MULTIMEDIA RETRIEVAL - STATE OF THE ART

Peter L. Stanchev

Abstract: In this keynote speech we discuss the history, the state of art, and the future of multimedia retrieval [6]. Wide access to large information collections is of great importance in many aspects of everyday life. For this reason, significant effort has been spent in studying and developing techniques that support effective and efficient retrieval of multimedia data. Every day we are overwhelmed by information of many types: TV channels, news feeds, web sites, to name a few. Without an efficient and effective retrieval and filtering support, much time and effort is required in finding the information that we really need in this highly dynamic information age.

The process of image description consists of extracting the global image characteristics, recognizing the image-objects, and assigning semantics to these objects. The image data can be treated as physical image representation and its meaning as a logical image representation. The logical representation includes methods for describing the image and image-objects characteristics and the relationships among the image objects. Several visual descriptors exist for representing the physical content of an image, such as the MPEG-7 standard [5]. Historically, semantic retrieval was frequently based on computer vision. To reduce the semantic gap, the low-level content-based media features are frequently being converted to high-level concepts or terms.

The MPEG-7 descriptors can be classified as general visual descriptors and domain specific descriptors. The former include color, texture, shape, and motion features. The latter includes face recognition descriptors. Color is one of the most widely used image and video retrieval feature. The MPEG-7 standard includes five color descriptors, which represent different aspects of the color and include color distribution, spatial layout, and spatial structure of the color. The image texture is one of the most important image characteristic in both human and computer image analysis and object recognition. Visual texture is a property of a region in an image. There are two texture descriptors in MPEG-7: a homogeneous texture descriptor and edge histogram descriptor. Both of these descriptors support search and retrieval based on content descriptions. MPEG-7 supports region and contour shape descriptors. Object shape features are very powerful when used in similarity retrieval. Although distance functions are not part of the standard, we will present the most used distance functions.

A technique for improving the similarity search process of images in a Multimedia Content Management System is analyzed. The content based retrieval process integrates the search on different multimedia components, which are linked in XML structures. Depending on the specific characteristics of an image data set, some features can be more effective than others when performing similarity searches [2]. Based on this observation we propose a technique that predicts the effectiveness of the MPEG-7 image features that depends on a statistical analysis of the specific data sets in the Multimedia Content Management System. This technique is validated through extensive experiments with human subjects.

We illustrate several aspects of the fine art databases [4]. We showed that MPEG-7 descriptors can be used, but they give different results than applying on other type of images. The use of the Color Structure descriptor only produces sufficiently efficient results in the query search. The new generation Semantic Web languages, such as

RDF(S) and OWL will play a major role. The integration of semantic understanding of pictures with personalized delivery raises new questions. The query language for this type of system is not yet scandalized but we hope that an emerging standard will come soon.

We discuss the problems witch arise working with Magnetic Resonance (MR) images. As an illustration of medical image processing tools we discuss MR brain segmentation problems. Functional analysis of different medical systems is made. We emphasize on the fact that working with medical images is different from working with other kind of images. As an illustration two systems are presented [3]. The first system is MEDIMAGE, which is a multimedia database for Alzheimer's disease patients. It contains MR images, text and voice data and it is used to find some correlations of brain atrophy in Alzheimer's patients with different demographic factors. The second system is Epilepsy system, which includes image data from MRI and SPECT, scans and EEG analysis results and it is used for patients with epilepsy.

We present a novel approach for efficient video stream filtering that is based on the use of the MPEG-7 descriptors and exploits the properties of metric spaces in order to reduce the computational load of the filtering receiver [1]. Among other types of information, Audio/Video can be considered today as a primary means of communication, due to its richness in informative content and to its appeal. This implies that the development of techniques supporting the retrieval of Audio/Video documents is of primary importance to enable the access for the general public as well as for professional users of significant asset of today's life. This process will gain more impetus from the adoption of standards to represent video content. The retrieval process is based on a simple schema: users specify their request needs (e.g. a set of keywords or a sample image) that are translated into a system query. The items in the archive are compared with the user's query, in order to determine if they are relevant for the user's request. To process this type of query, it is necessary to determine a set of properties of the objects stored in the archive (usually called features) and a similarity measure to compare queries and archive objects. Video features can be described by using the MPEG-7 standard. In case the similarity measure is metric, many possible approaches to create indexes can be adopted. These indexes allow improving the efficiency of the retrieval process, by comparing the query only with a limited number of objects in the archive. Our approach goes toward a solution of this problem, by proposing a novel approach to Audio/Video filtering that makes use of simple additional information sent together with the video. This allows avoiding the comparison of the filter with video features for many non-matching videos or video components that will not pass the filter in any case. The proposed approach requires that the measure of the similarity between the filter and the video representative is metric, and it is based on the use of the well known technique of pivots.

The following systems are discussed: Tamino, TV-Anytime, MILOS, PicToSeek, Marvel, imgSeek, IKONA, SIMPLIcity, ALIP, SIMBA, Viper, LCPD, Video Google, Cortina, Octagon, PicSOM, LIRE. Semantic retrieval in TREC is also presented.

The specific challenges in the field are highlighted. Conclusion remarks about the future of the multimedia retrieval are drowning.

Bibliography

[1] Falchi F., Gennaro C., Savino P., Stanchev P., Efficient Video Stream Filtering, accepted for *ACM Multimedia,* 2007

[2] Stanchev P., Amato G., Falchi F., Gennaro C., Rabitti F., Savino P., "Selection of MPEG-7 Image Features for Improving Image Similarity Search on Specific Data Sets", *7-th IASTED International Conference on Computer Graphics and Imaging, CGIM 2004*, Kauai, Hawaii, 2004, 395-400.

[3] Stanchev P., Fotouhi F., Siadat M-R., and Soltanian-Zadeh H., Multimedia Mining- "A High Way to Intelligent Multimedia Documents", *book chapter 7- Medical Multimedia and Multimodality Databases,* D. Djeraba (Eds) - Kluwer Accademic Publishers, 2002, 139-160.

[4] Stanchev P., Green Jr D.., Dimitrov B., Semantic Notation and Retrieval in Art and Architecture Image Collections, *Journal of Digital Information Management, Vol. 3, No. 4, December 2005*, 218-221

[5] Stanchev P., Introduction to MPEG 7: Multimedia Content Description Interface, International *Conference on "Visualization, Imaging and Image Processing (VIIP 2004),* September 6-8, 2004, Marbella, Spain

[6] Stanchev P., Semantic Video and Image Retrieval - State of the Art and Challenges, ACMSE 2007, *The 45th ACM Southeast Conference*, Winston-Salem, North Carolina, USA, March 24, 2007, 512-513

Author's Information

Peter Stanchev is a Professor of Computer science at Kettering University, Flint, Michigan, USA and Professor and Department Chair at the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria. He is also affiliated with the Institute of Information Science and Technologies, Italian National Research Council, Pisa, Italy. He has published two books, more than one hundred and fifty papers in monographs, journal, and peer-reviewed conferences that have more than 250 citations. His research interests are in the field of Image Processing, Multimedia Systems, Database Systems, and Expert Systems. He lectures courses in the areas of: Database Systems, Information Retrieval, Data Mining, Human-computer interaction, Computing and Algorithms, Web technology, Computer Architecture, Design of Information Systems, Computer Operating Systems, Image Databases, Fuzzy sets, Decision Support Systems, Multimedia Databases, Expert Systems, Image Processing, Computer Graphics and Game design. He is a Co-Chair of the annual multimedia semantics workshops.

Peter L. Stanchev – e-mail: pstanche@kettering.edu; www.kettering.edu/~pstanche
Kettering University; Flint, MI, USA 48504;
Institute of Mathematics and Informatics, Sofia-1113, acad. G.Bontchev, bl.8, Bulgaria;