Objects. Figure 5. Precision recall diagrams of general purpose approaches.

diagrams, which can depict the global behavior of every method during the retrieval. Precision is the fraction of the 3D objects retrieved that are relevant to the user's query, while recall is the fraction of the objects that are relevant to the query and have been successfully retrieved. Figure 4 depicts a comparative retrieval performance for the aforementioned approaches. It is obvious that the retrieval performance of all approaches, except from 3D Krawtchouk is comparable and only few details distinguish their efficiency. More specifically, the retrieval of STT outperforms all the methods when recall is less than 50%, while LFD outperforms all other approaches for recall greater than 50%. Thus, using the STT more similar objects are first retrieved, while using the LFD approach, the lowest ranked similar result appears earlier than the other approaches.

From the aforementioned analysis it is obvious that the content-based search and retrieval of 3D CAD objects is still an open problem. The methods that have been proposed so far are mainly expansions of general purpose 3D search approaches and, thus, their retrieval accuracy lacks when compared with their retrieval accuracy in the general purpose applications. Future work could involve combination of methods dedicated to CAD parts with sophisticated knowledge-based schemes (e.g., ontologies) to bridge the semantic gap.

References

1. D. Bespalov, C.Y. Ip, W.C. Regli, J. Shaffer, "Benchmarking CAD search techniques," Proceedings of the 2005 ACM symposium on solid and physical modeling, pp. 275–286, 2005.
2. S. Jayanti, Y. Kalyanaraman, N. Iyer, and K. Ramani, "Developing An Engineering Shape Benchmark For CAD Models," Computer-Aided Design, Vol. 38, No. 9, Shape Similarity Detection and Search for CAD/CAE Applications, pp. 939–953, September 2006.
3. R.C. Veltkamp and F.B. ter Haar, "SHREC2007: 3D Shape Retrieval Contest," Department of Information and Computing Sciences, Utrecht University, Technical Report UU-CS-2007–015.
4. H. Rea, J. Corney, D. Clark, J. Pritchard, M. Breaks and R. MacLeod, "Part sourcing in a global market," Proceedings of ICeCE' 01, China Machine Press, Beijing, 2001.
5. G. Cybenko, A. Bhasin, and K.D. Cohen, Pattern Recognition of 3D CAD Objects: Towards an Electronic Yellow Pages of

Mechanical Parts, Smart Engineering Systems Design, Vol. 1. pp. 1–13, 1997.
6. P. Daras, D. Tzovaras, A. Axenopoulos, D. Zarpalas, A. Mademlis and M.G. Strintzis, "The Spherical Trace Transform," SHREC 2007 Proceedings.
7. P. Daras, D. Zarpalas, D. Tzovaras, M.G.S trintzis, "Efficient 3-D model search and retrieval using generalized 3-D radon transforms," IEEE Transactions on Multimedia, Vol. 8, No 1, February 2006, pp. 101–114.
8. A. Mademlis, A. Axenopoulos, P. Daras, D. Tzovaras and M.G. Strintzis, "3D Content-based Search based on 3D Krawtchouk moments," 3D Data Processing, Visualization & Transmission (3DPVT 2006), University of North Carolina, Chapel Hill, June 2006.
9. T. Napoléon, T. Adamek, F. Schmitt, and N.E. O'Connor, "Multi-view 3D retrieval using silhouette intersection and multi-scale contour representation," SHREC 2007 Proceedings.
10. M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3D shape descriptors," In Symposium on Geometry Processing, 2003.
11. D. McWherter, W.C. Regli, "An approach to indexing databases of solid models," Technical report DU-MCS-01–02, Philadelphia (PA): Drexel University, 2001.
12. T. Tung and F. Schmitt, "Shape retrieval of watertight models and CAD models using aMRG," SHREC 2007 Proceedings.
13. N. Iyer, Y. Kalyanaraman, K. Lou, S. Jayanti, and K. Ramani, "A Reconfigurable, Intelligent 3D Engineering Shape Search System Part I: Shape Representation," Proceedings of ASME DETC' 03, 23rd Computers and Information in engineering (CIE) Conference, Chicago, IL, September 2–6, 2003.
14. K. Lou, S. Jayanti, N. Iyer, Y. Kalyanaraman, K. Ramani, and S. Prabhakar, "A Reconfigurable, Intelligent 3D Engineering Shape Search System Part II: Database Indexing, Retrieval and Clusterin," Proceedings of ASME DETC' 03, 23rd Computers and Information in engineering (CIE) Conference, Chicago, IL, September 2–6, 2003.
15. N. Iyer, K. Lou, S. Jayanti, Y. Kalyanaraman, and K. Ramani, "A Multi-Scale Hierarchical 3d Shape Representation for Similar Shape Retrieval," Proceedings of TMCE 2004, Vol. 2 – Tools and Methods for Competitive Engineering, Lausanne, Switzerland, pp. 1117–1118.
16. D.Y. Chen, X.-P. Tian, Y.-T. Shen, and M. Ouhyoung, "On visual similarity based 3D model retrieval," Computer Graphics Forum (EG 2003 Proceedings), Vol. 22, No. 3, 2003.
17. J. Kobayashi, A. Yamamoto, T. Shimizu, and R. Ohbuchi, "A database-adaptive distance measure for 3D model retrieval," SHREC 2007 Proceedings.
18. P. Jiantao and K. Ramani, "A 2D Sketch Based User Interface for 3D CAD Model Retrieval," Journal of Computer Aided Design and Application, CAD 2005, Vol. 2, June 20–24, 2005, pp. 717–727.
19. Z. Li, M. Liu, D.C. Anderson, and K. Ramani, "Semantics-based Design Knowledge Annotation and Retrieval," Proceedings of the ASME 2005 International Design Engineering Technical Conference & Computers and Information in Engineering Conference, Vol. 3, pp. 799–808, Long Beach, CA, September 24–28, 2005.
20. D. Bespalov, W.C. Regli, A. Shokoufandeh, "Reeb Graph based shape retrieval for CAD," Proceedings of ASME design engineering technical conferences, Computers and information in engineering conference, 2003.
21. H-P. Kriegel, P. Kröger, Z. Mashael, M. Pfeifle, M. Pötke, S. Seidl, "Effective Similarity Search on Voxelized CAD Objects," Proceedings of Eigth International Conference on Database Systems for Advanced Applications, pp. 27–36, 2003.
22. H-P. Kriegel, S. Brecheisen, P. Kröger, M. Pfeifle, M. Schubert, "Using sets of feature vectors for similarity search on voxelized CAD objects," Proceedings of ACM SIGMOD International Conference on Management of Data, pp. 587–598, 2003.

Three Dimensional Copyright Protection

► Three Dimensional Object Watermarking

Three Dimensional Face Identification

► Frequency Domain Representations for Three Dimensional Face Recognition

Three Dimensional Face Identification-authentication

► Face and Facial Expression Recognition Using Three Dimensional Data: an Overview

Three Dimensional Face Recognition

Iordanis Mpiperis, Sotiris Malassiotis, Michael G. Strintzis
Informatics and Telematics Institute, Thessaloniki, Greece

Synonyms

► Three-dimensional face identification; ► Three-dimensional face authentication; ► Three-dimensional face verification

Definition

3D face recognition is referred to as person identification using the three-dimensional geometry of the face.

Introduction

The authentication or recognition of an individual's identity is an integral part of numerous systems developed in the context of a variety of applications ranging from law enforcement and surveillance of public places to digital rights management and engaging human-computer communication. Biometrics, the physical and behavioral traits that characterize uniquely a human, such as fingerprints, signature, voice and face, have attracted growing research interest as an alternative means for reliable person identification over passwords, PINs, smart cards and similar mechanisms. In particular, person identification from face images has gained a prominent position among other biometrics. This is mainly due to the social acceptability of face as a biometric and the user-friendliness of face recognition systems, but is also motivated by the inherent ability of humans to recognize faces visually in their natural inter-personal communication with apparently minimum effort.

For many decades, face images and image sequences have been used as a basis for establishing identity and a plethora of algorithms have been proposed; an extensive review may be found in [1]. However, despite the great advances, face recognition from images in vivo is still far from being a reliable and robust way for establishing identity. Recorded images of the same face may have large variations in appearance, caused for example by changes in the illumination of the scene or orientation of the head. This contributes considerably to the performance of existing systems when applied to non-controlled conditions. To cope with these difficulties, a promising new technology, 3D face recognition, was recently investigated. This relies on discriminative features extracted from the 3D geometry of the face, which is captured by special 3D scanners or 3D cameras. Since 3D geometry is independent of appearance variations, 3D face recognition has demonstrated very good accuracy even in "difficult" conditions.

A typical face recognition system based on 3D data consists of four parts in general, which involve data acquisition, pre-processing, feature extraction and feature classification. 3D data (also called depth) may be captured directly, using a laser scanner, or indirectly, applying stereo photogrammetry techniques to color images captured by multiple cameras or using a single camera and a structured-light projector. The result of the acquisition is usually a "range image" with each pixel recording the distance of a point on the object from the camera. The accuracy and resolution of the data depends on the acquisition technology and environment. Laser scanners for example lead to more accurate 3D face models, but need several seconds to grab a single image, while being also relatively expensive. On the other hand, stereo-based and structured-light based techniques are faster and cheaper but produce noisy data.

Once 3D facial data is acquired, it undergoes standard pre-processing, such as noise removal, hole filling and smoothening, or more elaborate pre-processing, such as head pose compensation usually using variants of the Iterative Closest Point algorithm [2]. Then, several surface features that hopefully characterize unambiguously each facial surface are computed. These features, also called descriptors, may be local or global and they are usually defined in terms of surface curvature, distance between surface points, moments and other geometrical attributes of the surface. Point signatures [3], spin images [4] and snapshots [5] are examples of local surface descriptors, while contour-lines [6] and canonical images [7] are more global descriptors. Particular emphasis is placed on these features being pose-invariant to dispense with pose compensation and eliminate the associated errors. Exhaustive literature surveys for 3D and multimodal 2D and 3D face recognition algorithms may be found in [8, 9].

Feature classification usually follows vector-oriented pattern classification techniques such as Support Vector Machines, Neural Networks, Principal Components Analysis and Linear Discriminant Analysis, or statistic-based techniques, such as Bayes Classification and Hidden Markov Models [10]. Most of the classifiers proposed in literature require training with a representative number of training faces to adjust a set of parameters. Then, given this set of parameters and a gallery of already processed faces whose identity is known, recognition of new unseen faces, called probe faces, may be performed. The features of the probe face are extracted first and then they are matched against the corresponding features of the gallery faces to establish a measure of similarity. The probe face is finally classified to the most similar gallery face.

Quantitative evaluation and comparison between face recognition algorithms is usually performed under two experimental scenarios, face identification, which responds to the question "Who is he?", and face authentication, which responds to the question "Is he really who he claims to be?" Authentication is a 1:1 match process where the face in question is compared with the face whose identity is claimed. On the other hand, identification is a 1:N match process, since the face in question is compared with all faces stored in a database. Performance is assessed by means of the Cumulative Match Characteristic (CMC) and the Receiver Operating Characteristic (ROC) curves for the identification and the authentication scenario respectively. The CMC is the graph of recognition rate versus the ranking of similarity measures, while the ROC curve is the graph of False Rejection Ratio versus False Acceptance Ratio. Recently, Face Recognition Grand Challenge (FRGC) [11], has been established as a framework for evaluation of 3D face recognition algorithms. FRGC is a project trying to promote and advance 3D face recognition technology and to this end it offers a common data set and a common infrastructure for experiments and quantitative assessment. Its goal is to exceed by an order of magnitude the performance of state-of-the-art 2D face recognition algorithms promoted by a corresponding project, the Face Recognition Vendor Test (FRVT), and by this time many algorithms capable of achieving recognition rates greater than 90% have been proposed.

The availability of cheap and accurate 3D sensors has led to a booming of 3D face recognition technology over the past 10 years and research has already reached a sufficiently mature level. Although pitfalls inherent in 2D systems have been successfully encountered or alleviated, a new challenge has come up regarding 3D face recognition, facial deformations that alter the geometry of the face such as those caused by facial expressions. Most algorithms proposed so far are sensitive to the deformation of the surface caused by facial expressions and their performance decreases considerably when dealing with non-neutral faces. Of course, there are other problems that need to be encountered too, such as the age variation and occlusion, but current research is primarily focusing on expression-invariant 3D face recognition.

Cross-References

► Face and Facial Expression Recognition Using 3D Data: An overview
► Three Dimensional Face Identification
► Three Dimensional Face Verification

References

1. W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face recognition: A literature survey," ACM Computing Surveys, Vol. 35, No. 4, 2003, pp. 399–458.
2. P.J. Besl and N.D. McKay, "A method for registration of 3-D shapes," IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 14, No. 2, February 1992, pp. 239–256.
3. C.S. Chua, F. Han, and Y.K. Ho, "3D human face recognition using point signature," Proceedings of the fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 233–238, 2000.
4. A. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 5, 1999, pp. 443–449.
5. S. Malassiotis and M.G. Strintzis, "Snapshots: A Novel Local Surface Descriptor and Matching Algorithm for Robust 3D Surface Alignment," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29, No. 7, July, 2007, pp. 1285–1290.
6. C. Samir, A. Srivastava, and M. Daoudi, "Three-Dimensional Face Recognition Using Shapes of Facial Curves," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, 2006, pp. 1858–1863.
7. A.M. Bronstein, M.M. Bronstein, and R. Kimmel, "Expression-Invariant Representations of Faces," IEEE Transactions on Image Processing, Vol. 16, No. 1, January, 2007, pp. 188–197.
8. K.W. Bowyer, K. Chang, and P. Flynn, "A Survey of Approaches and Challenges in 3D and Multi-Modal 3D + 2D Face Recognition," Computer Vision and Image Understanding, Vol. 101, 2006, pp. 1–15.
9. K. Chang, K. Bowyer, and P. Flynn, "An Evaluation of Multi-Modal 2D + 3D Face Biometrics," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 4, April, 2005, pp. 619–624.
10. R.O. Duda, E. Hart, and D.G. Stock, "Pattern Classification," Wiley, London, 2001.
11. P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the Face Recognition Grand Challenge," Proceedins IEEE Computer Vision and Pattern Recognition, pp. 947–954, June, 2005.

Three Dimensional Face Verification

► Frequency Domain Representations for Three Dimentional Face Recognition

Three Dimensional Motion Picture

► Three Dimentional Television Services: Introduction

Three Dimensional Object Retrieval

► Three Dimensional Content-Based Search and Retrieval of CAD Objects

Three Dimensional Object Watermarking

Athanasios Mademlis, Petros Daras, Dimitiros Tzovaras, Michael G. Strintzis
Aristotle University of Thessaloniki, Thessaloniki, Greece

Synonyms

► Three-dimensional Copyright protection; ► Steganography

Definition

An appropriate signal (called watermark) is hidden into the original 3D content being protected so that the copyright or the integrity of a 3D object can be verified.

During the last decade, the enormous number of publicly available multimedia data (e.g. digital photographs, 3D objects, etc.) along with the growth of the Internet, increased the demand for effective and practical approaches suitable for protecting the multimedia data. The most common way to achieve the latter is the use of watermarking techniques. Watermarking is a process where a given amount of information (which is called watermark), is embedded in the original content in such a way that is noticeable from the appropriate algorithms and imperceptible from the user. The 3D objects, in contrast to images and videos, are commonly represented as polygon meshes. As a consequence the principles of 3D object watermarking approaches differ from image and video watermarking, since essential information used in the 2D case, such as colour and texture, cannot be used in the 3D.

Depending on the application, two main categories of 3D object watermarking: the *robust watermarking*, which is utilized for copyright protection, and the *fragile watermarking*, which is used for content protection. Concerning the first category, the watermark should be detectable even if the initial content has been suffered various alterations (which are called attacks), while for fragile watermarking, the watermark should not be detectable if the object has been altered in any way. The most common attacks in a watermarked 3D object are:

- The *geometric attacks*, i.e., rotation, translation, and uniform scaling
- The *mesh attacks*.
  - Mesh simplification or subdivision where the 3D object is processed in order to produce a more coarse or detailed object respectively,
  - Mesh retriangulation where the original triangulation is discarded and a new one is formed.
  - Mesh resampling, where new points and triangles are formed based on the initial 3D object.
  - Mesh clutter, where a part of the object is removed

Every robust watermarking scheme can be further classified depending on the watermark detection procedure. In the *non-blind* systems, the original object is required during detection while in the *blind* systems, is not. Finally, every watermarking scheme can be *secure* if the original watermark information is required during the detection process or *non-secure* if the watermark information is extracted from the examined object.

There are three properties that can be observed in a watermark scheme: the capacity, which represents the maximum number of bits that can be embedded inside the media, the imperceptibility, i.e. the human eye should not be able to understand that the media is watermarked and the security, thus, people who do not know the necessary secret key or the watermark should neither detect nor read the watermark.

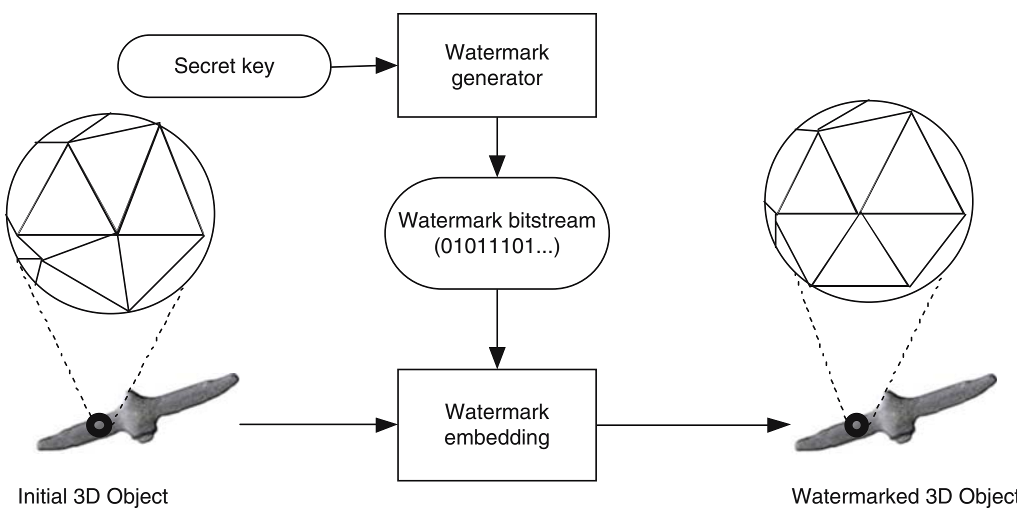
Each watermarking scheme is composed of two main and discrete steps: the watermark embedding procedure, where a secret key which identifies the copyright owner is hidden in the original object and the watermark extraction/verification procedure, where the watermark is extracted or verified in a possibly watermarked object. The watermark embedding procedure followed by the most approaches is depicted in Fig. 1. The watermark bitstream is hidden in the original 3D mesh by slightly changing the position of selected points of the original mesh. These alterations are detected in the watermark detection procedure, where the object is examined for the watermark existence. If the watermark is verified in a fragile scheme, then the author is sure that the initial object has not been modified. In the case of the watermark verification using a robust scheme, the author of the content can claim his copyright.

A very interesting part in 3D object watermarking is the way that the watermark bitstream alters the original mesh. Many approaches have been proposed so far which can be classified into two broad classes depending on the way of the embedding underlying principle: the methods that embed the watermark directly in the spatial domain and the methods that embed the watermark indirectly, using an intermediate transformation domain. Various methods process the 3D object in the spatial domain. The first methods proposed were relied on traditional mesh properties, like triangle similarity, tetrahedral volumes and surface normals [1–3]. The watermark embedding process affects the position of the vertices and the mesh connectivity so as to uniquely encode the desired bitstream in a detectable way.

In the transform-based methods (also called spectral methods), the 3D object is transformed in another domain and the watermark is embedded in the spectral coefficients of the mesh. The majority of the approaches presented in the literature focus on the wavelet transformation of the 3D meshes [4]. In accordance with image wavelet transform, the mesh is decomposed into a mesh of lower resolution (having less points and triangles) and the wavelet coefficients. However, approaches that rely on other transforms, e.g. the spherical harmonics transformation [5] have also been presented.

The major difference in the spatial and spectral approaches is the way that they affect the original mesh. Generally, the spatial domain methods modify specific points and polygons and result in a more imperceptible watermarked object when compared to the transform-based approaches. However, they are more vulnerable in the majority of watermark attacks.

To sum up, while many methods have been proposed so far, no robust approach exist that is able to handle the majority of the possible attacks. More specifically, although most of the presented blind methods are robust against geometric attacks, only few of them can handle intuitive mesh attacks. The non-blind approaches are more robust compared to blind approaches; however, the requirement of the original 3D mesh poses a major security risk.



Three Dimensional Object Watermarking. Figure 1. Typical watermark embedding by altering vertex positions.

Cross-References

► Digital Watermarking

References

1. O. Benedens, "Geometry-based watermarking of 3D models," Computer Graphics and Applications, IEEE, Vol. 19, No. 1, 1999, pp. 46–55.
2. O. Benedens, "Watermarking of 3D polygon based models with robustness against mesh simplification," Proceedings of SPIE: Security and Watermarking of Multimedia Contents, Vol. 3657, pp. 329–340, 1999.
3. R. Ohbuchi, H. Masuda, and M. Aono, May 1998, "Watermarking three dimensional polygonal models through geometric and topological modifications," in IEEE JSAC, pp. 551–560.
4. S. Kanai, H. Date, and T. Kishinami, "Digital watermarking for 3D polygons using multiresolution wavelet decomposition," in Sixth IFIP WG 5.2 GEO-6, 1998.
5. L. Li, D. Zhang, Z. Pan, J. Shi, K. Zhou, and K. Ye, "Watermarking 3D mesh by spherical parameterization," Computers and Graphics, Vol. 28, No. 6, December 2004, pp. 981–989.

Three Dimensional Search Engine

► Three Dimensional Content-Based Search and Retrieval of CAD Objects

Three Dimensional Television Services

Hari Kalva, Lakis Christodoulou, Liam Mayron, Oge Marques, Borko Furht
Florida Atlantic University, Boca Raton, FL, USA

Synonyms

► Three-dimensional Motion Picture; ► Three-dimensional TV; ► Three-dimensional Videos

Definition

This article explores the challenges and opportunities in developing and deploying 3D TV services. The 3D TV services can be seen as a general case of the multiview video that has been receiving significant attention lately. The keys to a successful 3D TV experience are the availability of content, the ease of use, the quality of experience, and the cost of deployment.

Introduction

The recent interest in 3D and multiview point (MV) TV can be attributed, in part, to the success of the MPEG-4 AVC/H.264 video coding standard. The coding gains made possible by H.264 can be applied to provide enhanced services such as multiview point TV and 3D television. Another reason for the increasing interest in 3D TV is the recent advances in the display technologies that have lowered the cost of projectors and 3D displays. While these technological advances have renewed interest in 3D/multiview coding, the successful deployment of 3D services still faces key challenges.

Ever since the invention of the Wheatstone stereoscope in 1938, there has been a constant interest in stereo imaging and this extended to the 3D motion pictures and television. There have been a large number of 3D movies and the technology has been steadily improving. One of the recent successes is "Chicken Little" shown in 3D using special digital projectors and polarized glasses to separate the left and right eye view. Alternative technologies use synchronized/active eyeglasses to present the proper images to the left and the right eye. The successes of the 3D movies, however, have not translated into successes for 3D TVs. The primary obstacles have been the quality of the displays and the cost of deployment. The TV industry had been experimenting with 3D TV by offering some special programming in 3D. In November 2005, NBC showed portions of an episode of "Medium" in 3D. To experience the 3D program, the glasses were distributed in that week's issue of *TV Guide* magazine. The world had seen another 3D TV experiment during the summer of 2006 World cup Soccer in Germany. The soccer matches were expected to be simulcast in 3D for users equipped with autostereoscopic displays with eight view points. Autostereoscopic displays provide a 3D experience without any need for glasses, and the advances in autostereoscopic display technology is expected to be one of the key drivers for the successful deployment of 3D TV.

One of the reasons for the lack of success of the 3D TV so far is the ease of use and viewing comfort. Most of the displays today use standard TV with anaglyph video and a pair of glasses to generate 3D perception. Watching such TV for long periods causes eye strain. Even the current generation autostereoscopic displays have limited viewing angle and are not suitable for viewing for longer periods. The applications where 3D TV have had reasonable success are the applications where viewing comfort is secondary to the objective;

applications such as security, medicine, design automation, and scientific visualization.

The digital video revolution launched by the MPEG-1 and MPEG-2 video coding standards also resulted in an active 3D and multiview video coding research [1, 2]. The MPEG-2 multiview profile is a form of temporal scalability that encodes left view of the stereo pair as a base layer and the right view is coded as a temporal enhancement. The existing studies on the quality of 3D video are based on MPEG-2 view coding and not directly applicable to the H.264 based coding that is expected to be used in 3D TV services [3]. The studies also did not use autostereoscopic displays that are expected to be the dominant display types for 3D TV. The MPEG-2 based coding is inefficient compared with the H.264 based view coding; furthermore, the coding artifacts in MPEG-2 and H.264 are different and are likely to have different effects on the 3D perception. One of the goals of our work in 3D coding area is to understand the impact of the coding artifacts on asymmetric view coding. Since each compression algorithm causes different artifacts, especially in the view coded at a lower bitrate, the effect of the compression algorithm on the quality of the 3D video would be different. The quality of 3D video experiences is also influenced by the type of displays used. A good summary of the perceptual quality requirements and evaluations for 3D video is presented in [4]. Our current focus is on developing efficient coding and representation algorithms for 3D and multiview video. We are using H.264 as the basis for view coding and autostereoscopic displays for rendering the 3D video.

In this article we discuss the key challenges to the success of 3D TV and discuss ways to overcome some of these challenges. We describe our ongoing research in the area of 3D and multiview video coding with emphasis on video coding for 3D TV that exploits the potential of asymmetric view coding. We report the results of a user study to evaluate the quality of 3D video on autostereoscopic displays with H.264 coded stereo views. The study provides an understanding of the bounds of the asymmetric coding and additional bandwidth necessary to provide 3D TV services.

3D TV Challenges

While public fascination with 3D technology remains, the success of the 3D TV will depend on the user adoption and the willingness of service providers to deploy new services. The key factors that influence the success of 3D TV are: (i) effortless viewing, (ii) content availability, (iii) compatibility with the existing infrastructure, and (iv) role of multiview video coding.

Effortless Viewing

Viewer comfort is the most important factor that can make or break the 3D TV. Depth perception, new displays, and experiences all affect the viewing experiences. TV watching is different from watching movies and so is 3D TV experience from 3D movie experiences. Users tend to watch TV for longer periods of time with breaks in between while movies are on average less than 1 ½ h long. Most of the users also tend to do other activities while watching TV. These factors rule out the use of glasses to experience 3D TV. Active eyewear or other types of glasses would be cumbersome and "unnatural" for users to wear continuously. Users will expect to watch some videos in 2D and the display technologies have to support this seamless mode switching.

There are a variety of 3D display technologies that are still in their early stages of development [5]. The complexity and cost considerations make most of these technologies unsuitable for 3D TV applications at this time. Autostereoscopic displays are good candidates to provide realistic 3D experiences in the home environment. Autostereoscopic displays based on lenticular imaging have been relatively successful. Display manufacturers such as Sharp and Philips have developed autostereoscopic LCD displays based on lenticular techniques at a reasonable cost. These first-generation displays have restricted viewing angles and hence restrict the movement of users. This is expected to change with more viewing freedom as the technologies mature.

The effortless viewing experiences also depend on the 3D content and how the usage patterns evolve. There have been no long-term studies on the impacts of 3D TV on user experiences, usage patterns, and most importantly, vision. One of the common complaints about 3D experiences is eye strain. Eye strain is attributed to a number of factors including the type of displays, content, and the constant vergence/accommodation adjustments in the human eye. There has been work on developing algorithms that attempt to reduce the strain on the eye when watching 3D videos [6]. The long-term studies are necessary to understand the effects of 3D viewing on the eye strain and human vision. Users can be expected to adjust their viewing patterns based on their personal experiences.

Content Availability

Content availability is another factor that will influence the success of 3D TV. Content production is a difficult task and it is even more difficult to create 3D effects in the content that are pleasing to the eye. Since most of the available TV content today is available as 2D video, ability to convert 2D content to 3D will make large amounts of 3D content available. Dynamic digital depth (DDD) is a pioneer in this area and has been successful in developing 2D to 3D conversion [7]. While 2D to 3D conversion makes available large amounts of content in 3D quickly, such converted content was not originally created for 3D viewing and effectiveness of such content is unclear.

Infrastructure Compatibility

The 3D TV deployment should be seen as an evolution similar to the black and white to color TV migration. The compatibility with the existing systems cannot be compromised. The TV industry is currently beginning a slow migration from analog to digital and to HD TV. Introduction of 3D TV will perhaps begin after migration to digital/HD TV, which will put away 3D TV deployment at least for 5 years .

Technology compatibility is another factor that has to be considered. The two main approaches to delivering 3D video are (i) stereo coding where the left and right views are encoded and (ii) depth image based rendering (DIBR) where a single view and an associated depth map are transmitted to the receiver [8]. DIBR systems generate the left and right views at the receiver based on the single view and the depth information. These two approaches have their advantages and disadvantages. However, from a production and compatibility point of view the stereo coding methods are more suitable as studios do not need new depth-sensing cameras. Furthermore, the free viewpoint TV (FTV) based on multiview video coding (MVC) is gaining momentum and this makes DBIR approaches unnecessary as the MVC would enable view synthesis to generate the left and right views necessary for the 3D TV.

The Deployment Costs and Benefits

The success of the 3D TV depends to a large extent on cost of deployment and additional revenues that will justify the investments. Hardware manufacturers will have an incentive as they can sell more hardware. However, the TV networks and content providers have a harder challenge. Advertisements in 3D are expected to be the initial drivers as it is generally believed that a 3D ad will make a better impression on the consumers. Broad studies supporting this general belief are yet to be conducted. In April 2006, Philips announced a 42″ flat panel 3D TV with a price tag of $18,000 intended for advertisements and demos by commercial establishments. Another recent development is a 3D advertising network launched in Thailand that is expected to be deployed in other countries. These developments are a pointing in the direction of advertiser-driven 3D TV.

The Role of Multiview Video Coding

Video coding technologies have matured sufficiently over the last few years to make possible new generation of video applications. Multiview video coding has been receiving significant renewed attention among researchers and the industry [9]. Multiview video coding (MVC) is also being standardized by the MPEG committee [10]. The goal of MVC is to allow coding of multiple camera views such that the user has the freedom of choosing the view point within a small field of view. The biggest challenge here is in developing compression technologies that can exploit the redundancies among the multiple views to achieve a high compression ratio.

The 3D video can be seen as a subset of multiview video coding where the stereo pair required for the 3D TV can be synthesized at the receiver. Multiview video systems require a multiview decoder at the receiver and can use existing TV monitors. Since the multiview video encoding is expected to be based on H.264 video coding, the incremental cost of a multiview receiver is small. These factors perhaps will result in faster deployment of multiview video compared with the 3D. With 3D being a special case of multiview with two coded views, migration from multiview to 3D will be an incremental step. Efficient compression of the multiple views or stereo views continues to be important. For the special case of stereo views intended for 3D TV, the human visual may be exploited to improve the compression efficiency further.

Efficient 3D Coding and Quality Evaluation

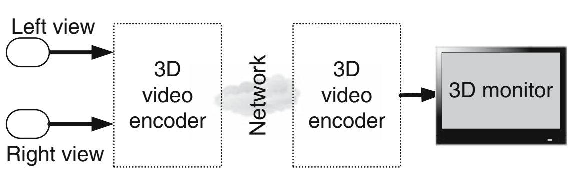
The human visual system has the property of binocular mixture where the left and right eye views are compared and combined to generate a single 3D percept.

The binocular mixture makes possible encoding of left and right eye views at different bitrates. This asymmetric view coding has been exploited before to improve the compression efficiency [3, 11]. The H.264 video coding used in our system is much more efficient than MPEG-2 and also has support for de-blocking that improves the perceptual quality of video. The effects of these improved compression algorithms on video quality cannot be understood from the past MPEG-2 based studies.

Figure 1 shows the general architecture of a 3D video system used to evaluate the efficiency of the asymmetric view coding. The stereo views are encoded at the sender by exploiting the large amount of redundancies among the views. We use H.264 as the core compression engine with inter-view prediction to increase compression efficiency [12]. The coded views are communicated to the receiver where the decoded views are rendered on an appropriate display. The 3D displays use a pair of coded views to display 3D video with depth perception.

We used the Sharp LL-151–3D autostereoscopic display that uses lenticular imaging techniques to test the rendering of stereoscopic videos. The display is 15-inches, XGA resolution (1024 by 768 pixels). The perception of depth is achieved by a parallax barrier that diverts different patterns of light to the left and right eye. It should be noted that our player architecture accommodates a variety of formats for 3D playback and can be extended to include others.

We are currently conducting a large user study to evaluate the impact of asymmetrically coded 3D views on the quality of the 3D video rendered on the Sharp autostereoscopic display. The goal of this study is to understand the bounds of asymmetric coding, relationship between the eye dominance and 3D quality of asymmetrically coded video, and to understand the effects of the H.264 coding artifacts. The results are reported based on the evaluations from 14 users who have evaluated the subjective quality.



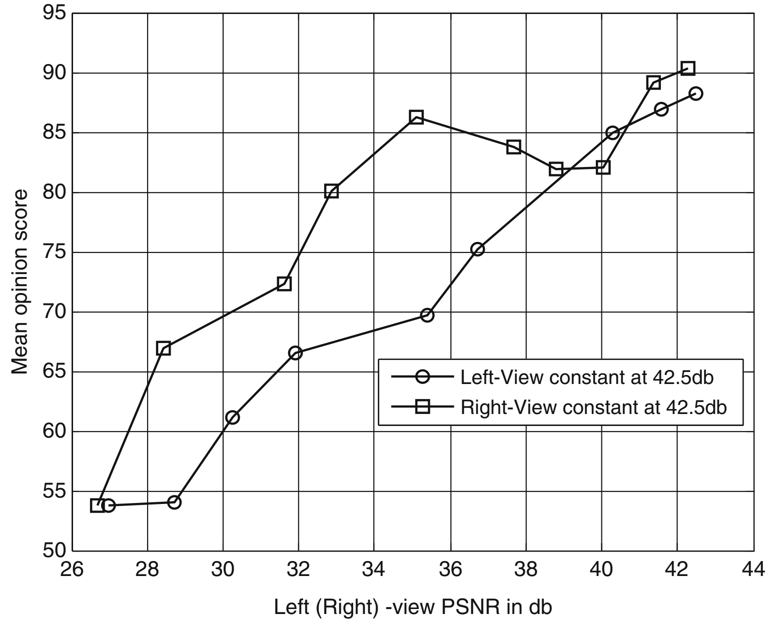
Three Dimensional Television Services. Figure 1. 3D/Multiview Video System.

The subjective video quality evaluation was done using two stereo video sequences (Akko & Kayo and Ballroom). The left and right eye views are encoded at variety of qualities. The test sequence pairs were created with one view at a high quality and the other at lower qualities. Each 3D sequence is 10 s long with a 5-s mid-gray displayed between the sequences. The sequences at different qualities are shown in a random order on the 15-inch Sharp autostereoscopic 3D displays. The users evaluated the subjective quality on a scale of 1 (unacceptable) to 100 (excellent). Figure 2 shows the mean opinion scores (MOS) for the Akko and Kayo sequence with one view kept at a constant quality while the quality of the other view is varied. The figure shows a plot of MOS versus the PSNR of the left (right) view while keeping the right (left) view PSNR constant at 42.5 dB.

The results show that asymmetric coding can be exploited in encoding 3D video sequences. The MOS is significantly higher when the right eye view is coded at a higher quality. This may be due to the eye dominance as most people are right-eye dominant and hence perceive a better quality when the right eye view is at a higher quality. The role of eye dominance in visual perception is not well understood. A recent study found that eye dominance improves the performance of visual search tasks by perhaps aiding visual perception in binocular vision [13]. Further evaluation with additional sequences and careful user screening is necessary to fully validate our claim.

Conclusion

3D TV has the potential to become the next big revolution in television. The realization of this potential is influenced by the technological advances as well as the human factors. The key factors that influence the success of 3D TV are: (i) effortless viewing, (ii) content availability, (iii) compatibility with the existing infrastructure, and (iv) role of multiview video coding. Autostereoscopic displays are likely to become the dominant mode of displaying 3D video. The autostereoscopic displays based on lenticular imaging techniques have advanced sufficiently to be able to provide large-scale user tests for 3D TV services. Larger-scale studies are necessary to gain proper understanding of the 3D TV experiences.



Three Dimensional Television Services. Figure 2. Mean opinion scores for asymmetric view coding.

References

1. B.L. Tseng and D. Anastassiou, "Multi-Viewpoint Video Coding with MPEG-2 Compatibility," IEEE Transactions Circuits and Systems for Video Technology, Vol. 6, No. 4, August 1996, pp. 414–419.
2. A. Puri, R.V. Kollarits, and B.G. Haskell, "Stereoscopic Video Compression Using Temporal Scalability," Proceedings of the SPIE Visual Communications and Image Processing'95, Taiwan, May 1995.
3. L.B. Stelmach and W.J. Tam, "Stereoscopic image coding: Effect of disparate image-quality in left- and right-eye views," Signal Processing: Image Communication, Vol. 14, 1998, pp. 111–117.
4. L.M.J. Meesters, W.A. Jsselsteijn, and P.J.H. Seuntiens, "A Survey for Perceptual Evaluations and Requirements of Three-Dimensional TV," IEEE Transactions On Circuits Systems and Video Technology, Vol. 14, No. 3, March 2004, pp. 381–391.
5. N.A. Dodgson, "Autostereoscopic 3D Displays," Computer, Vol. 38, No. 8, August 2005, pp. 31–36.
6. J.S. McVeigh, M.W. Siegel, and A.G. Jordan, "Algorithm for automated eye strain reduction in real stereoscopic images and sequences," Proceedings of the SPIE Vol. 2657: Human Vision and Electronic Imaging, pp. 307–316, 1996.
7. P. Harman, J. Flack, S. Fox, and M. Dowley, "Rapid 2D to 3D conversion," Proceedings of the SPIE, Vol. 4660, pp. 78–86, 2002.
8. C. Fehn, "A 3D-TV approach using depth-image-based rendering (DIBR)," Proceedings of VIIP 03, Sept. 2003.
9. ISO/IEC JTC1/SC29/WG11, "Survey of Algorithms used for Multi-view Video Coding (MVC)," MPEG Document MPEG2005/N6909, January 2005.
10. ISO/IEC JTC1/SC29/WG11, "Call for Proposals on Multi-view Video Coding (MVC)," MPEG Document MPEG2005/N7327, July 2005.
11. H. Kalva and B. Furht, "Hypercube based inter-view prediction for multi-view video coding," Proceedings of the Second Workshop on Immersive Communication and Broadcast Systems (ICOB), October 2005.
12. D.V. Meegan, L.B. Stelmach, and W.J. Tam, "Unequal Weighting of Monocular Inputs in Binocular Combination: Implications for the Compression of Stereoscopic Imagery," Journal of Experimental Psychology: Applied, Vol. 7, No. 2, June 2001, pp. 143–153.
13. E. Shneor and S. Hochstein, "Effects of Eye Dominance in Visual Perception," International Congress Series, Vol. 1282, Vision 2005, September 2005, pp. 719–723.

Three Dimensional TV

► Three Dimentional Television Services: Introduction

Three Dimensional Video Compression

► Hypercube-Based Multi-View Video Coding

Three Dimensional Videos

▶ Three Dimensional Television Services: Introduction

Three-Dimensional Video Visualization and Navigation System

Borko Furht
Florida Atlantic University, Boca Raton, FL, USA

Synonyms

▶ Navigation of Video Data

Definition

This article describes an interactive three-dimensional system for visualization and navigation of video data, which is based on hypercube structures. The described interactive system allows content creators to connect video data in space and time as well as to combine video data from different sources.

Introduction

The recent advances in Internet and video technology, including mobile and wireless systems allow us to receive a high-quality resolution video on our desktops as well as on the variety of mobile systems. High-speed networks provide high bandwidth, which will soon allow us to receive multiple real-time videos. We are at the verge of the Web-based video revolution, in which we will begin receiving and processing video applications in a completely new way. Besides just watching videos, the users are also becoming content creators; they create their own videos, upload videos on their own channels.

There are a number of companies, which are currently offering a variety of video-based applications, predominately in the area of entertainment. Basically, they all allow a similar functionality and similar user experience. Various videos are stored on the company's servers and a set of videos are presented sequentially (or linearly) on the selected Webpage. The user can click on a selected video, a larger window will open, and the user can then watch the video. There is a VCR-type of functionality for controlling the video (Start, Stop, Forward, etc.). Examples of screen shots from several companies are shown in Figs. 1 and 2.

There is a simple navigation through the system by clicking various categories, or by using Search function to find and retrieve videos using key words. However, this sequential (or linear) structure does not provide exciting user's experience, especially for young generation, which grew up with video games. There is no specific links between various videos, which connect topics in space and time.

There are a number of novel video visualization techniques proposed in literature, which provide more effective visual analysis of large-scale videos [1–5]. This article describes a hypercube-based system,

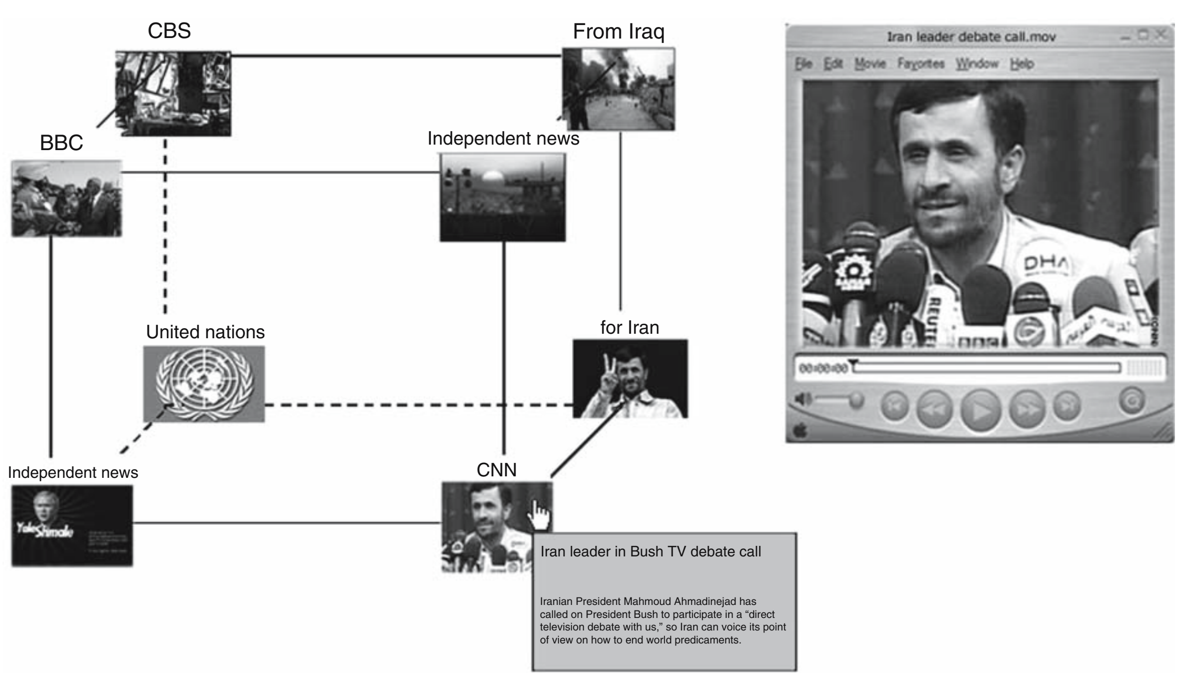
Three-Dimensional Video Visualization and Navigation System. Figure 1. Two screen shots from www.youtube.com.

which allows an effective visualization, navigation, and retrieval of group of video clips [6].

Hypercube-Based User Interface

The basic element of the described method for creating an interactive user interface is an interactive cube with eight nodes, as presented in an example in Fig. 3. Each node is linked to a specific video (or eventually to another level of a cube interface). In this specific example, the interactive cube presents a set of current news videos on Iraq war, combined from different parts of the world. Each node contains a small video icon. The content creator can combine videos from different sources and create a three-dimensional view

Three-Dimensional Video Visualization and Navigation System. Figure 2. Screen shots from www.current.tv and www.lulu.tv.



Three-Dimensional Video Visualization and Navigation System. Figure 3. (a) Cube-based interface, (b) Information about the video, and (c) The user plays a video.

of videos. The user can navigate through the cube space and the cube will rotate accordingly. As the user traverses a node (video icon) with the mouse, textual information about the related video will appear on the screen, as indicated in Fig. 3. The user can select and play the video by clicking on a node. The selected video will begin playing on the screen, as illustrated in Fig. 3.

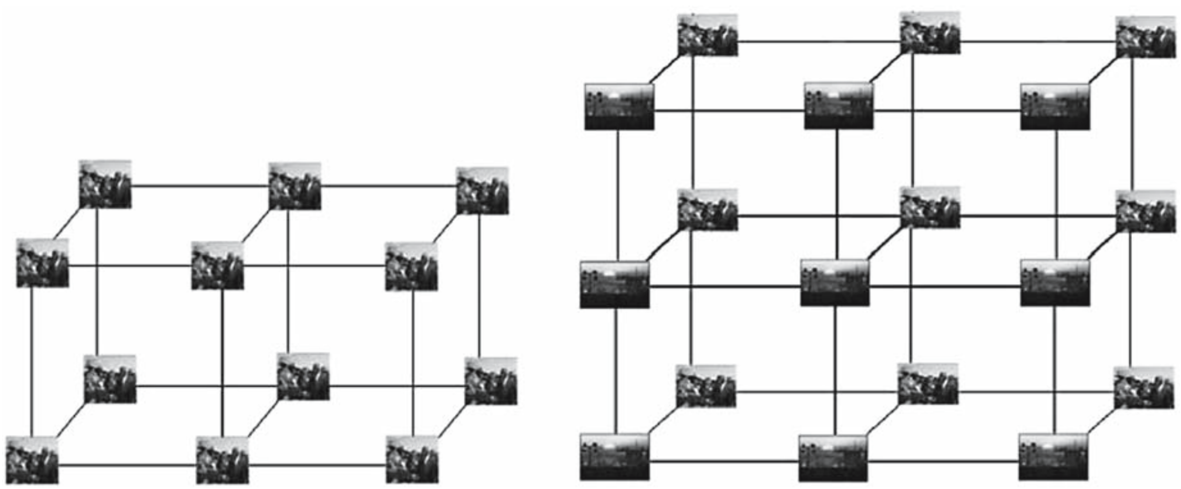
Similarly, a time cube can be created, where video clips on the same topic are linked in time. The user can travel through the cube in time and select and watch the video.

Hypercube Extensions

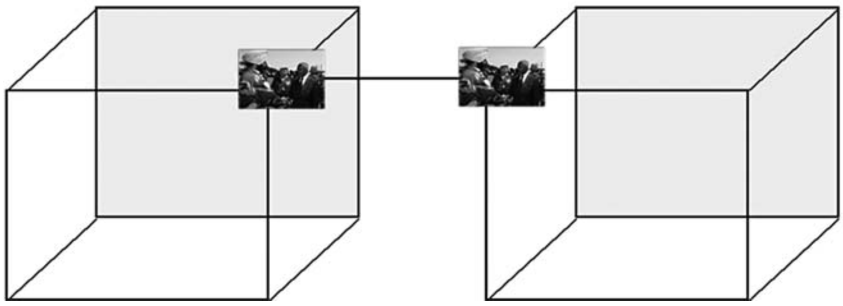
The basic cube can be extended to provide more information (additional videos and additional multidimensional links). This allows an effective visualization of a large number of video data, including large video archives. The user gets more exciting experience in traveling through the space or time and looking for videos.

Figure 4 shows a dual cube with 12 nodes and quadratic cube with 18 nodes. These structures allow more than eight videos to be linked together in space or time.

Another possible extension is connecting two or several cubes and building hypercubes. An example in Fig. 5 illustrates connecting two cubes. A possible scenario is that from the video clip, which connects these two cubes, one cube to the left is a space-connected cube and another on the right is a time-connected cube. This method allows content creators to efficiently



Three-Dimensional Video Visualization and Navigation System. Figure 4. Extensions of the basic cube.

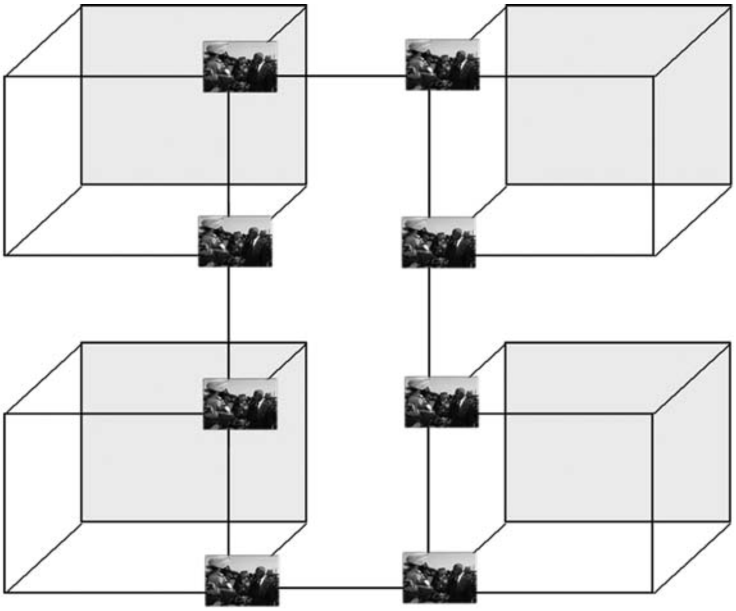


Three-Dimensional Video Visualization and Navigation System. Figure 5. Hypercube extension consisting of space and time cubes.

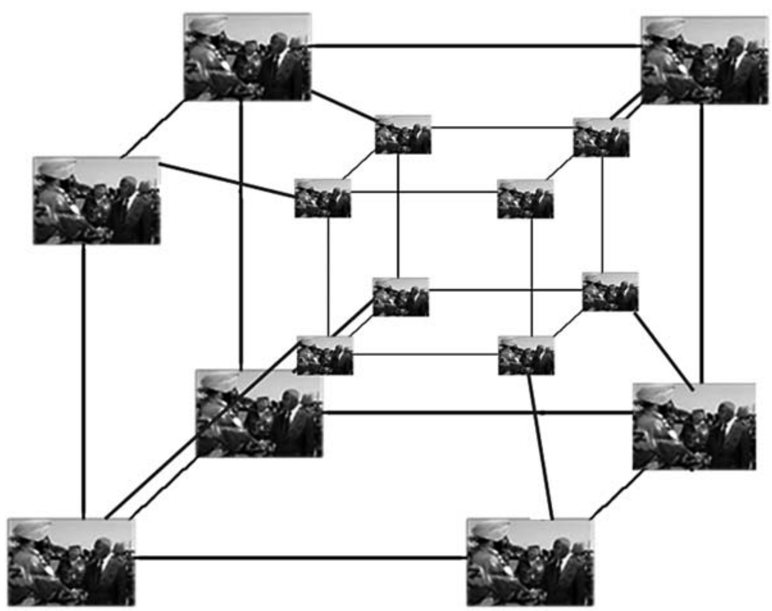
create, combine, and link video clips in space and time. The user can then travel through the space and time, and look for and select related video clips.

This model can be further extended to multi-cube, as illustrated in Fig. 6. Finally, multidimensional hypercubes can be created, as illustrated in Fig. 7. Two-dimensional cube in Fig. 7 provides an interesting link of an inner and outer cube, where related nodes (video clips) are interconnected.

The basic principles of navigation through the proposed hypercube structures are already described in Sections ”Hypercube-Based User Interface” and



Three-Dimensional Video Visualization and Navigation System. **Figure 6.** Multilevel hypercube.



Three-Dimensional Video Visualization and Navigation System. **Figure 7.** Two-dimensional hypercube.

"Hypercube Extensions." The user travels through the hypercubes in space and time, gets basic descriptions of the videos in the hypercube, and selects the video for view. The hypercube model allows the user to get the comprehensive view of available videos on specific topics, their space and time relationship.

Applications

In the emerging field of Internet applications relating to entertainment and news, there is a great number of applications which can take advantage of the proposed hypercube-based user interface. Practically, all the companies, which offer video applications (entertainment, news, etc.) currently use tremendous number of videos. The hypercube structure can help them to better organize and link these videos by their theme, and in space and time, and present to the user. The user will have a completely new experience in traveling through time and space hypercubes, and selecting and watching the videos.

The interactive system provides a new experience both for content creators and for users. Content creators can organize and visualize a complex video data structure in a well-structured way, and can combine video data in space and time. The users can easily navigate through a number of video clips in space and time.

References

1. Y. Wang, D.M. Krum, E.M. Coelho, and D.A. Bowman, "Contextualized Videos: Combining Videos with Environment Models to Support Situational Understanding," IEEE Transactions on Visualization and Computer Graphics, Vol. 13, No. 16, November/December 2007, pp. 1568–1575.
2. Y.-F. Ma and H.-J. Zhang, "Video Snapshot: A Bird View of Video Sequence," Proceedings of the 11th International Multimedia Modeling Conference, 2005.
3. H. Luo, J. Fan, and J. Yang, "Exploring Large-Scale Video News via Interactive Visualization," Proceedings of the IEEE Symposium on Visual Analytics Science and Technology, Baltimore, MD, pp. 75–82, October/November 2006.
4. A. Akutsu, Y. Tonomura, and H. Hamada, "Video Interface for Spatiotemporal Interactions Base don Multi-dimensional Video Computing," Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 191–194, April 1997.
5. A.C. Gallo, "Method of Representing Information on a Three-Dimensional User Interface," US Patent No. 2002/0163546 A1, November 2002.
6. B. Furht, "Interactive System for Visualization and Retrieval of Video Data," Provisional US Patent, 2007.

Three GPP

► Networking for Massively Multiuser Online Gaming

Threshold Schemes with Minimum Pixel Expansion

Definition

In visual secrete sharing schemes, attempts have been made to restrict the pixel expansion while images with high contrast are produced.

Although results obtained via a (k,n)-visual secret sharing (VSS) scheme can be easily revealed by a human observer without any computations, the size of shares is usually much larger compared to the size of the original input image. By encrypting the secret pixel into m_1m_2 blocks of share pixels, the expansion is determined by the so-called expansion factor m_1m_2 which denotes the number of columns of the $n \times m_1m_2$ basis matrices. In the recent past, attempts have been made to restrict the pixel expansion while images with high contrast are produced. The designer most often has to trade pixel expansion constraints to the recovered image quality and/or the order of (k,n)-threshold schemes [1–3].

A probabilistic VSS scheme which offers no pixel expansion ($m_1m_2 = 1$) which can be regarded as a VSS scheme with the minimum pixel expansion is briefly reviewed in the sequence. The reconstruction of the secret image is probabilistic. In a deterministic scheme, any given secret pixel is encrypted into m_1m_2 subpixels. Hence the reconstructed image is m_1m_2 times larger in size compared to the original binary input. To obtain no pixel expansion, each secret pixel is reconstructed with a single pixel in the probabilistic scheme. The binary secret pixel is correctly reconstructed with a given probability. However the quality of the reconstructed binary image depends on how large the image areas of pixels showing the same binary value are.

A probabilistic VSS scheme can be constructed using the two sets, white set P_1 and black set P_0, each consisting of $n \times 1$ matrices, respectively. When sharing a white (or black) pixel, the dealer first randomly chooses one $n \times 1$ column matrix in P_1 (or P_0), and then randomly selects one row of this column matrix to a relative shadow. The chosen matrix defines the pixel color in every one of the n shadows. A probabilistic VSS scheme is considered valid if the contrast and security conditions are met [2].

An image secret sharing (ISS) scheme in [3] can be viewed as an alternative to minimal pixel expansion based VSS schemes. The scheme is a (2,2)-ISS solution suitable for cost-effective encryption of the natural images. It encrypts the decomposed "black" bit of the original secret pixel into a black (or white) bit in each of the two shares. To differentiate between the "black" and "white" bits of the original secret pixels, the decomposed "white" bits are encrypted into black and white (or white and black) pixels in the corresponding shares. Such an approach satisfies the essential perfect reconstruction property and can serve as the private-key cryptosystem.

Cross-References

- ► Image Secret Sharing
- ► Private-key Cryptosystem
- ► Visual Cryptography

References

1. P.-A. Eisen and D.-R. Stinson, "Threshold visual cryptography schemes with specified whiteness levels of reconstructed pixels," Designs, Codes and Cryptography, Vol. 25, No. 1, January 2002, pp. 15–61.
2. C.-N. Yang, "New Visual Secret Sharing Schemes Using Probabilistic Method," Pattern Recognition Letters, Vol. 25, No. 4, 2004, pp. 481–494.
3. R. Lukac and K.-N. Plataniotis, "A New Encryption Scheme for Color Images," Computing and Informatics, submitted.

Trends in Multimedia E-Commerce

Definition

Trends in multimedia e-commerce will always be a function of market factors and the cultural landscape. As such, several surprising and lucrative trends have dominated land-line and wireless multimedia e-commerce in recent times.

Ringtone E-Commerce

Ringtones are the sounds that are activated on cell phone when incoming calls arrive. In early cell phones tones were limited to a small set of just nine. Today's cell phone have the capability to play *monophonic* (a single voice), *polyphonic*, and *sampled* (e.g. clips of actual recordings) ringtones. Initially a sort of "geek sub-culture" activity, ringtone e-commerce broke-out in 2003 as service providers and third party communications providers established e-commerce sites for ringtone download. Some experts estimate that consumers put down near $2 billion worldwide in that year alone for tones [1]. Ringtone e-commerce revenue generally gets split between many business entities. Some of these entities are: the wireless provider, the music label, publishers, and the third party technology middlemen who provide [1]. Validating the fact that ringtone e-commerce has come of age is the news that Billboard magazine now charts the 20 top-selling ringtones in its "Hot Ringtones" chart.

Other ringtone-related trends are currently in the pipeline. One is the trend towards the insertion of humorous "samples" into conversations (e.g., the voice of a celebrity impersonator). Another newer trend for 2005 is the combination of pleasing images to go along with a ringtone. For example, a pretty image such as a sunset of landscape pops up on the phones LCD screen simultaneously with the ring. Such clips and graphics will enable a similar download economy.

Audio and Music

Although music download e-commerce was happening before Apple introduced its *iPOD* line of mp3 players in 2001, Apple's chic devices, marketing approach, and easy to use online (legal) music download site called *iTunes* are widely seen as sea-change events. Over 10 million iPODS have been sold since 2001 and while 2003 saw about 20 million legal song-downloads, 2004 saw 200 million. Research firms believe that 2004s $330 million market for online music will double in 2005. While iTunes gets most press, there are many other ways to legally acquire audio and music online including: Napster, Real, download.com, eMusic, Wal-mart.com, mp3.com, and Amazon.com. Such stores replace the "first generation" of illegal file sharing systems which allowed users to trade high-quality versions of songs online at no cost. Illegal file sharing ultimately led to the Recording Industry Association of America's (RIAA) much-publicized lawsuits with individual uploaders. RIAA has thus far targeted principle sources (e.g., those users with more than 1,000 songs on their hard drives) rather than casual users; however, their efforts are aimed to deter future abusers of copyrighted media.

Cross-References

► Online Multimedia E-Commerce

References

1. M. Maier, "The Changing Ringtone Economy, in Business 2.0," Available online at: http://www.business2.com, October 2004.

Two-level Meta Data Management

► Availability and Scalability of Lomss