

Information Retrieval in Geneva: the *Viper* research group



Since more than 10 years, the *Viper* research group of the Computer Science Department of the University of Geneva works at advancing the field of Information Retrieval. The group is part of the Computer Vision and Multimedia Lab (CVMLab), originally specializing on Computer Vision problems. From there, the *Viper* group has spawned, with an initial motivation of transferring the lab's expertise in Computer Vision for reaching efficient Content-based Image Retrieval (CBIR). It now addresses a wider range of scientific challenges, from Content-based Multimedia Information Retrieval to Multimedia Content exploration and annotation via Knowledge Management and Discovering developments, all based on formalized and robust strategies, essentially derived from Machine Learning.

1. From classical information retrieval to multimedia information retrieval

Text information retrieval has a long history of development since the initial SMART system proposed in the late 60's. It has now reached a rather mature stage and exported itself in general contexts such as Digital Libraries and even expanded successfully to the Web context in the late 90's.

Multimedia Information Retrieval has a younger history and is less attached to commercially successful systems. Multimedia Information here generalizes text to all other types of information, from visual content (*eg* images, drawing, 3D), audio content (*eg* speech, music) to the combination of all such as Web pages (*eg* considered as a compound of images and text) or video (*eg* a news broadcast as a compound of a spatiotemporal visual stream, speech, possibly music and text caption or teletext).

Multimedia Information Retrieval offers several challenges:

- First, most of the multimedia content is not text. The automated processing, indexing and retrieval of that content therefore calls for advanced techniques for the automated interpretation of audio, visual contents and alike. Hence, classical limitations of fields such as Computer Vision or Speech Understanding are directly or partly inherited by the field of Multimedia Information Retrieval. The lack of suitable interpretation capabilities is accepted as the *semantic gap*, the difference between the best interpretation reached from the physical features of the information at hand (color, pitch,...) and what a human operator would capture from that content ;
- Further, Multimedia Information is often complex and composed of several unit streams of information (*eg* audio and video streams), referred to as *modalities*. Hence, Multimedia Information Retrieval adds to the above interpretation challenge the problems of mixing correctly various simultaneous streams of information. This problem is that of Information Fusion where information gathered from multiple modalities should be fused into one unique characterization of the multimodal content. It is also generally recognized that further reductions of the semantic gap may only come from a good understanding of fusion mechanisms;
- The complexity of Multimedia Information retrieval also influences the efficiency of its handling. While text data may be stored and handled efficiently and document unit sizes are within acceptable ranges for storage (memory and disks), transfer, and analysis, this is less so with Multimedia Information. In the case of video for example, the size of one document only may well be equivalent to that of a complete

text corpus. The scalability of processing and access strategies therefore become critical in this context;

- Multimedia Information Retrieval follows its textual counterpart in that it should be interactive. In most usage scenario, the acceptable delay between a query and a response is within the same range of that for textual systems. To scalable indexing and access, MIR therefore adds fast and accurate query processing. It is further recognized that proper user interaction is critical to a correct semantic interpretation of his/her queries via the concept of *relevance feedback*, where the user “trains” the system to his/her view of the information need;
- Finally, while text Information is interacted with easily through summaries for example (*eg* snippets, tagclouds), the case of MIR again makes the interaction more complex. Temporal streams (video, audio) should also be handled via summaries to enable proper relevance feedback acquisition. For example, if provided with a page of 20 video documents of 5 minutes each as initial result, the user may have to spend about 30 minutes before returning some list of positive and negative matches to better the results. Clearly, video summaries such as the joint use of keyframes and tagclouds are of help here;

An initial temptation is that of transferring directly the experience text IR to MIR. Models for text IR such as the classical TF-IDF indexing based on counting word occurrences in text documents are rather simple and have largely proved their efficiency. It is therefore natural to find MIR systems directly mapping these models onto other content. The Viper group has done so eight years ago when releasing the GNU Image Finding Tool (GIFT) as the first complete GNU opensource package¹ to enable Content-based Image Retrieval. Here, image parts characterized by color and texture at several scales become words and are indexed as text would be. The query mechanism is the Query-by-Example (QBE) paradigm where the user shows positive and negative examples of the sought content as a query. In this context, the Relevance Feedback mechanism is just a system to construct incrementally the proper query as the appropriate set of positive and negative examples.

2. Information retrieval as a learning problem

Models designed for text IR adapt from relevance feedback, essentially using linear or Bayesian learning. As result, their learning capabilities are limited to well-behaved classes. That is groups of relevant documents that are rather similar in their content. This is largely not the case for MIR where relevance is a complex notion. For example, the various possible visual aspects of a given object make it impossible for any visual feature to group all views of that object into one compact cluster. This motivates the search for advanced learning techniques and the Viper group has developed an expertise in modeling retrieval problems using Machine Learning tools. We have developed a fully operational model considering positive and negative examples provided as query as samples of the positive (relevant) and negative (non-relevant) classes. The challenge is therefore to interactively train a flexible Machine Learning algorithm with few provided samples of a (potentially small) positive class against provided samples of a (potentially large and diverse) negative class. Our strategy is based on non-linear learning using

¹ Still available at <http://www.gnu.org/software/gift>

generalization algorithms such as Support Vector Machines (SVM), Discriminant Analysis (xDA) and Boosting. We have adapted a number of techniques to this interactive unbalanced context.

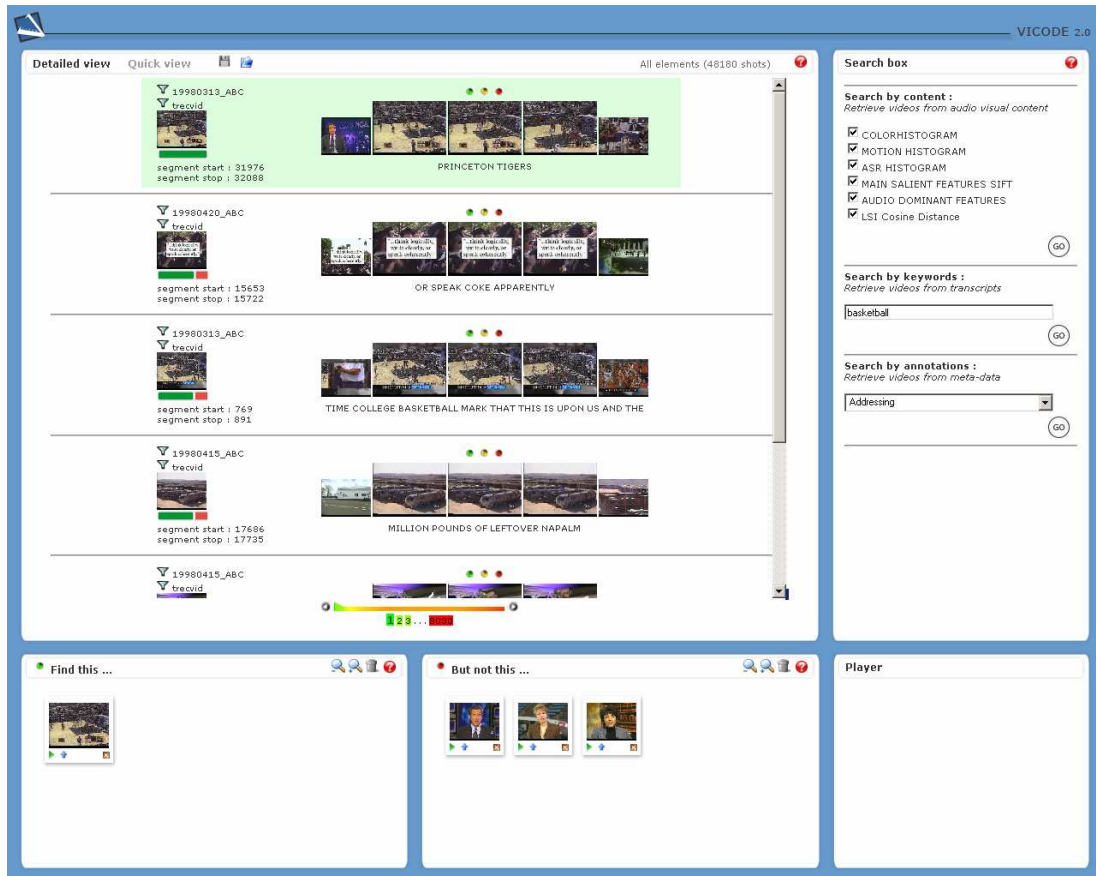


Figure 1: *Viper*'s ViCoDE: a video Query-by-Example system

Learning further enables various formal ways of performing efficient Information Fusion. We have employed either early or late fusion strategies to create image and video search engines. One recent achievement of the *Viper* group is the development of a large-scale multimodal retrieval engine called CMSE (Cross-modal Search Engine²) that acts over combination of audio, visual and text content. It has successfully been applied to the retrieval of Cultural Heritage (CH) material (essentially pictures of art pieces with descriptions from CH websites and museums) in the MultiMATCH European project³. The classical Open Computer Vision library is used to extract features such as color, texture, face occurrence and our engine mixes accurately all needed modalities at query time. In such a context, the rather limited QBE paradigm is completed with keyword-based queries, the currently most natural way to formulate queries.

3. Extending the information access strategy

² CMSE: Cross-modal Search Engine: <http://viper.unige.ch/cmse>

³ MultiMATCH : MultiLingual/MultiMedia Access to Cultural Heritage. <http://www.multimatch.eu>

Content-based Retrieval is core for enabling scalable information access. However, not all scenarios are query-based. It may well be the case that the user simply wishes to explore the corpus at hand for *eg* opportunistic browsing (*ie* expressing his/her information need along the collection navigation). The Viper group is active in developing such a framework and has proposed the Collection Guiding principle as a way to guide the user within the virtual collection. That is, the collection of multimedia documents is preprocessed and a visit strategy for that (originally unorganized) content is automatically proposed. Our base strategy relies on viewing the collection as a network of inter-connected items and a visit over this network is modeled as a path reaching some optimality. For example a Traveling Salesman Tour over the similarity networks will provide the user with a path based on “smooth” content transition from one item to the next. Adapted browsing interfaces may therefore be developed to make the best of that data organization (see *eg* Figure 1).



Figure 1: *Viper*'s browser for Cultural Heritage data in the MultiMATCH project. The browser adds visual similarity (vertically) to collection browsing (horizontally)

Organizing the content is not only good for its exploration, it also helps its description (group annotation, summarization, sampling). As active participant of the PeTaMedia⁴ European Network of Excellence, the *Viper* group develops mining strategies for semi-automated annotation and content management over distributed networks. Our current developments focus on exploiting long-term interaction with multimedia content. We particularly exploit the context of social communities (in relation to the Web 2.0) to extract semantics from user interaction logs and project them onto shared items.

4. Summary

Providing efficient access to large-scale multimedia collections is a multi-faceted challenge, from content processing and indexing to annotation and exploration issues. We

⁴ PeTaMedia : Peer-to-Peer Tagged Media : <http://www.petamedia.eu>

view all these issues as necessary and complementary themes to address to reach yet unavailable accepted performances. The *Viper* group is addressing these issues globally and has already provided software solution materializing its findings.

5. Further reading:

The *Viper* group's scientific publication list is available at <http://viper.unige.ch/publications>

Selected items:

- [1] Bruno, E., Moënne-Loccoz, N., & Marchand-Maillet, S. (2008). Design of multimodal dissimilarity spaces for retrieval of multimedia documents. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(9), 1520-1533.
- [2] Kludas, J., Bruno, E., & Marchand-Maillet, S. (2008). *Exploiting Synergistic and Redundant Features for Multimedia Document Classification*. In 32nd Annual Conference of the German Classification Society - Advances in Data Analysis, Data Handling and Business Intelligence (GfKI 2008), Hamburg, Germany.
- [3] Marchand-Maillet, S., & Bruno, É. (2005). *Collection Guiding: A new framework for handling large multimedia collections*. In Proceedings of the First Workshop on Audio-visual Content And Information Visualization In Digital Libraries, (AVIVDiLib05), Cortona, Italy.
- [4] Moënne-Loccoz, N., Janvier, B., Marchand-Maillet, S., & Bruno, E. (2006). Handling Temporal Heterogeneous Data for Content-Based Management of Large Video Collections. *Multimedia Tools and Applications*, 31, 309-325
- [5] Morrison, D., Bruno, E., & Marchand-Maillet, S. (2009). Capturing the semantics of user interaction: A review and case study. In *Web Emergent Intelligence*, Springer.

Contact

Dr. Stephane Marchand-Maillet
Viper group – CS Department – University of Geneva
Center for Computer Science (CUI)
Route de Drize 7 – CH-1227 Carouge - Switzerland
URL: <http://viper.unige.ch>
Email: Stephane.Marchand-Maillet@unige.ch

Bio



Dr. Stephane Marchand-Maillet received his PhD on theoretical image processing from Imperial College, London in 1997. He then joined the Institut Eurecom at Sophia-Antipolis localization and recognition. Since 1999, he is Assistant Professor in the Computer Vision and Multimedia Lab at the University of Geneva, where he is working on content-based multimedia retrieval as head of the *Viper* research group.

He has authored several publications on image analysis and information retrieval, including a book on low-level image analysis.

He and his group are currently involved in several joint international efforts for benchmarking content-based multimedia retrieval systems (including the Benchathlon and ImageCLEF).