P2P comes to the aid of audiovisual search (w/ Video)

November 18 2009

(PhysOrg.com) -- Current methods of searching audiovisual content can be a hit-and-miss affair. Manually tagging online media content is time consuming, and costly. But new 'query by example' methods, built on peer-to-peer (P2P) architectures, could provide the way forward for such data-intensive content searches, say European researchers.

A team of researchers have turned to peer-to-peer (P2P) technology, in which data is distributed and shared directly between computers, to power potent yet data intensive audiovisual search technology. The technique, known as query by example, uses content, rather than text, to search for similar content, providing more accurate search results and reducing or even eliminating the need for pictures, videos and audio recordings to be laboriously annotated manually. However, effectively implementing content-based search on a large scale requires a fundamentally different approach to the text-based search technology running on the centralised systems of the likes of Google, Yahoo and MSN.

"Because we're dealing with images, video and audio, content-based search is very data intensive. Comparing two images is not a problem, but comparing hundreds of thousands of images is not practical using a centralised system," says Yosi Mass, an expert on audiovisual search technology at IBM Research in Haifa, Israel. "A P2P architecture offers a scalable solution by distributing the data across different peers in a network and ensuring there is no central point of failure."

Currently, when you search for photos on Flickr or videos on YouTube, for example, the keywords you type are compared against the metadata tags that the person who uploaded the content manually added. By comparison, in a content-based search, you upload a picture or video (or part of it) and software automatically analyses and compares it against other content analysed previously.

Working in the EU-funded SAPIR project, Mass led a team of researchers in developing a powerful content-based search system implemented on the back of a P2P architecture. The software they developed automatically analyses a photo, video or audio recording, extracts certain features to identify it, and uses these unique descriptors to search for similar content stored across different peers, such as computers or databases, on a network.

"In the case of a photograph, five different features are used, such as the colour distribution, texture and the number of horizontal, vertical and diagonal edges that appear in it," Mass explains.

In the case of videos, different frames are captured and analysed much like a photograph to build up a unique descriptor. Audio is converted into text using speech-to-text software, while music is analysed by its melody. The extracted features are represented in standard formats such as XML, MPEG7, MPEG21, MXF and PMETA, allowing complex queries from multiple media types.

Peering here, peering there in search of content

Processing and data transmission demands are kept in check by ensuring that searches target specific groups of peers on the network.

"When someone initiates a search, the system will analyse their content and compare it to other content across specific peers rather than across

the entire network. For example, if an image has a lot of red in it, the system will search the subset of peers that host a lot of images in which the dominant colour is red," Mass notes. "This helps ensure the search is faster and more accurate."

In the network, each peer - be it a home user's personal computer or a media group database - can be both a consumer and producer of content. All push data for indexing by the P2P network and make it searchable.

To further enhance the search capabilities, the SAPIR team developed software that compares a newly uploaded image to similar images and then automatically tags it with keywords based on the most popular descriptions for the similar images in the database. This automated tagging technique, based on metadata generated by the "wisdom of the crowd," is being further researched by IBM and may find its way into commercial applications, Mass says. It could, for example, automatically and accurately tag photos uploaded to Flickr from a mobile phone, eliminating the need for users to battle a small screen and keypad in order to do so manually.

Mass sees additional applications in security and surveillance by incorporating face recognition and identification into the image and video analysis system, as well as, evidently, for media companies looking for a better way to organise and retrieve content from large audio, video and image collections.

"IBM and the other project partners are looking at a variety of uses for the technology," Mass notes.

Project partners Telefónica and Telenor are also looking to use the audiovisual search commercially.

One scenario envisaged by the SAPIR researchers is that of a tourist

visiting a European city. They could, for example, take a photo of a historic monument with their mobile phone, upload it to the network and use it to search for similar content. The city's municipal authorities and local content providers, meanwhile, could also act as peers, providing search functionality and distributing content to visitors. Combined with GPS location data, user preferences and data from social networking applications, the SAPIR system could constitute the basis for an innovative, content-based tourist information platform.

More information: www.sapir.eu/

Provided by ICT Results