Video

YING LI
IBM T.J. Watson Research Center, Hawthorne,
NY, USA

Definition

Video, which means ''I see'' in Latin, is an electronic representation of a sequence of images or frames, put together to simulate motion and interactivity. From the producer's perspective, a video delivers information created from the recording of real events to be processed simultaneously by a viewer's eyes and ears. For most of time, a video also contains other forms of media such as audio.

Video is also referred to as a storage format for moving pictures as compared to image, audio, graphics and animation.

Historical Background

Video technology was first developed for television systems, but it has been further developed in many formats to allow for consumer video recordings. Generally speaking, there are two main types of video: analog video and digital video. Analog videos are usually recorded as PAL (Phase Alternating Line) or NTSC (National Television System Committee) electric signals following the VHS (Video Home System) standard and stored in magnetic tapes. Digital videos, on the contrary, are usually captured by digital cameras and stored in digital video formats such as DVD (Digital Versatile Disc), QuickTime and MPEG-4 (Moving Picture Experts Group).

Launched in September 1976, VHS became a standard format for consumer recording and viewing by the 1990s. Since then, it has dominated both home and commercial video markets. In March 1997, the DVD format was introduced to American consumers, which gradually pulled consumers away from VHS in the following years due to its much better quality. In June 2003, the DVD's market share exceeded that of the VHS for the first time. Since then, it has been steadily expanding its consumer market, and by July 2006, most major film studios have stopped releasing new movie titles in VHS format, opting for DVD-only releases. Now, VHS is gradually disappearing from both rental and retail stores, and DVD has dominated the whole commercial market. Nevertheless, VHS is still popular for home recording of television programs, due to the large installed base and the lower cost of VHS recorders and tape.

For the last few decades, as video technology quickly advances and the cost of storage devices rapidly decreases, digital videos have become widely available in diverse application areas such as medicine, remote sensing, entertainment, education and online information services. This has thus led to very active research in various video-related areas.

Foundations

The last three decades have witnessed a significant amount of research efforts on various aspects of video technologies. Roughly speaking, they fall into the following three general categories: video representation, video content analysis, and video application. Specifically, video representation deals with the way a video is represented, in another word, the file format. Video content analysis, on the other hand, aims to automatically structure and ultimately understand the video by analyzing its underlying content. Due to the difficult nature of this problem, such process usually involves the analysis of multiple media modalities including visual, audio and text information. Finally, video application applies what is has learned from the analysis engine, and facilitates various types of content access including video browsing, summarization and retrieval. A brief discussion on each of these three research domains is given below.

Video Representation

A video sequence with accompanying sound track can occupy a vast amount of storage space when represented in digital format. As estimated in [6], a 1-min video clip could possibly occupy up to 448 MB. Consequently, compression has been playing an important role in modern schemes for video representation.

A wide variety of methods has been proposed to compress the video stream. Nevertheless, almost all of them build their approaches upon the fact that video data contains both spatial and temporal redundancy. Specifically, to reduce the spatial redundancy, an intra-frame compression is applied which registers differences between parts of a single frame. Such a task is more closely related to image compression. Likewise, to reduce the temporal redundancy, an inter-frame compression is exploited which registers differences between neighboring frames. This involves discrete Cosine transform (DCT), motion compensation and other techniques. Some popular video compression mechanisms include H.261, H.263, H.264, MPEG-1, MPEG-2, MPEG-4 and MJPEG (Motion-Joint Photographic Experts Group). Specifically, H.261 is a 1990 ITU-T (Telecommunication Standardization Sector of International Telecommunication Union) video coding standard originally designed for transmission over

ISDN lines. Later on, H.263 and H.264, which provide more capabilities and mainly target at video-conferencing applications, were standardized in 1995 and 2003, respectively. In 1998, the Moving Picture Experts Group (MPEG) was formed to establish an international standard for the coded representation of moving pictures and associated audio on digital storage media. Currently, there have been three established MPEG standards from this effort: MPEG-1, MPEG-2, and MPEG-4. Each of them targets at different commercial applications. For instance, MPEG-1 is usually used as the Video CD (VCD) format, MPEG-2 for High Definition Television (HDTV), and MPEG-4 for streaming video applications. Finally, to facilitate mobile appliances such as digital cameras, MJPEG was developed in 1990s which uses intra-frame coding technology that is very similar to those used in MPEG-1 or MPEG-2. However, it does not use inter-frame prediction, which on one hand, results in a loss of compression capability, yet on the other hand, it makes the degree of compression capability independent of the amount of motion in the scene. Moreover, it also eases video editing as simple editing can now be performed at any frame.

Video Content Analysis

Video is a type of rich media as it often consists of other media types such as audio and text. Consequently, research on video content analysis can be grouped into three classes: visual content analysis, audio content analysis, and audiovisual content analysis. A general goal of video content analysis is to extract the underlying video structure so as to facilitate convenient and nonlinear content access. Yet a more aggressive goal is to automatically understand video semantics so as to support applications such as video summarization and retrieval that require an in-depth understanding of the video content.

Video Application

Besides the large amount of research efforts on video content analysis, there are also many attentions on studying various video applications. After all, making the bulky and unstructured video content convenient and efficient to access, present, share, search and deliver is the ultimate goal of the entire research community in this area.

(Abridged)

Recommended Reading

1. Cheung S. and Zakhor A. Efficient video similarity measurement with video signature. IEEE Trans. Circ. Syst. Video Tech., 13(1):59–74, 2003.

2. Li Y. and Dorai C. SVM-based audio classification for instructional video analysis. In Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, 2004.

3. Li Y. and Kuo C.-C. Video Content Analysis Using Multimodal Information: for Movie Content Extraction, Indexing and Representation. Kluwer, MA, USA, 2003.

4. Li Y., Narayanan S., and Kuo C.-C. Content-based movie analysis and indexing based on audiovisual cues. IEEE Trans. Circ. Syst. Video Tech., 14(8):1073–1085, 2004.

5. Mahmood T.S. and Srinivasan S. Detecting topical events in digital video. In Proc. 8th ACM Int. Conf. on Multimedia, 2000, pp. 85–94.

6. Mitchell J., Pennebaker W., Fogg C., and LeGall D. MPEG Video Compression Standard. Chapman & Hall, New York, NY, USA, 1992.

7. MPEG Requirements Group, MPEG-7 Applications Document v.8, ISO/MPEG N2860, MPEG Vancouver Meeting, July 1999.

8. MPEG Requirements Group, MPEG-7 Context, Objectives and Technical Roadmap, ISO/MPEG N2861, MPEG Vancouver Meeting, July 1999.

9. MPEG Requirements Group, MPEG-7 Requirements Document V.15, ISO/MPEG N4317, MPEG Sydney Meeting, July 2001.

10. Nock H., Adams W., Iyengar G., Lin C., Naphade M., Neti C., Tseng B., and Smith J. User-trainable video annotation using multimodal cues. In Proc. 26th Annu. Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, 2003, pp. 403–404.

11. Oh J. and Hua K. Efficient and cost-effective techniques for browsing and indexing large video databases. In Proc. ACM SIGMOD Int. Conf. on Management of Data, 2000, pp. 415–426.

12. Pfeiffer S., Lienhart R., Fischer S., and Effelsberg W. Abstracting digital movies automatically. J. Vis. Comm. Image Represent., 7(4):345–353, 1996.

13. Yan R., Hauptmann A., and Jin R. Negative paeudo-relevance feedback in content-based video retrieval. In Proc. 11th ACM Int. Conf. on Multimedia, 2003, pp. 343–346.

14. Yeung M., Yeo B., and Liu B. Extracting story units from long programs for video browsing and navigation. In 1996, pp. 296–305.

15. Zhang T. and Kuo C.-C. Audio content analysis for on-line audiovisual data segmentation. IEEE Trans. Speech Audio Process., 9(4):441–457, 2001.

16. Zheng W., Li J., Si Z., Lin F., and Zhang B. and Using highlevel semantic features in video retrieval. In Image and Video Retrieval. Springer, Berlin Heidelberg, New York, 2006, pp. 370–379.

Answer the following questions:

1) What is *video?*
2) What is the difference between *analog* and *digital* video?
3) What makes VHS still popular for home recording?
4) What is meant by *remote sensing?*
5) What do the terms *video representation, video content analysis,* and *video application* refer to?
6) What do most compression formats build on?
7) What is MPEG-1?
8) What does the term *rich media* refer to?

Match the following terms and their definitions:

1) video representation
2) video content analysis
3) intra-frame compression
4) inter-frame compression
5) likewise

a) a way of reducing spatial redundancy
b) deals with the file format
c) a way to reduce temporal redundancy
d) involves structuring the video
e) means the same as "similarly"

Mark the following statements as *true* or *false*:

1) Video was first developed for home use.
2) H.261 is a video coding standard originally designed for transmission over ISDN lines.
3) MPEG-4 is used as a high definition television standard.
4) MJPEG has been designed for use in mobile appliances.
5) MJPEG has nothing in common with MPEG-1 and MPEG-2 formats
6) A general goal of video content analysis is to facilitate convenient and linear content access.

Vocabulary

aim to do st [eɪm] 🔊 – snažit se něco dělat, být zaměřen na dělání něčeho

analysis [ə'næl.ə.sɪs] 🔊 – analýza, pl. analyses

appliance [ə'plaɪ.ən*t*s] 🔊 – zařízení

audio ['ɔː.di.əʊ] Ⓤ ['aː.di.oʊ] 🔊 – audio, zvuk

audiovisual [ˌɔː.di.əʊ'vɪʒ.u.əl] Ⓤ [ˌaː.di.oʊ–] 🔊 – audiovizuální

capability [ˌkeɪ.pə'bɪl.ɪ.ti] Ⓤ [–ə.t̬ i] 🔊 – schopnost

compression [kəm'pres] 🔊 – komprese

consumer [kən'sjuː.məʳ] Ⓤ [–'suː.mɚ] 🔊 – spotřebitel

to disappear [ˌdɪs.ə'pɪəʳ] Ⓤ [–'pɪr] 🔊 – zmizet, mizet

discrete Cosine transform [dɪ'skriːt] 🔊 ['kəʊ.saɪn] Ⓤ ['koʊ–] 🔊 [træns'fɔːm] Ⓤ [–'fɔːrm] 🔊 – diskrétní kosinová transformace

diverse [daɪ'vɜːs] Ⓤ [dɪ'vɝːs] 🔊 – rozdílný

domain [dəʊ'meɪn] Ⓤ [doʊ–] 🔊 – doména

dominate ['dɒm.ɪ.neɪt] Ⓤ ['daː.mə–] 🔊 – dominovat

due to st [djuː] Ⓤ [duː] 🔊 – kvůli něčemu, díky něčemu

format ['fɔː.mæt] Ⓤ ['fɔːr–] 🔊 – formát

frame [freɪm] 🔊 – rám, rámec

to involve st [ɪn'vɒlv] Ⓤ [–'vaːlv] 🔊 – zahrnovat něco

to launch st [lɔːn*t*ʃ] Ⓤ [laːn*t*ʃ] 🔊 – vypustit něco, vydat něco

linear ['lɪn.i.əʳ] Ⓤ [–ɚ] 🔊 – lineární

market share ['maː.kɪt] Ⓤ ['maːr–] 🔊 [ʃeəʳ] Ⓤ [ʃer] 🔊 – podíl na trhu modality

motion ['məʊ.ʃⁿn] Ⓤ ['moʊ–] 🔊 – pohyb

nevertheless [ˌnev.ə.ðə'les] Ⓤ [–ɚ–] 🔊 – nicméně

non-linear [non'lɪn.i.ə] – nelineární

present ['prez.ⁿnt] 🔊 – současný (compare the pronunciation with that of verb *to present*)

research [rɪ'sɜːtʃ] Ⓤ ['riː.sɜːtʃ] 🔊 🔊 – výzkum

retrieval [rɪ'triːv] 🔊 – vyhledávání, vyzvedávání

sequence ['siː.kwən*t*s] 🔊 – sekvence

spatial and temporal redundancy ['speɪ.ʃəl] 🔊 ['tem.pər.əl] Ⓤ[–pɚ.əl] 🔊 [rɪ'dʌn.dənt .si] 🔊 – prostorová a časová nadbytečnost (redundance)

steady ['sted.i] 🔊 – stálý

thus [ðʌs] 🔊 – tak, a tak

to browse st [braʊz] 🔊 – listovat, procházet něčím

to capture st ['kæp.tʃəʳ] Ⓤ [–tʃɚ] 🔊 – zachytit něco

to deal with st [dɪəl] 🔊 – zabývat se něčím

to develop [dɪ'vel.əp] 🔊 – vyvinout, vyvíjet

to ease st [iːz] 🔊 – udělat něco jednodušší, zjednodušit

to exceed st [ɪk'siːd] 🔊 – přesahovat, překročit něco

to extract st [ɪk'strækt] 🔊 – extrahovat, vytáhnout něco

to facilitate st [fə'sɪl.ɪ.teɪt] 🔊 – zjednodušit něco, umožnit něco

to fall into st [fɔːl] Ⓤ [faːl] 🔊 – spadat do něčeho

to occupy ['ɒk.jʊ.paɪ] Ⓤ ['aː.kjʊ–] 🔊 – zabírat, okupovat

to opt for st [ɒpt] Ⓤ [aːpt] 🔊 – rozhodnout se pro něco, zvolit něco

to present st ['prez.ⁿnt] 🔊 – prezentovat něco

to refer to st [ri'fəː] 🔊 – odkazovat na něco, označovat něco

to release st [rɪ'liːs] 🔊 – vypustit, vydat něco

to simulate ['sɪm.jʊ.leɪt] Ⓤ [–t̬ ɪd] 🔊 – simulovat, napodobovat

to witness st ['wɪt.nəs] 🔊 – být něčemu svědkem

ultimate ['ʌl.tɪ.mət] Ⓤ [–t̬ ə–] 🔊 – konečný, nejzazší

underlying [ˌʌn.də'laɪ.ɪŋ] Ⓤ [–dɚ–] 🔊 – základní

Phrases

consequently – následně

generally speaking – obecně řečeno

independent of st – nezávislý na něčem

likewise – podobně

on the contrary – naopak

on (the) one hand ..., on the other hand ... – na jedné straně ..., na druhé straně ...

to play an important role in st – hrát důležitou úlohu v něčem