

Le Duy-Dinh, Assistant Professor, Digital Content and Media Sciences Research  
Division

Building a truly usable video search engine

My research focuses on building a truly usable video search engine.

To make video searchable, we first break down a video into a number of short shots. We then tag each shot with a title and information: such as the people, things, and acts depicted. This is called indexing. After this processing, the shots can be searched and just their highlights extracted. I'm studying technologies for having computers handle this indexing to achieve high speed and high precision.

I'm considering using information from the Internet in this indexing. For example, I've thought of using resources like a list of the top 10 news stories of the day in indexing news video broadcast by NHK. I'm also thinking about extracting information from resources such as blogs written by people who watched the news and using it to analyze the video.

Of course, the interests of an individual don't necessarily coincide with those of the public at large. What's considered important news varies from person to person. While something like the soccer World Cup is included in most top 10 stories, most Japanese viewers would probably prefer to see only matches in which the Japanese team competed. I think we can make possible searches tailored to such personal tastes by providing (recommending) optimal shots based on user search histories and other information.

I'm also taking part in a project whose purpose is to study video analysis technologies for social analysis. Social analysis involves needs like, for example, examining how many

times former Prime Ministers Hatoyama and Aso appeared in all news video in the run-up to an election during a certain month. We're studying technologies to meet such needs.

At the same time, by its nature, a textual search of video content always involves the issue of semantic gaps: how best to process the meaning contained in video using computers.

As one approach to this issue, we're considering specifying as conditions shots already in the user's possession and shots the user has searched for before, then searching for similar shots. Being able to search videos on a shot-by-shot basis without textual intermediation should make it possible to avoid semantic gaps and make searching richer than through specifying conditions using text alone. For example, currently, 32 terms at most can be specified as conditions in a Google search. Specifying one or more shots as the search conditions should make it possible to search in even greater detail using the people, things, and acts depicted in the shots

While we're increasing the precision of video search, we also need to ensure that processing is done within a practical timeframe. Since the task requires massive computing power, we need to consider using grid architecture, although my preferred approach to resolving this tradeoff would be to develop smart algorithms. Right now, I'm building a new multimedia search engine with Professor Shin'ichi Satoh as an NII Grand Challenge project. Our goal is to provide a practical search engine within the next five to 10 years.

Interviewed and summarized by : Atsushi Saito