WHY AND HOW DO WE EXPLORE FUNCTIONALITIES IN THE IMAGE/VIDEO COMPRESSED-DOMAIN?

Shih-Fu Chang

Department of Electrical Engineering &
Center for Telecommunications Research
Columbia University, New York, NY 10027, U.S.A.
sfchang@ctr.columbia.edu
http://www.ctr.columbia.edu/~sfchang

With more and more visual material used in multimedia applications, effective and efficient image indexing and manipulation techniques are needed. People have envisioned applications which will allow users to search hundred of thousands of images/videos in real time, or allow users to manipulate/edit videos faster than real time. One example of such time-critical applications are video broadcast journalists, who usually need to finish processing news projects under extremely tight deadlines. Both the tasks of searching large video archive to get the right material and editing them into the final video sequence need to be completed within just few minutes. or even seconds.

The biggest obstacle for the above task is the large data amount and rate of visual material. One hour of uncompressed digital video (say CCIR601 digital TV) is more than 100 GB. Using MPEG compression, we will still have about 1GB per hour. Manipulation of such a large amount of data in real time usually needs high-end dedicated hardware processors. Speed and storage are two issues of the highest concern in dealing with such real-time video applications.

Searching in a large image/video database also stresses the technology to the extreme. Drawbacks of traditional keyword or manual semantic descriptions have been well known. A recent approach is to explore the content-based approach [2], which allows users to find subjectively similar images from the database based on the visual content, e.g., object shape, color, texture, layout, motion, and temporal relationships. One critical task in this approach is image content (or feature) extraction in an automatic or semi-automatic fashion. Content extraction may be done off-line. But in a large dynamic image/video archive in which new material keeps coming in (e.g., satellite picture database), the issue of processing speed is still critical.

One promising direction for this critical task is to explore solutions in the image/video compressed domain. In applications where videos and images are compressed, there will be great advantages if content extraction and video manipulations can be done directly in the compressed domain. Compressed videos and images do not need to be decoded back to original pixel-by-pixel files, which consume large space and computing power. Also, if re-encoding of images/videos is not needed, visual quality will be maintained at the original level without causing additional degradation introduced by the re-encoding process.

Compressed-domain image technologies have attracted much interest recently, including our continuing research work. Manipulation functions such as geometrical transformation and video compositing have been developed in the JPEG-based compressed domain [3, 4] and MPEG-based compressed domain [3]. We have also demonstrated techniques for compressed-domain nonlinear video editing [6]. Feature extraction from the compress images/videos have been demonstrated for content-based query as well [5,6]. How are these functions achieved in the compressed domain? Usually answers are not trivial even for simple functions. One example is simple cut and paste. Cut and paste of images may cause block structure misalignment in the JPEG and MPEG compressed domain. Cut and paste of videos may cause buffer overflow or underflow in the MPEG-compressed domain. Solutions need to be provided to handle these issues (which may not have equivalent counterparts in the uncompressed domain).

Then it comes the interesting and important question: will it pay off to pursue the compressed-domain direction? It depends. The factors we need to carefully consider include the specific functions, the cost model of resources (speed, space, quality), and the compression algorithms. We can achieve very impressive improvement (speedup of over 10 or even 100) by using the compressed-domain techniques if the requested function matches the characteristics of the compression algorithm. One case is image overlap (i.e. picture in picture) at restricted locations. Another case is extracting using video features (e.g., scene changes, zooming, panning, scene grouping) from motion vectors in the MPEG compressed stream. However, for sophisticated operations and visual content, it will be hard to find efficient, accurate counterparts in the existing compressed domain. Examples in this category include special, fancy image morphing effects and accurate object motion.

We said above that content access functionality is needed and all current compression technologies impose various constraints. One one hand, finding efficient solutions to handle these constraints leads to exciting research. On the other hand, one would hope to have a fundamental solution — design a new compression technique that gives us easy content access! Fortunately, the need of content access functionality has gradually been recognized by many research communities, including image processing, image coding, and computer vision. The new international video coding standardization effort, MPEG-4, actually has this as its major theme, with an ambitious completion schedule (1998). Let's work hard and hope for its success!

1. S.-F. Chang, "Compressed-Domain Techniques for Image/Video Indexing and Manipulation", IEEE Intern. Conf. on Image Processing, ICIP 95, Special Session on Digital Image/Video Libraries and Video-on-demand, Oct. 1995, Washington DC. (invited paper)
2. W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin, "The QBIC Project: Querying Images by Content Using Color, Texture and Shape", SPIE 1993 Intl. Symposium on Electronic Imaging: Science and Technology, Conf. 1908, Storage and Retrieval for Image and Video Databases, Feb. 1993. (also in IBM Research Report RJ 9203 (81511), Feb. 1, 1993, Computer Science.)
3. S.-F. Chang, *Compositing and Manipulation of Video Signals for Multimedia Network Video Services*, Ph.D. Dissertation, U.C. Berkeley, Aug., 1993.
4. B.C. Smith and L. Rowe, "A New Family of Algorithms for Manipulating Compressed Images," *IEEE Computer Graphics and Applications*, pp. 34-42, Sept., 1993.
5. B.-L. Yeo and B. Liu, "A Unified Approach to Temporal Segmentation of Motion JPEG and MPEG Compressed Video," *IEEE Int. Conf. on Multimedia Computing and Systems*, May 1995.
6. J. Meng and S.-F. Chang, "Tools for Compressed-Domain Video Indexing and Editing," to appear on SPIE Conference on Storage and Retrieval for Image and Video Database, San Jose, Feb. 1996.