

Ergonomic Minding of Media Collections

Personal or professional collections of media, such as photos, music, movies and homevideos tend to increase quickly in size, mainly due to the ease of collecting lots of digital content using a variety of capture devices (digital cameras, mobile phones, camcorders) or over the Internet. However, the level of management (ie organization and search) of those multimedia collections remains largely behind the ease of content creation.

Ergonomic Minding of Media Collections

In spite of advances in contentbased retrieval and automatic multimedia indexing, multimedia content management is still difficult. For example, when it comes to personal photography, the frequent procedure is to place items into folders, often organized by date or periods, possibly events, and loosely annotated. As a result, images and events are difficult to find and search becomes a frustrating or even painful operation because tools for browsing personal image collections are not appropriate to the user's needs. In some cases, it is possible to organize a collection by accurate date/time or location/place (thanks to the EXIF and GPS data recorded by modern capture devices). However, this is still limited and it remains desirable to search using the multimedia content itself.

The current challenge of multimedia information systems is thus to design and provide professional or non-professional users with new interactive tools that:

- improve the browsing experience to access both personal or professional collections;
- make the search easier and more natural than the folderstyle layout;
- allow finding images and videos easily, rapidly, and accurately.

Media management can be significantly improved using the current results of lowlevel content abstraction techniques (e.g., color, contours, texture) and highlevel content abstraction techniques (e.g., image classification, face recognition) by combining those results with information obtained by minding a particular collection and developing novel browsing interfaces guided by the user's personal preferences.

The *Ergonomic Minding of Media Collections* is thus facing the following challenges: – extract relevant information from the media content for efficient indexing, search and retrieval. More precisely, the following technologies should be investigated: (i) robust extraction of invariant visual descriptors for image classification and object recognition, (ii) accurate and reliable unconstrained face detection and recognition as the human face is recognized to be an important semantic cue in visual content.

– develop novel multimedia minding strategies and techniques to optimally prepare and enrich the collection content for constructing new user interaction models.

– propose and validate in front of users new interaction models for image/ video search and browsing based on clustering or topology-preserving dimensionality reduction and projection techniques. The main goal is to offer the user a global and precise access to the multimedia collection at minimal cost.

In an example scenario, the user navigates within a flat 2dimensional arrangement of his/her media collection, organized by clusters. Each cluster corresponds to a given search criterion (query) and is represented by a statistical visual summary of the query result. One main challenge is to make available data interoperable at all levels. Thus, the construction of clusters would use and combine EXIF information (e.g., date/time, location), metadata (e.g., tags, events) and/or content (e.g., dominant color, people occurrence). In a complementary faceted-like approach, if the current facet combination isolates a subset of the collection, then it is possible to further refine the search by seamlessly navigating within the clusterbased representation of this subset.

Examples of Existing Systems

As examples of relevant directions in the development of the *Ergonomic Minding of Media Collections* methodology, we detail two applications illustrating different perspectives.

Collection Guiding

The Collection Guide proposes an alternative approach to many current information management systems, which are centered on the notion of a query. This is true over the Web (with all classical Web Search Engines), and for digital libraries. In the domain of multimedia, available commercial applications propose rather simple management services whereas research prototypes are also looking at responding to queries. The notion of browsing comes as a complement or as an alternative to query-based operations in several possible contexts.

In the most general case, multimedia browsing is designed to supplement search operations. This comes from the fact that the multimedia querying systems largely demonstrate their capabilities using the Query-by-Example (QBE) scenario, which hardly corresponds to any usable scenario. Multimedia search systems are mostly based on content similarity. Hence, to fulfill an information need, the user must express it with respect to relevant and non-relevant examples. The question then arises of how to find the initial examples themselves. Researchers have therefore investigated new tools and protocols for the discovery of relevant examples. These tools often take the form of browsing interfaces whose aim is to help the user exploring the information space in order to locate the sought items.

In Marchand-Maillet, 2005, the principle of *Collection Guiding* is introduced. Given the collection of images, a path traversing the complete multimedia collection is automatically created so as to “guide” the visit of the collection. For that purpose, image intersimilarity is computed and the path is created via a Travelling Salesman tour of the collection. The aim is to provide the user with a base exploration strategy based on a minimal variation of content at every step. This implicitly provides a dimension reduction method from a high dimensional feature space to a linear ordering. The Collection Guide methodology provides also several multidimensional arrangements and is therefore directly related to information visualization.

The figure on page 23 shows an example opposing the classical unordered gridbased display and a 3D display strategy exploiting content similarity and clustering. Similar images are automatically arranged around their representative image in a planet metaphor. The user may thus obtain a global over view of the collection (interplanet arrangement) or visit a specific subset of the collection (i.e. visit a specific “solar system”).

It is now clear that browsing comes as a necessity to closely adapt information inspection and retrieval to the specific user’s needs. There is no doubt that future information systems will comprise this emergent aspect, as a complement to currently dominating search operations.

Google Portrait

In 2007, S. Marcel and al. proposed *Google Portrait* (<http://www.idiap.ch/googleportrait>) to retrieve and browse images from the Internet containing only one particular object of interest: the human face. The goal is to filter the images provided by a standard image retrieval system with a face detector and to present portraits as a result instead of the complete image.

Image search starts with a text query. The Google Image engine is used to retrieve images matching the query. Each image URL is extracted from the Google Image result page, then images are processed in parallel. This processing includes download and face finding. Images with detected faces are presented on a new result page listing face portraits together with a confidence and direct links to the image URL and to the source page. The result page is a table with 5 columns and with as many rows than images with detected faces. The first column contains image close ups ("portraits"), the second column contains a confidence on the likelihood of the portrait to be a face, the third column contains the size of the original image and the last columns contain links on the the original image and on its website.

Google Portrait includes a module for manual annotation. Indeed, portraits (detected faces) are very likely to correspond to the query, but there is no guarantee as Google Portrait uses a face detection system, not a face recognition system. It is then possible to edit a result and to change the tag of the portrait (name of the person). Tags are saved in a database which can then be populated based on collaborative working.

Google Portrait has been released on November 27, 2006. Later in spring 2007, Google Image was providing an "unofficial" face finding search mode. Google Image face finding is now directly available in Advanced image search at least since June 24, 2007. Nearly at the same time, probably during summer 2007, Microsoft added also face detection to Microsoft's Live Search. The comparison with Google Portrait in terms of performance is impractical since both Google and Microsoft (those companies have incomparable computing facilities) online face-detectors give results on already batch processed images. Conversely, Google Portrait is performing live face detection on images downloaded on the fly from the Internet.

Conclusion

The *Ergonomic Minding of Media Collections* implicitly acknowledges the limits of current information access systems and paves the way to new solutions and challenges in the Multimedia Information Retrieval and Management community. The emphasis is placed on shifting the focus from the traditional contentprocessing and indexing view point to a knowledge and data minding approach complemented with a strong involvement of users in the construction of interactive systems.

We believe that such a joint data processing and usercentric approach will demonstrate that the strong implication of users, as a source of semantic information via dedicated interfaces adapted to efficiently capture useful information, is a robust and scalable solution to the problem of highlevel management of multimedia information.

Various European and Swiss research projects in the area of Multimedia Information Retrieval and Management are currently active. Additionally, a project initiative on the *Ergonomic Minding of Media Collections* is currently under setup to target the development of these new generations of interactive multimedia management systems and to encourage their transfer to usable commercial solutions.

References

- S. Marchand-Maillet and E. Bruno, “Collection Guiding: A new framework for handling large multimedia collection”, *International Workshop on Audio-Visual Content and Information Visualization in Digital Libraries (AVIVDiLib'05)*, 2005 (<http://vipер.unige.ch/collecti...>).
- S. Marcel, P. Abbet and M. Guillemot, “Google Portrait” (<http://www.idiap.ch/googleportrait>), *Idiap Communication*, IdiapCom072007, 2007.



Sébastien Marcel

Idiap Research Institute, Centre du Parc, Rue Marconi 19, CP 592,
1920 Martigny, Switzerland, phone: +41 (27) 721 77 27



Stéphane Marchand-Maillet

Viper group – CUI – University of Geneva, Battelle Building A, 7,
Route de Drize, 1227 Carouge, Switzerland, phone: +41 (22) 379 01
54

Abstract

Français

Il est aujourd'hui très simple de se constituer des collections de médias (photos, musique, films, etc.). La gestion et le maniement de grandes quantités de données, qui deviennent rapidement trop volumineuses, sont néanmoins toujours aussi difficiles. La gestion des médias peut être considérablement améliorée si l'on combine des techniques *low-level* pour abstraire des contenus (p.ex.: couleurs, contours, surfaces) avec des techniques *high-level* (p.ex.: classification des images, sélection des visages).

Tandis que le «browsing» classique dépend d'indications de recherche précises, le *Collection Guiding* renverse le processus de recherche: les collections sont en effet computérisées, les vues multidimensionnelles sont réduites à une présentation linéaire le long de laquelle le «browser» peut se mouvoir; il apprend ainsi à reconnaître le portrait simplifié de la collection complète et peut décider dans quelle direction, respectivement dans quel espace il veut poursuivre sa recherche. De tels systèmes de *browsing*, qui peuvent être adaptés aux besoins de tout un chacun, remplaceront de plus en plus à l'avenir les systèmes de recherche rigides.

Google Portrait procède d'une autre manière: le système délaisse toutes les autres composantes au profit d'un seul et unique critère de recherche: le visage humain. Les images sont filtrées au moyen d'un programme de détection de visages (à ne pas confondre avec un programme de reconnaissance des visages; en anglais: *to detect*, resp. *to recognize*) où uniquement le portrait et non pas l'image dans son ensemble est intégré dans les résultats de la recherche.

Microsoft a suivi entretemps avec une détection des visages en mode «Live Search». Globalement, on constate dans la gestion des objets multimédias un glissement du traitement classique des contenus avec indexation vers une approche qui privilégie le *knowledge management* et le *data minding* et qui implique davantage l'utilisateur dans la construction de systèmes interactifs. En Europe et en Suisse, de nombreux projets sont actuellement en cours dans le domaine de la recherche sur les contenus multimédias et leur organisation.

Deutsch

Es ist heutzutage einfach, digitale Mediasammlungen (Fotos, Musik, Filme etc.) anzulegen. Die Verwaltung und der Umgang mit den rasch wachsenden und damit unübersehbaren Datenmengen gestalten sich aber immer schwieriger. Das Mediamanagement kann erheblich verbessert werden, wenn einfache Techniken der Bildbeschreibung, z.B. Farben, Konturen, Oberfläche, mit entsprechenden anspruchsvollen Techniken, z.B. Bildklassifizierung, Gesichtsdetektion, gemeinsam eingesetzt werden.

Während herkömmliches «Browsen» von präzisen Sucheingaben abhängig ist, kehrt das «*Collection Guiding*» den Suchvorgang um: Sammlungen werden informatisiert, mehrdimensionale Ansichten auf eine lineare Darstellung heruntergebrochen, deren entlang

sich der «Browse» bewegen kann. Der Nutzer lernt so das vereinfachte Abbild der gesamten Sammlung kennen und kann sich entscheiden, in welche Richtung resp. in welchem Raum er weitersuchen möchte. Derartige Browsersysteme, die sich den Wünschen des Einzelnen anpassen können, werden starre Suchsysteme in naher Zukunft vermutlich in den Hintergrund drängen.

Einen anderen Weg geht *Google Portrait*: Das System vernachlässigt alle übrigen Suchelemente bis auf ein einziges Suchkriterium: das menschliche Gesicht. Bilder werden mittels eines Gesichtsdetektionsprogramms (nicht zu verwechseln mit Gesichtserkennungsprogramm; engl. to detect resp. to recognize) gefiltert, anstelle des gesamten Bildes wird ausschliesslich das Portrait in das Suchergebnis einbezogen. Microsoft hat in der Zwischenzeit mit einer Gesichtsdetektion in «Live Search» nachgezogen.

Insgesamt zeigt sich beim Multimedia-Management eine Verschiebung des Ansatzes von der traditionellen Content-Verarbeitung mit Indexierung hin zu einem neuen Ansatz, der das Knowledge-Management und das Data Mining in den Vordergrund stellt und damit den Nutzer stärker als bislang in den Aufbau von interaktiven Systemen einbezieht. Zur Zeit sind in der Schweiz und in Europa zahlreiche Projekte zum Thema Suche nach und Organisation von multimedialen Inhalten am Laufen.