
Multimedia Tools and Applications, 27, 187–193, 2005

Multimedia Algorithmics

REMCO C. VELTKAMP Remco.Veltkamp@cs.uu.nl
Institute of Information and Computing Sciences, Utrecht University, Padualaan 14, 3584 CH, The Netherlands

Abstract. If perceptually relevant multimedia methods with guaranteed performance are not developed soon, there is no hope that the problem of multimedia information overload is effectively solved. Research is needed to handle images, music, video, and 3D models, with methods that guarantee robustness, invariance, efficiency, etc., and are also perceptually and cognitively relevant. The invention of algorithms that provably satisfy such properties is a new field of research: multimedia algorithmics.

Keywords: multimedia, algorithmics, perception

1. Introduction

Multimedia research has been going on since the nineteen-sixties, even if it was not called like that. A key aspect of multimedia research is interfacing: establishing a seamless interaction and communication between the user and the computer. In that respect it represents an important ingredient of current developments which are denoted by buzz phrases such as ubiquitous computing, ambient intelligence, context awareness, the disappearing computer, video at your fingertips, anything, anyone, anywhere, anytime. Multimedia retrieval is essential for coping with the problems of information overload, in production and content management, and in personalized usage. Indeed, the reason that email and web search engines have become so immensely popular are precisely that they cope with these issues with respect to text. However, if perceptually relevant multimedia methods that guarantee performance are not invented soon, there is no hope that similar problems are effectively solved with respect to images, music, video, and 3D models.

Since the first pictorial information systems in the early nineteen-eighties, research has come a long way in developing various methods to handle visual information by its content, as opposed to processing by keywords [2]. However, these content descriptions consist of low level color, texture, and shape features [3], and they often miss perceptual relevance. The methods for extracting and comparing these features are primarily heuristic, which, although they are clever themselves, miss guaranteed properties. In contrast, an algorithmic approach is focused on provable properties, see Section 1.1.

By way of example, looking at a particular multimedia framework as in figure 1, we see that those processes that are of an algorithmic nature are the extraction of features from the multimedia documents, the matching of the query features with the database features, the construction of the indexing data structure to speed up the searching, and the visualization of the resulting retrieved multimedia documents.

The big challenge in multimedia for the next fifteen years is the processing of information in a way that is perceptually and semantically relevant. Because of the need for personalized

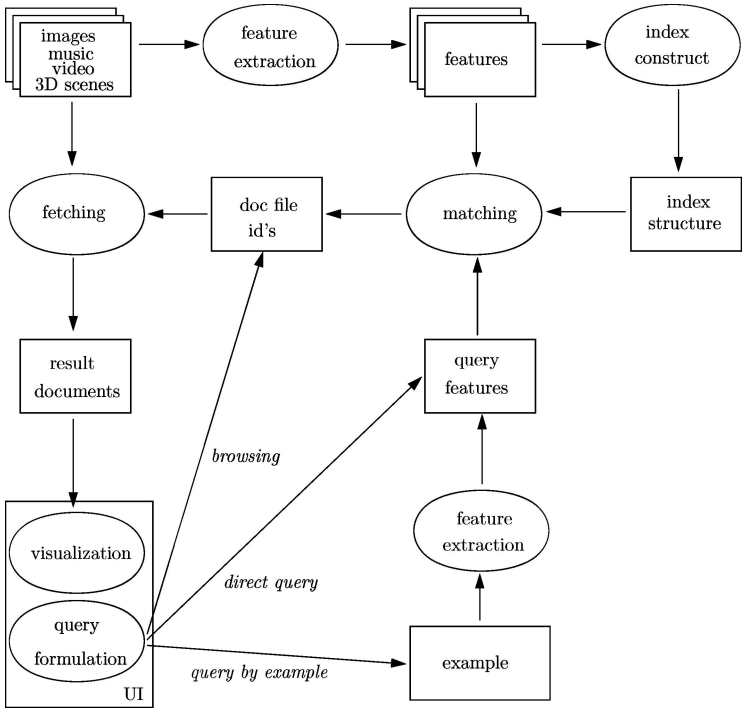

Figure 1. Multimedia retrieval framework.

information access and searching, processing should be done in a manner that is guaranteed effective. Because the searching and filtering is performed on very large databases of multimedia information, it must be done with guaranteed efficiency also. The holy grail is not yet within reach. What makes this difficult is the gap between the high level semantic information and the low level features of current multimedia systems. For example, if one is looking for an image of the holy grail (such as figure 2) on the basis of image content features, one may query for a chalice shape and a star shape. However, simple edge detection yields a set of unconnected lines, not a star. Therefore, low level features will fail miserably for this purpose.

A concrete listing of research issues in multimedia is the following. Firstly, in order to arrive at semantic access, a necessary step is the identification of what is perceptually and cognitively important in the multimedia documents.

Secondly, in order to cope with the data and information overload, it is becoming essential that effective and efficient searching techniques are developed. Indeed, not only company archives contain huge amounts of media. The success of mobile phone with sms (short message service) shows that as soon as consumer groups adopt devices like mobile phones with built-in digital cameras and mms (media message service) via broad band communication like GPRS or UMTS, massive amounts of images and video are produces and stored. Digital music and movies are already causing a very large amount of Internet traffic. The so-called fourth wave in multimedia (after images, video and music), consisting

Figure 2. M. L. Kirk, "And Down the Long Beam Stole the Holy Grail", 1912.

of 3D models and scenes, is showing more and more on the web. Together these media form an enormous amount of data, and it is essential to provide tools to match and filter, and to retrieve personalized information from it.

Thirdly, to make retrieval feasible from such large quantities, efficient searching methods must be invented. In particular, indexing data structures and algorithms must be designed that avoid the need to scan whole collections from front to back, but instead refer the user in a few steps to the right place in the collection.

Fourthly, any successful fully-fledged system needs to provide a combination of image, video, music, and 3D model handling, with text capabilities. The integrated system engineering is far from trivial, and challenging in itself.

The algorithmic aspects of these items form an area of research, *multimedia algorithmics*, which requires a combination of theoretical algorithms design and application oriented experimentation.

1.1. Algorithm design

Like all computer systems, all multimedia systems are built on algorithms. They are the crucial mechanisms for working with information in any representation: computing, deriving,

deciding, checking, storing, searching, learning, managing, modeling, visualizing, comparing, optimizing, transforming, sending, protecting, etc. Any system in modern information and communication technology is an algorithmic system. Although not always clear to people outside computing science, algorithmics, the invention of algorithms, is a science, involving the derivation of solutions to unsolved problems, and proving properties such as correctness and efficiency. When it comes to the design of algorithms for multimedia, this involves for example algorithms for extracting and grouping perceptually relevant patterns, computing the similarity, indexing and searching in large collections, and visualizing retrieval results in a way that is meaningful for relevance feedback. Research issues are the invention of new algorithms that solve problems in an efficient way, guaranteeing provable properties in a rigorous way, taking an axiomatic approach, basing derivations on first principles.

1.2. Experimental algorithmics

Apart from fundamental modeling, design and analysis of perceptually relevant algorithms, implementations and experimentation play a crucial role to show proof-of-concepts in practice. Implementation was characteristic of early work in algorithmics. Donald Knuth, one of the most influential researchers in early computer science, insisted on implementing every algorithm he designed, and on conducting rigorous analysis of the resulting code. Since then, appreciation has faded. For some while, the standard journals in the algorithmics community, such as the J. Algorithms, J. ACM, SIAM J. Computing, and Algorithmica, have not published much experimental studies. Since a few years, the algorithms community has shown signs of returning to implementation and testing as an integral part of algorithm development [1].

1.3. Multimedia algorithmics

The above-stated aspects are combined into a line of research that is rooted in the discipline of fundamental algorithm design, and applied to the domain of multimedia: multimedia algorithmics.

In order to bring science a significant step further, new models and algorithms must be designed that provide solutions to the above-mentioned problems, and that are provably efficient, robust, and perceptually relevant. What makes this difficult is the gap between the high level semantic information and the low level features of current multimedia systems. A challenging research agenda for the next ten years is to invent algorithms for multimedia along the following orthogonal axes:

1. The tasks in a typical multimedia framework that are of an algorithmic nature: perceptual feature extraction, pattern matching, indexing, and visualization.
2. The different media (images, music, video, 3D models and scenes) to which these task are applied.

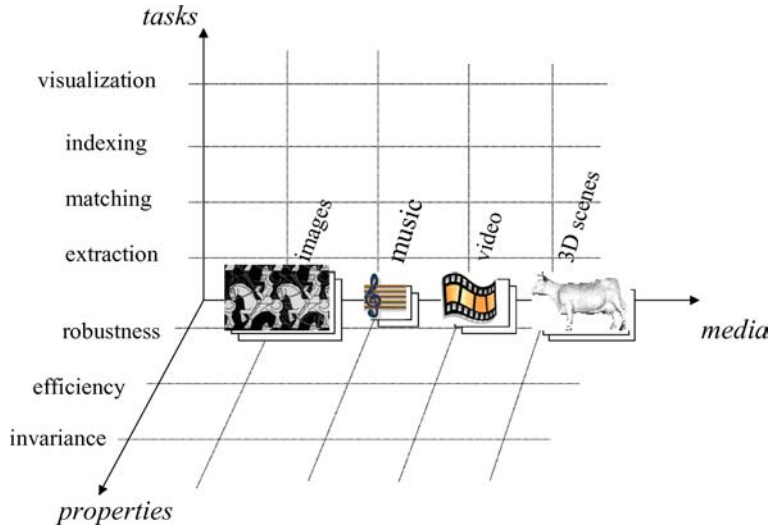

Figure 3. Multimedia research space.

3. The desired properties of algorithms that must be invented: robustness, invariance, efficiency, etc.

Together, these aspects span a whole research space, as illustrated in figure 3.

2. Contributions

The current issue contains four papers, covering two types of media: video and images, two papers on each type.

The paper by Adams, Venketesh, Bui, and Dorai is about identifying the location of the boundaries between different parts of movies, assuming a 3-act narrative structure. This allows for segmentation film at a higher level than at the level of shots and scenes. Their algorithm is based on the formulation of the moment of the act boundary given certain evidence, using Bayes rule. The experimental verification is performed on 25 complete movies.

Bertini, Del Bimbo, and Nunziati present an algorithm for the detection of highlights in sport video's, such as a corner kick in soccer, a free throw in basketball, a drop kick in rugby, a volley serve in tennis, and a turn in swimming. The particular highlights are modeled in finite state machines to encode the temporal sequence of events. The algorithm is based on visual cues only, such as playfield zones, camera motion, and position of sporters. These cues are automatically detected and fed into the decision algorithms for the particular hightlights.

The contribution of Hermes, Miene, and Herzog is a system for content-based image retrieval, incorporating algorithms for comparing regions that take into account that different

colors spread differently in color space, as well as the location of the regions in the image. The user is able to delete regions that are perceptually less relevant from the query, thus changing the overall similarity measure between query and database images.

The last paper is by Zhou, Ma, Celenk, and Chelberg. They present a new algorithm for content-based image retrieval based on the detection of regions of interest and relevance feedback. Their region detection algorithms are based on color saliency, the wavelet transform to represent texture, and the Radon transform to represent shape. The relevance feedback from the user is translated into weights of each detection feature. The improvement of relevance feedback over retrieval without relevance feedback is experimentally verified. Also the convergence ratio is experimentally demonstrated.

These four contributions together form a nice snapshot of the current research directions in multimedia algorithmics. With respect to video, there is a movement from the lower level units such as shots and scenes, to higher level units such as acts and highlights, exploiting domain knowledge of the type of video. In the image retrieval field there is a shifting from low level features towards higher level regions, exploiting user interaction to achieve more perceptual and semantic relevance.

Acknowledgment

I want to thank all the peer reviewers for their indispensable work in checking correctness and assessing the novelty and overall quality of all the submissions. They did a great job.

References

1. B. Moret, "Towards a discipline of experimental algorithmics," in *Data Structures, Near Neighbor Searches, and Methodology: Fifth and Sixth DIMACS Implementation Challenges*, M. Goldwasser, D. Johnson, and C. McGeoch (Eds.), 2002, pp. 197–213.
2. A.W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, 'Content-based image retrieval at the end of the early years," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 12, pp. 1349–1380, 2000.
3. R.C. Veltkamp and M. Tanase, "A survey of content-based image retrieval systems," in Content-Based Image and Video Retrieval O. Marques and B. Furht (Eds.), 2002, pp. 47–101.

Remco Veltkamp is associate professor at Utrecht University, and principal investigator of the Multimedia and Geometry Group. He obtained his Ph.D. in 1992 from Erasmus University Rotterdam. His current research focuses

on the algorithmic aspects of multimedia retrieval and geometric pattern recognition. Aspects like the algorithmic design and analysis, and experimental verification play an important role in his work. His Ph.D. dissertation has been published as a book, and he has written over 40 refereed papers on indexing, music retrieval, shape matching, shape reconstruction, geometric constraint management, and variational curve and surface design. He is editor of Pattern Recognition and the International Journal on Shape Modelling, and is organiser of the Dagstuhl Seminars on Content-Based Retrieval.