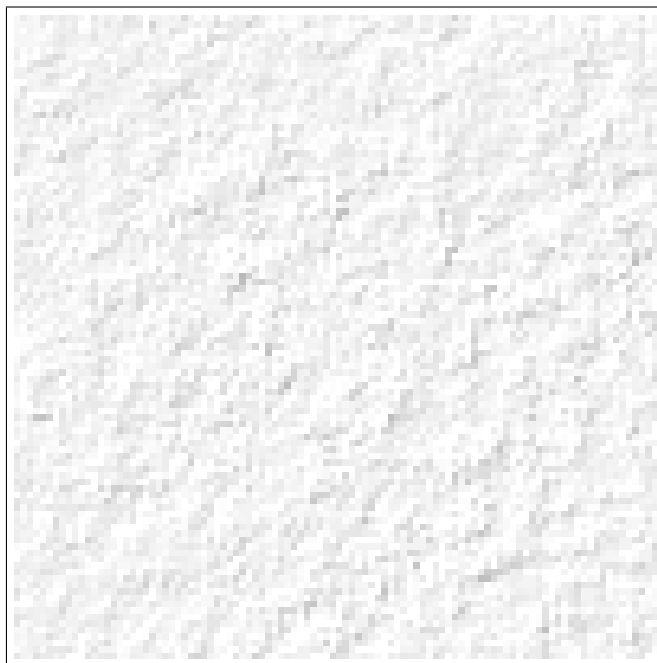


Æliens



topical media & game development

<http://www.cs.vu.nl/~eliens/media>

preface

This book provides a concise and comprehensive introduction to multimedia. It arose out of the need for material with a strong academic component, that is material related to scientific research.

Indeed, studying multimedia is not only fun. Compare it with obtaining a driver license. Before you are allowed to drive on the highway, you have to take a theory exam. So why not take such an exam before entering the multimedia circus.

Don't complain, and take the exam. After all it makes you aware of the rules governing the (broadband) digital highway.

themes and variations

So, who is this book meant for? It is meant for the student or reader who is looking for a quick introduction to the main topics in multimedia. The twelve chapters provide a concise overview of the themes and trends in current multimedia practice and research.

The themes and variations addressed in this book may be summarized as follows.

themes and variations

- *digital convergence – all for one, one for all*
- *broadband communication – entertainment*
- *multimedia information retrieval – as an afterthought?*
- *multimedia and game application(s) – from design to development*

To explain in somewhat more detail, *digital convergence* may be characterized as the coming together of data (including audio, video and information) in a possible multitude of platforms, to which these data are delivered by a variety of (broadband) communication channels. In fact, the increasingly powerful communication infrastructure due to the popularity of the Internet and the World Wide Web, leads to an almost universally accessible multimedia (information) repository, for which (unfortunately) the notion of (multimedia) information retrieval seems to have occurred only as an afterthought. Digital content design is only one step in the process of multimedia application development. Important issues in multimedia application development are, apart from project management, data representation, navigation, presentation and usability.

An underlying thought that motivated the writing of this book is that somehow the gap between *authoring* and *retrieval* should be bridged. In other words, either by developing the technology for extracting features or attributes from multimedia objects, or by applying content annotation for such objects, multimedia information retrieval should be considered as a necessary asset to make a multimedia web an effective information repository. In multimedia applications, such as the *digital dossier* we introduce in chapter 10, the data representation must accomodate meta-information, to support effective navigation and search.

Another line of thought, that became more clear during the writing of the book is concerned with the aesthetics of (interactive) applications. You will find more on this in chapters 11 and 12, that deal with game development.

what do you need to learn

When taking up multimedia as a subject of study, you may ask yourself what you need to know and learn about it. In general, what this book presents is

a collection of concepts, a number of facts, some history, potential applications and application areas, a brief overview of standards (some of which are still being developed), technology issues, as well as some scattered insights on visual design, application development and the relevance of multimedia and games.

Let me be frank with you. There is too much information to be digested in a first course. Nevertheless, after studying this book you will have an introduction to multimedia that should be viable for the rest of your (academic) career.

Now, don't hesitate, put yourself to the test and check which phrases and acronyms you are familiar with in the lists given for the subjects of *digital convergence*, *broadband communication* and *information retrieval*.

digital convergence

- concepts – *digital revolution*
- facts – *from the entertainment industry*
- history – *from Pong to Big Brother*
- applications – *infotainment*
- standards – *MPEG, RM3D, SMIL*
- technology – *TV, PC, DVD*

How did you succeed thus far? If you did well, try the second round and test yourself in what detail you have have knowledge about technologies mentioned.

broadband communication

- concepts – *Quality of Service*
- facts – *compression is needed*
- history – *the internet*
- applications – *entertainment and communication*

- standards – *HTTP, TCP/IP, RTP*
- technology – *cable, (X)DSL*

Finally, check to what extent you master the vocabulary of multimedia information retrieval.

multimedia information retrieval

- concepts – *features, precision, recall*
- facts – *the problem is utterly complex*
- history – *from text to multimedia*
- applications – *digital libraries*
- standards <– *distance metrics*
- technology – *indexing & algorithms*

If you are working online, you may click back to the text in the book that explains these notions. Just to make sure whether your impression of familiarity was justified.

assignment(s)

I strongly believe that practical work is necessary, also for academics, to get a good grasp on multimedia and game development. Even if your interest is purely intellectual, it pays off to make your virtual hands dirty and indulge in making a compelling presentation.

As an assignment, consider making a presentation that offers an

Annotated Tour in Amsterdam

Amsterdam is the place where I live, and where our students take their courses. You may find it more convenient or natural to replace Amsterdam with a location of your choice.

Online, you will find an elaborated version of the assignment, including an extended description, a working plan, deliverables and hints. In essence though, the intent of the assignment is to make a compelling, not to say artistic, presentation, and to explore the realm of multimedia rethorics.

As a tool you may choose, for example, Flash or the flex 2 SDK, which is freely available.

examination

Despite the fact that some consider the practical aspects of multimedia to be exclusively relevant, the intellectual aspects of multimedia should not be ignored.

Consider the following question, which is directly related to one of the themes underlying this book, that is the complementarity of authoring and retrieval:

Give a short description of the contents and structure of your presentation. Indicate how the information contained in your presentation can be made accessible (for example in search).

This question can only be answered when the student has a sufficient level of experience, insight and knowledge of the field, and is able to relate theory and practice.

Each chapter contains a brief list of questions that may be used as a checklist, to see if you have sufficient knowledge of a particular area. These questions may also be used to prepare exams! The questions are meant to test for insight, that is the ability to discuss a somewhat broader theme, and knowledge of concepts and technology, covering definitions, applications, historical facts, as well as the technological infrastructure enabling the deployment of multimedia applications.

In addition to the regular material, the book also contains a number of examples and sections indicating *research directions*. These sections are not meant to be part of the exam, but might provide the student with suggestions for projects or further research. Moreover, both the discussions in the *research directions* and the material in the appendices presents a vision on what multimedia should be. In effect, I have a strong preference for a programmatic approach to (intelligent) multimedia, as outlined in appendix E. Nevertheless, the bulk of the (regular) material is relevant also for readers with a rather different opinion on what constitutes the *essence of multimedia*.

how to use this book

The intended audience for this book is

intended audience(s)

- students (beginning and advanced)
- instructors
- professionals and interested laymen

The course notes were explicitly written for first year Computer Science and Information Science students. The Information Science students are expected to choose the specialisation *Multimedia and Culture*, a curriculum provided by the department of Mathematics and Computer Science of the Faculty of Sciences of the VU University Amsterdam. The course has a practical part and a theoretical part, which in combination takes 2-4 weeks, full time study. The book covers the theoretical part. The online version gives a skeleton assignment that may be adapted by the one responsible for the course. The online version contains all the material needed for giving a multimedia course, that is

multimedia course

- presentations for all chapters, including the preface in dynamic HTML slides
- presentable versions of the MPEG-4 standard, and other relevant material
- possible exam questions, with back links into the text for quick learning and review
- seven sample lectures, with additional explanation for the instructor

One additional remark may be made. This is (so to speak) 'a book with an attitude'. It is slightly authoritative and directive towards the students, telling them to learn the facts and 'do the exam'. Some students take refuge to learning the 'keywords and phrases'. They are even helped in this respect, since the text

uses a 'graphic' layout to emphasize important points, and to allow for a quick recognition of chunks of relevant material.

the artwork

Although a book about multimedia does not need to be a multimedia artefact itself, it seemed better to include illustrations, to avoid the impression of a 'dry' book. Since I did not want to include any redundant diagrams or pictures, I decided to use a personal selection from the history of visual design, games, computer art and video art, not only to spice up the book but also to give the reader a collection of interesting samples. Each chapter starts with illustrations setting the *visual theme* of the chapter. All other illustrations are, in one way or another, related to the examples or the text of that chapter. Brief comments about the artwork, and an explanation of the visual theme, can be found at the end of each chapter.

about the author

At some point you may wonder whether the author is qualified or authorized to write about a particular subject and, in this particular case, to publish a book about such an elusive notion as multimedia.

Let me give you some personal history. Way back in the seventies, I did a degree in painting at the Gerrit Rietveld Academy in Amsterdam. At the same time, I did a master's philosophy, where I graduated in the field of aesthetics on a comparative study on theories of imagination and creativity, reading writers such as Kant, Husserl and Sartre. Then I got an interest in computer music, after listening to a concert of Xenakis in Paris, and started to work on a PDP-15 (with 4K of memory) at the Institute of Sonology in Utrecht. Leaving all philosophy and traditional art behind, I learned programming, studied AI and theoretical computer science. Some eight years later, I obtained my Ph.D. in computer science and started my academic career. After working in software engineering, and in particular object-oriented software development, I was asked, at the end of the millenium, to set up a collection of multimedia courses, since by then multimedia was coming in vogue as an academic subject. These courses, which include the introduction multimedia, Web3D authoring, intelligent virtual environments, a multimedia casus, and recently also visual design, are reflected in this book.

about the book

What started as a (not so) gentle introduction to multimedia, has grown into a rich (at times somewhat idiosyncratic) collection of topical material about multimedia and game development. Borrowing a phrase from the politics of the seventies, at some point, apparently, the professional became personal, and the personal professional. Nevertheless, the book may still be read as an introduction. It is written in a concise and compact manner, supported by the slides format, which

allows for presentation of the material in class, and is illustrated by a variety of images, taken from the arts, design, and multimedia and game projects.

acknowledgements

This book is the result of developing the course notes for an *introduction to multimedia* for first year Computer Science and Information Science students. Hence, first of all, I like to thank the students that had to endure all the rough drafts of this material, and perhaps not to forget my experiment(s) with the presentation format of it. Further I like to thank Harrie van der Lubbe and Sander Lammers for developing the manual for Director and their support in developing the practical assignment. Also, I like to thank Martin Kersten from CWI for allowing me to join his Multimedia Database Systems research group as a guest for a period of about two years, and Alex van Ballegooij for his active involvement in the RIF project and his coding effort for the *slide* PROTOs, used to produce the presentation slides for this book and described in appendix B. Also from CWI, I like to thank Lloyd Rutledge, Lynda Hardman and Jacco van Ossenbruggen, for their effort in thinking about the multimedia course in its initial stages, and Lloyd and Jacco for their involvement in some of the practical work, and Jacco in particular for his knowledge of hypermedia systems that he shared with me during the period that he was my Ph.D. student. From CWI, I like to thank Zsofi Ruttkay for her general interest in 'my projects'. From the VU, I like to thank Andy Tanenbaum for allowing me to use his material on digital video, Gerrit van der Veer for taking the initiative for *Multimedia and Culture*, Zhisheng Huang for his excellent contributions to the WASP project, Johan Hoorn for our spirited cooperation, and Claire Dormann, for our discussions on the direction the *Multimedia and Culture* curriculum should take, and for sharing her thoughts on persuasive technology with me. I also like to thank Tatja Scholte and IJsbrand Hummelen from ICN (Netherlands Institute for Cultural Heritage) for their contributions to the *multimedia casus*, and Gaby Weijers and Bart Rutten from Montivideo for their cooperation and all the video material they so generously provided. Thanks are due to Mark Veldhuijzen van Zanten, Jaap Stahlie, Peter van Kessel, Federico Campanale and Katelijne Arts for providing me with material. Special thanks goes to the student members of the 2004 autumn *multimedia casus* group, for their collective work on the *abramovic dossier*, and to Rutger van Dijk for rekindling my interest in C++/DirectX programming by his youthful enthusiasm. I also should not forget two students from the Computer Science master Multimedia, who were exceptional in their dedication and skills, Winoe Bhikharie, who acted (among others) as the manager in the development of the VU-Life game described in 11.2, and Marco Bouterse, who excelled in both DirectX and Half-life shader programming. Also I need to mention Marek van de Watering, a student from Multimedia and Culture, who often surprised me by his sincerity and sensitivity. No doubt, I owe thanks to Gaynor Redvers-Mullon who made a serious attempt to encourage me to get the best out of this manuscript, even though at some point I decided to do it in my own way. Further, I also wish to thank Dhaval Vyas, at that time Ph.D. student at

VU, for his interest in my projects and for sharing his thoughts on the *panorama* project, and Zeljko Obrenovic, postdoc at CWI, for his enthusiasm, his technical skills and his willingness to assist me with my teaching. I thank Pier Vellinga, who I googled as the spin-doctor of climate change, for involving me in Clima Futura, and not in the least for the pleasant camaraderie we enjoyed during the effort to obtain the dutch national prize for academic communication, and Frans Feldberg, for involving me in the VU @ Second Life effort, an equally refreshing interdisciplinary partnership, with which we got into the 8 o'clock nation TV news, as planned! Last but not least, I owe Yiwen Wang (王依文) my deepest gratitude for providing the motivation to continue the work, and to strive for beauty and pleasure, in a serious way.

Finally, I must mention that I owe much insight and material to (among others) the following books and articles: Subrahmanian (1998), Forman and Saint John (2000), Chang and Costabile (1997), Ossenbruggen (2001), Vasudev and Li (1997), Hughes (2000), Grau (2003), Kress and van Leeuwen (1996), and not to forget Zielinski (2006). As in any intellectual endeavor, intellectual ancestry can hardly be praised enough. So let me briefly indicate, for each chapter, some of the sources that provided me with inspiration, insight and material:

1. Forman and Saint John (2000), Davenport (2000), Jain (2000).
2. Chang and Costabile (1997), Ossenbruggen (2001), Hughes (2000).
3. Vasudev and Li (1997), Koenen (2000), Visser and Eliens (2000).
4. Luna (2003), Adams (2003), Fernando and Kilgard (2003)
5. Subrahmanian (1998), Baeza-Yates and Ribeiro-Neto (1999).
6. Subrahmanian (1998), McNab et al. (1997), Kersten et al. (1998).
7. Subrahmanian (1998), Fluckiger (1995),
8. Fluckiger (1995), Ballegooij and Eliens (2001), Huang et al. (2002).
9. McCuskey (2002), Bolter and Grusin (2000),
10. Chapman and Chapman (2004a), Chapman and Chapman (2004b), Hughes (2000),
11. Sherrod (2006), Grau (2003),
12. Juul (2005), Arnheim (1957), Hawkins (2005), Kress and van Leeuwen (1996).

The material in sections 4.3, 7.1, 7.3, chapter 8, sections 9.3 and 10.2, and section 11.2 reflect my own research efforts. The other material has all been diligently collected from (among others) the sources mentioned.

topical media & game development

preface	i
 part i. digital convergence	 1
reading directives	1
perspectives	1
essay topics	2
 1. digital culture	 3
learning objectives	3
1.1 entertainment and experience	3
entertainment	5
digital experience	6
example(s) – <i>VR for pain relief</i>	8
research directions– <i>the face of cyberspace</i>	9
1.2 technological developments	10
TV or PC	11
example(s) – <i>visible world</i>	14
research directions– <i>technological determinism</i>	15
1.3 multimedia applications	15
new media	16
TV meets the Web	17
mobile multimedia	17
the academic perspective	18
example(s) – <i>personal memex</i>	20
research directions– <i>the information society</i>	20
1.4 development(s) – convergence in second life	21
questions	22
projects & further reading	22
the artwork	22

2. hypermedia information spaces	25
learning objectives	25
2.1 information and data	26
logical information spaces	27
example(s) – <i>e-flux</i>	29
research directions– <i>universal interchange</i>	29
2.2 hypermedia	31
classification of hypermedia	32
Dexter Hypertext Reference Model	33
Amsterdam Hypermedia Model	34
example(s) – <i>hush</i>	35
research directions– <i>computational models</i>	36
2.3 multimedia authoring	36
visualization	37
persuasive technology	38
(re)mediation	38
example(s) – <i>mobius</i>	41
research directions– <i>narrative structure</i>	41
2.4 development(s) – semantic mashups	43
questions	48
projects & further reading	49
the artwork	49
 part ii. delivery & presentation	 51
reading directives	51
perspectives	52
essay topics	52
 3. codecs and standards	 53
learning objectives	53
3.1 codecs	53
compression methods	55
compression standards	56
example(s) – <i>gigaport</i>	56
research directions– <i>digital video formats</i>	57
3.2 standards	59
MPEG-4	60
example(s) – <i>structured audio</i>	66
SMIL	66
RM3D – not a standard	69
example(s) – <i>rich internet applications</i>	73
research directions– <i>meta standards</i>	74
3.3 a multimedia semantic web?	75
Resource Description Framework – the Dublin Core	76
research directions– <i>agents everywhere</i>	78

3.4 development(s) – rethorics of change	81
questions	84
projects & further reading	84
the artwork	85
 part iii. information retrieval	 87
reading directives	87
perspectives	88
essay topics	88
 5. information retrieval	 89
learning objectives	89
5.1 scenarios	89
research directions– <i>information retrieval models</i>	91
5.2 images	93
descriptive information	93
similarity-based retrieval	95
example(s) – <i>match of the day</i>	98
research directions – <i>multimedia repositories</i>	98
5.3 documents	99
precision and recall	100
frequency tables	100
research directions– <i>user-oriented measures</i>	101
5.4 development(s) – inconvenient truth(s)	103
questions	104
projects & further reading	104
the artwork	104
 6. content annotation	 107
learning objectives	107
6.1 audio	107
musical similarity	109
example(s) – <i>napster</i>	111
research directions – <i>musical similarity matching</i>	111
6.2 video	113
video libraries	116
example(s) – <i>video retrieval evaluation</i>	117
research directions– <i>presentation and context</i>	118
6.3 feature extraction	120
anatomy of a feature detector	121
example(s) – <i>modern art: who cares?</i>	124
research directions– <i>media search</i>	124
6.4 development(s) – expert recommendations	126
questions	130
projects & further reading	130

the artwork	131
7. information system architecture	133
learning objectives	133
7.1 architectural issues	133
research directions– <i>the information retrieval cycle</i>	134
7.2 media abstractions	136
research directions– <i>digital libraries</i>	138
7.3 networked multimedia	140
virtual object	141
networked virtual environments	142
example(s) – <i>unreal</i>	144
research directions– <i>architectural patterns</i>	145
7.4 development(s) – living in a virtual economy	146
questions	148
projects & further reading	149
the artwork	149
 part iv. applications & tools	 151
reading directives	151
perspectives	152
essay topics	152
 8. virtual environments	 153
learning objectives	153
8.1 virtual context	153
research directions– <i>augmented virtuality</i>	156
8.2 navigation by query	157
case study – CWI	158
information in virtual worlds	159
presentation issues	160
research directions– <i>extended user interfaces</i>	161
8.3 intelligent agents	162
multi-user soccer game	163
agents in virtual environments	164
research directions– <i>embodied conversational agents</i>	168
8.4 development(s) – the metaverse revisited	169
questions	174
projects & further reading	174
the artwork	175

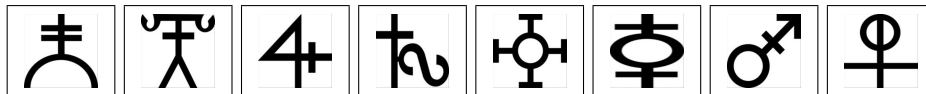
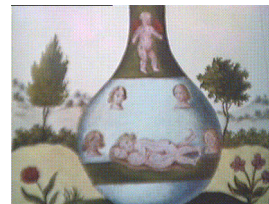
part v. game development	177
reading directives	177
perspectives	177
essay topics	178
 afterthought(s)	 178
 appendix	 182
D. resources, tools and technology	185
E. write an essay!	193
 references	 194
 index	 206

part i. digital convergence

more than the art of turning base metals into gold, alchemy is a system of cosmic symbolism
perfect solutions

chapters:

1. digital culture
2. hypermedia information spaces



2

reading directives In these first chapters, we will explore the notion of multimedia, look at it from a historical perspective and discuss in somewhat more detail the issue of convergence, the mix and interchange of media that has become possible in the digital era. In the second chapter, we will give a precise definition of information spaces, and an overview of the history of hypermedia.

Essential sections are section 1.2, which characterizes digital convergence in a more precise way, and sections 2.1 and 2.2, which respectively characterize information spaces and hypermedia. Section 2.3 may safely be skipped by readers not interested in the philosophy of media and creation.

perspectives The topics treated in this part can be looked at from multiple perspectives. When you write a paper about any of these topics, as suggested in *projects* paragraph, you should be aware of from which perspective you tackle your subject.

In summary, we can in a non-exhaustive way, distinguish between the following perspectives:

perspectives – digital convergence

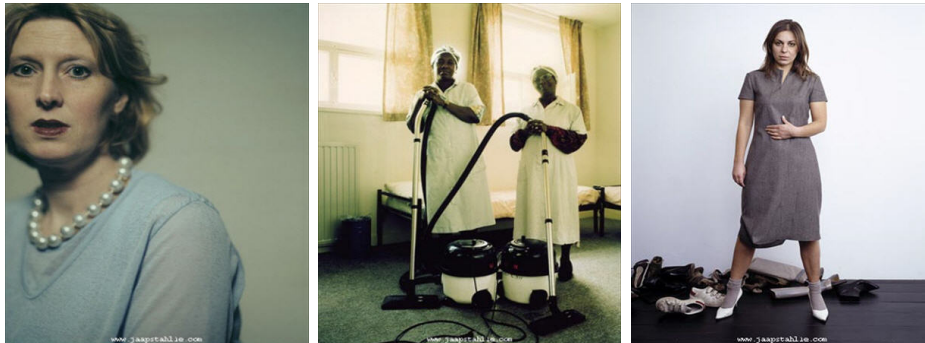
- *historical* – media development timeline
- *sociological* – communication relations
- *psychological* – experience limits
- *aesthetical* – dynamics of creation
- *technical* – divergence & competition
- *philosophical* – re-mediation
- *commercial* – what is the economic model?

As an example, the psychological perspective deals with an interesting issue, namely how much information can we digest and what are the limits to our perceptual system that determines whether the experience offered by a virtual reality interface is really effective.

essay topics The issues treated in these chapters may be used as topics for an essay. As a hint, here are a few titles:

- digital convergence and the future of mobile multimedia
- media @ home – the windows media center
- media art – merging technology and aesthetics

For a first essay, I would suggest a paper no longer than 5 pages. If there are technical details that you do not want to omit, then consider an appendix of 2 to 3 pages. For hints on how to approach writing a paper, see appendix 5.



the artwork

1. alchemy – an illustration from a book about alchemy, from which also the quote is taken, the quote is explained in the *afterthoughts*.
2. signs – ancient chemical symbols, van Rooijen (2003), p. 171, 172.
3. photographs – Jaap Stahlie¹, from portrait series.

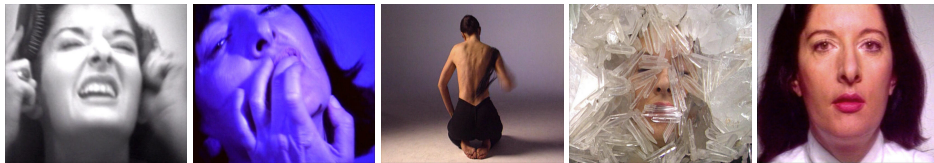
¹www.jaapstahlie.com

1. digital culture

life is becoming digital <XX>

learning objectives *After reading this chapter you should be able to define the notion of multimedia, recount the history of digital entertainment, explain the concept of digital convergence, discuss the future of cyberspace, and speculate about the commercial viability of mobile multimedia.*

We live in the digital era, Negroponte (1995). We are surrounding ourselves with gadgets and we are consuming immense amounts of information, that is increasingly being delivered to us via the Internet. We play games, and we still watch (too much) television. Some of us watch television on our PCs, and may be even looking forward to watch television on their mobile phone. This is multimedia. For others, the PC is still a programmable machine. Being able to program it might earn you a living. Understanding multimedia, however, might even provide you with a better living. In this chapter, we study what trends may currently be observed in the creation and delivery of multimedia information, and we explore what impact the digital revolution may have from a commercial perspective.



1

1.1 entertainment and experience

The question of *what is multimedia* is rather elusive. We may, nevertheless, look at how the phrase *multimedia* is used, and how the concept *multimedia* is related to other concepts. as in the concept graphs that may be obtained with the Visual Thesaurus², providing as input *multimedia*.

²www.visualthesaurus.com

We then see that the notion of multimedia is related to *systems*, in particular interactive and hypermedia systems, and indirectly also to the notion of *transmission*, which will even become more apparent when we inspect the graph for the concept of *medium*, depicted in figure X below.

However, although this gives us some indication of how to position *multimedia* in the larger area of computer applications, in particular when exploring the *systems* node, it does not so much tell us what multimedia is all about.

From the perspective of human cognition, we may look at how multimedia contributes to our understanding of ourselves and the world around us. Traditionally, three levels of cognitive functioning are distinguished, Bruner (1972), corresponding with three levels of meaning:

levels of meaning

- actionary level – action and movements
- sensory/iconic level – images and impressions
- symbolic level – language and mathematics

Multimedia is clearly (most strongly) related to the sensory/iconic level, although for games one could say there is also a strong relation with the actionary level, and to some extent (for both multimedia and games) with the symbolic level.

For a more serious and deep understanding of how multimedia artefacts provide meaning and what role they play in our daily life, or how that meaning is affected by social contexts, we need to take recourse to *semiotic theory*, which is now one step too far, both which we will look at in chapter 12.

Another perspective from which to understand the meaning of *multimedia*, is to look at the function of media in our society, or, in other words, how *multimedia* is situated in our cultural institutions.

Consider this quote from the preface of all of all MIT books in the *Leonardo* series:

cultural convergence

The cultural convergence of art, science, and technology provides ample opportunity for artists to challenge the very notion of how art is produced and to call into question its subject matter and its function in society.

Although the quote is about *art*, it is essentially related to *multimedia*, to the extent that the quote refers to *media art*. The MIT Media Lab³ is one of the world's most famous institutes in the field of multimedia. The *Leonardo* series is a collection of authoritative books on multimedia and related topics, which includes Zielinski (2006), Grau (2003), Wilson (2002).

To understand the position of (computer supported) media in our society, we may observe following Zielinski (2006): there are two forces, political and technological, and there is, currently, a trend towards standardization and uniformity

standardization and uniformity

1. Telematic media were incorporated very quickly in the globalization strategies of transnational corporations and their political administrators and they became increasingly dependent on existing power structures.

³www.media.mit.edu/

2. At the other end of the scale, there were individuals, or comparatively small groups, who projected great hopes onto these networks as a testing ground for cultural, artistic and political models that would give greater prominence and weight to divergence and plurality.

This reflects what Zielinski (2006) calls the *advanced media paradox*, facilitating heterogeneity and immersion on the one hand, and striving for universalisation on the other hand, as demanded by the centers of technological and political power.

Leaving the socio-political arena, we may in some sense predict the tension between *convergence* and *divergence*, by looking at the meaning context of the concept of *convergence*, again using the Visual Thesaurus, where we find that not only notions such as *overlap* and *occurrence* are related to it, but also the complementary concept of *divergence*. However, instead of speculating on the meaning of words, it might be more worthwhile to look at what we may consider to be the recent history of multimedia, entertainment.

entertainment

In november 2000, a theme issue of the Scientific American appeared, featuring a number of articles discussing (digital) entertainment in the era of digital convergence. Let's start with a quote:

Scientific American (november 2000)

The barriers between TV, movies, music, videogames and the Internet are crumbling. Audiences are fetting new creative options. Here is what entertainment could become if the technological and legal hurdles can be cleared ...

Moreover, the editors made some wildly speculative claims, such as *digitizing everything audio and video will disrupt the entertainment industry's social order, and the whole concept of holding a CD or movie in your hand will disappear once d-entertainment is widely available*. To some extent this seems already to be true, as for example the music industry can painfully testify to.

Underlying the importance of entertainment in the era of digital convergence is the premisses governing an entertainment economy, which may be stated as

there is no business without show business

Additionally, the authors of the introduction to the theme issue speculate that *the creation of content will be democratized*, due to the availability of low cost digital movie cameras and PC video editors. Producing a video movie is now possible for just a few thousand euro or dollars. However, given the aesthetic ignorance of the average individual making video movies, it seems doubtful that this will hold true for entertainment in general.

In that same issue of the Scientific American, Gloria Davenport, a pioneer in the field of multimedia, presents list of applications characterizing the evolution of digital entertainment, Davenport (2000):

evolution of digital entertainment

- 1953: Winky Dink (CBS) – interactive television, drawing exercise
- 1972: Pong (Atari) – ping-pong on computer screen
- 1977: Adventure – text-based interactive fiction
- 1983: Dragon's Liar – laser-disc technology 3D game
- 1989: SimCity – interactive simulation game
- 1989: Back to the Future – the Ride
- 1993: Doom – 3D action game
- 1995: The Spot – interactive web-based soap opera (Webisodic)
- 1999: IMAX3D – back to Atlantis (Las Vegas)
- 2000: Big Brother – TV + around the clock Web watch + voting
- 2001: FE Sites – fun enhanced web sites

It is interesting to note that *Big Brother*, which was originally created by a Dutch team, has become a huge success in many countries. Although the integration with the web was limited, it may be seen as the start of a number of television programs with web-based interaction facilities.

digital experience

The list compiled by Gloria Davenport suggests, a convergence towards an 'ultimate digital experience', Now, what does *digital experience* mean?

In a special issue of the Communications of the ACM, about the next 1000 years of computing, Ramesh Jain makes the following observation, Jain (2000):

The desire to share experiences will be the motivating factor in the development of exciting multimedia technology in the foreseeable future.

Considering the variety of means we have at our disposal to communicate, as reflected in the list below, we may wonder whether our current technology really stands out as something special.

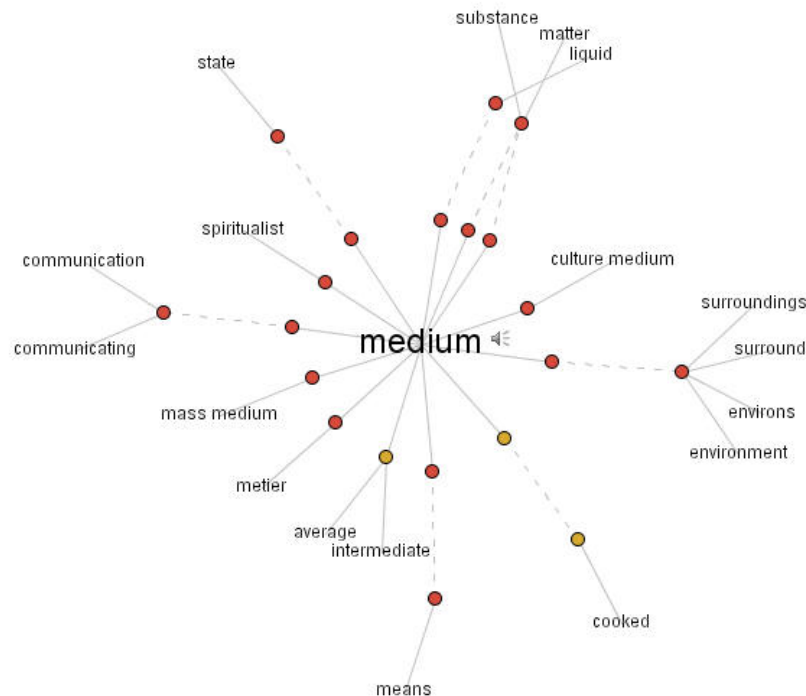
communication technology

- *oral* – communicate symbolic experiences
- *writing* – record symbolic experiences
- *paper* – portability
- *print* – mass distribution
- *telegraph* – remote narrow communication
- *telephone* – remote analog communication
- *radio* – analog broadcasting of sound
- *television* – analog A/V broadcasting
- *recording media* – analog recording
- *digital processing* – machine enhancement
- *internet* – multimedia communication

According to Ramesh Jain, internet-based multimedia communication differs from earlier communication technology in that it somehow frees the message from the medium. Reflecting on Marshall McLuhan phrase – *the medium is the message* – he observes that:

the medium was the message when only one medium could be used to communicate messages.

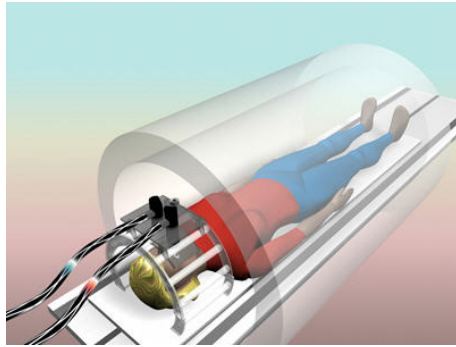
Now, that the Internet allows the synthesis and rendering of information and experiences using whatever is the most appropriate media to convey the message, the message is, as Jain phrases it, just the message, and the medium is just the medium. In other words, the medium itself does not seem to constrain what message can be conveyed. Looking at the documentary *Fahrenheit 9/11* though, we may seriously doubt whether this is true. Although it is possible to gain knowledge about the alliances that underly politics, even in the age of the internet, the television campaigns seem to be more dominant in affecting the general public's opinion about global politics than anything else, due to the conventional formats of presentation and editing.



Let's once more look at a graph, above, indicating the concept relations for the notion of *medium*. What strikes me as important are the relations with the distinct concepts of *substance*, *communication*, *environment*, and *intermediate*. In

some respects the notion of *medium*, underlying the plural use of it in *multimedia* is comparable to the notion of *ether*, which was once seen as a vehicle for the transport of broadcasted information. But I also like to stress the 'substantial' aspect of multimedia, as a material for design and creation, similar to paint.

The basic issue here is what is a medium and how does it affect, or even shape our experience(s). Following Ramesh Jain, we may speculate that the range of sensory information offered by multimedia applications may become much richer than is currently the case, and we may then predict that there will be a tremendous progress in presentation technology, multisensory presentation technology! Clearly, from a technological perspective there seems to be no limit, except those imposed by our own phantasy. However, it should be equally obvious that compelling experiences rely on carefully staged presentations, and as such require an entirely new discipline of design.



VR for pain relief

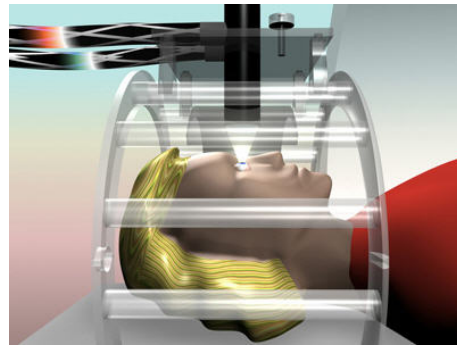


image delivery system

3

example(s) – *VR for pain relief*

The research project fMRI Research on Virtual Reality Analgesia⁴ at the Human Interaction Laboratory (Washington) has explored the use of VR to reduce the agony of taking MRI scans. The U.W Radiology Digital Imaging Science Centers wide field of view magnet-friendly virtual reality image delivery system makes it possible for volunteers and patients to have the illusion of going into virtual reality during fMRI brain scans. As explained on the website, the image on the left above, shows a woman in virtual reality during an fMRI brain scan, looking into a custom magnet-friendly virtual reality goggles. VR images from projectors in another room are carried to the participant in the form of light (photons, no electrons) via optic fiber image guides. The participant has the illusion of going inside the virtual world, allowing researchers to measure what happens to her brain when she reports reductions in pain during VR. The white cage-like structure around the womans head, in the image on the right, shows fMRI receiver coils used by the

⁴www.hitl.washington.edu/research/magnet

fMRI brain scanner to collect the information about changing patterns of brain activity.

Another project investigating the use of VR techniques for pain distraction can be found at the site of the Virtual Environments⁵ of the Georgia Institute of Technology, Atlanta.

research directions– *the face of cyberspace*

The notion of *cyberspace* was introduced in William Gibson's novel *Neuromancer*, that appeared in the early 1980's, signifying a vast amount of (digital) data that could be accessed only through a virtual reality interface that was controlled by neuro-sensors. Accessing data in *cyberspace* was not altogether without danger, since data protection mechanisms (including firewalls, as we call them nowadays) were implemented using neuro-feedback. Although the vision expressed in *Neuromancer* is (in our days) still futuristic, we are confronted with a vast amount of information and we need powerful search engines and visualisation techniques not to get lost. So what is the reality of *cyberspace* today?

... cyberspace is a construct in terms of an electronic system.

as observed by Vivian Sobschack, 1996, quoted from Briggs and Burke (2001), p. 321. On reflection, our (electronic) world of today might be more horrendous than the world depicted in *Neuromancer*. In effect,

cyberspace

television, video cassettes, video tape-recorder/players, video games, and personal computers all form an encompassing electronic system whose various forms interface to constitute an alternative and absolute world that uniquely incorporates the spectator/user in a spatially decentered, weakly temporalized and quasi-disembodied state.

All these gadgets make us dizzy, stoned with information and fried by electromagnetic radiation. However, the reality of everyday computer use is (fortunately?) less exciting than the images in *Neuromancer* suggest. User interfaces are usually tiresome and not at all appealing. So except for the fanatic, the average user does easily get bored. Would this change when virtual reality techniques are applied pervasively? What is virtual reality?

virtual reality

virtual reality (is) when and where the computer disappears and you become the 'ghost in the machine' ...

In other words, virtual reality is a technology that provokes immersion, sensuous immersion, supported by rich media and powerful 3D graphics. In our age of information, we may wonder how all that information should be presented. Rephrasing the question, we may ask what are the limits of the digital experience, or more importantly, what should be the norm: 3D virtual environments, plain text, or some form of XP?

⁵www.gvu.gatech.edu/virtual

1.2 technological developments

Let's see if we are able to give a more precise characterization of *digital convergence*. In their introduction to the theme issue of the Scientific American, Forman and SaintJohn locate the beginning of digital convergence, historically, at the 1939 New York World Fair, and more in particular the RCA Pavillion, which should be considered as the formal debut of television broadcast. They observe that

history

the receiver at the RCA Pavillon was way ahead of its time, it was a combination of television - radio - recorder - playback - facsimile - projector ...

Moreover, they remark that this *in hindsight suggests that we humans have a fundamental desire to merge all media in one entity*.

By way of definition we may state, following Forman and SaintJohn, that digital convergence is:

digital convergence

the union of audio, video and data communication into a single source, received on a single device, delivered by a single connection

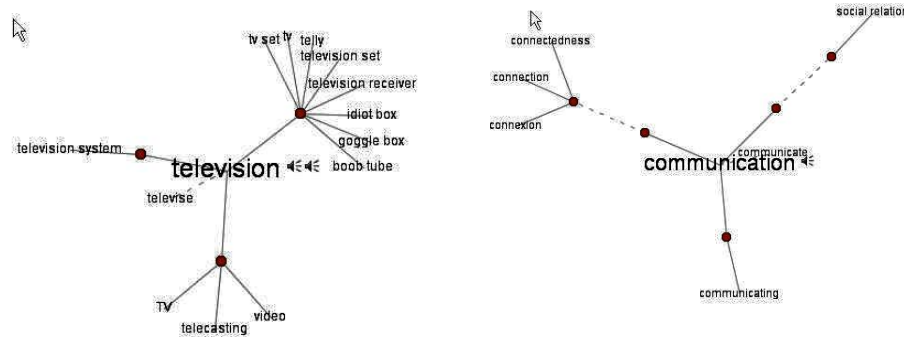
And, as they say, *predicted for decades, convergence is finally emerging, albeit in a haphazard fashion*.

Taking a somewhat closer look, we may discern subsidiary convergences with respect to content, platform and distribution:

subsidiary convergences

- *content* – audio, video, data
- *platform* – PC, TV, internet, game machine
- *distribution* – how it gets to your platform

Here, Forman and SaintJohn continue by speculating that if compatibility standards and data protection schemas can be worked out, all d-entertainment will converge into a single source *that can shine into your life on any screen, wherever you are ...* However, observe that the number of competing standards and architectures is enormous, and that apart from the technical issues involved it is not entirely clear what business model should underly such convergence. In computer shops, there PCs with TV receivers are sold in the range of 1000-2000 euro. This does not include the screen. They come with either the XP Home or Windows Media Center. One of the first in this line of machines, in the higher prices range, was the Sony W1.



4

TV or PC

It is fair to say that no device has changed the way we live so dramatically as television. Television, for one, has altered the way we furnish our living rooms, not to speak about the time we waste watching the thing. Comparing the graphs for *television* and *communication*, we immediately see that their underlying concepts are very different. And more specifically, the association of television with a phrase such as *idiot box* may raise doubt whether the promise of convergence, which does include communication as an essential feature, will ever become a reality.

Now, we may wonder what interactive television and enhanced television have to offer us. Looking back, we may observe that it takes some time for the new possibilities to catch on. For example, interactive television was introduced in 1970, but apparently people did not want to communicate with the broadcaster. As another example of enhanced television, take Big Brother. Although many people watched Big Brother when it first appeared on television, the willingness of the audience to react other than by phone was (apparently) somewhat disappointing. Perhaps, in the Netherlands this was due to the fact that only a fraction of the PC owners was, at that time, permanently online.

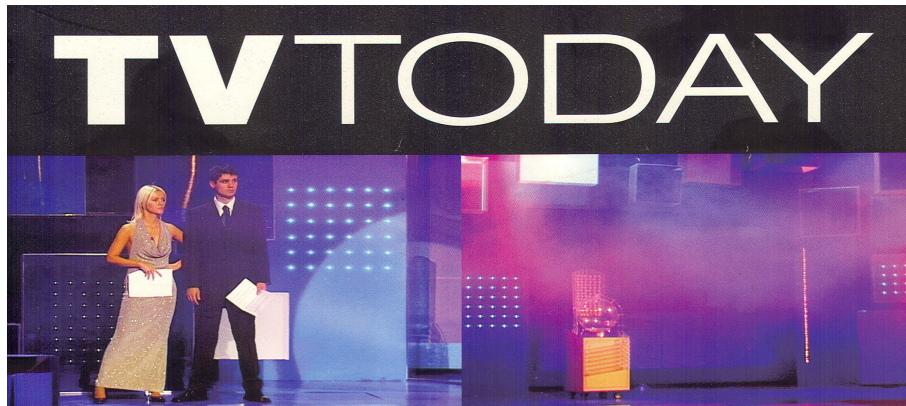
In spite of the failed experiments, Forman and SaintJohn state, somewhat optimistically, that *the convergence of digital content, broadcast distribution and display platforms create the big convergence of d-entertainment and information with feedback supporting human interactivity*.

Before looking at *digital television* more closely, let's summarize what digital convergence involves:

convergence

- *content* – 2D/3D graphics, data, video, audio
- *distribution* – broadcast, wireless, DVD, internet, satellite, cable
- *platform* – PC, television, game machine, wireless data pad, mobile phone

This summary indicates the technical opportunities, and the possible functional extensions that may enhance our use of television, computer, game console and mobile phone. As concerns digital television, we may come up with some immediate advantages, such as enhanced resolution, a multiplication of channels, and (more relevant to the issue of convergence) interactive television.



exposition on the history of TV in Institute for Time-based Arts/Montevideo⁶

5

To get you familiar with some common acronyms, when speaking about (digital) television, we must make a further distinction between:

- HDTV – high definition television
- SDTV – standard definition television
- ITV – interactive television

As further discussed in chapter 3, we have (standard) codecs for d-TV, in particular MPEG-2, for recording digital video, and MPEG-4, for high-quality streaming video on the internet, both from the Motion Picture Expert Group, that enable the effective delivery of digital video, possibly in combination with other content.

Unfortunately, experts disagree on what might become the most suitable appliance or platform to consume all those digital goodies. Here is a list of possible choices:

a *killer* d-TV appliance ...

- DVD player/recorder – 400.000 sold in 2 years, 2h of MPEG-2 video
- personal television – TiVo, Replay-TV (MPEG-2 cache)
- game machine – Sony PS 2, X-Box

Will we prefer to watch stored video, instead of live television broadcasts? Will the Internet be able to compete with traditional television broadcasting. Will DelayTV or Replay-TV, which allows you to watch previous broadcasts at a time that suits you become popular? Will an extended game machine or PC replace your television? Currently, we must observe that streaming media (still) have rather poor resolution.

Leaving game machines aside, will it then be the TV or PC that will become our platform of choice? Forman and SaintJohn observe:

TV or PC

The roadblock to the Entertainment PC could be the PC itself. Even a cheap TV doesn't crash or freeze. The best computers still do.

However, they conclude that it might make sense to adopt a programmable PC that can support competing TV standards, rather than construct a stack of TV peripherals. Nevertheless, there are a number of problems that occur when we (collectively) choose for the PC as our platform for d-entertainment. Should we have thin clients, for example based on the Sun/Java platform or so-called fat clients based on some version of Microsoft windows? How do we handle the fact that the current internet protocols are not robust, and how can we provide what is known as *quality of service*? Should we adopt any of the proprietary architectures and codecs, such as RealVideo, QuickTime, Windows media, or should we adhere to an open standard such as MPEG-4?

Evidently, the situation becomes even more complex when we just consider the range of alternatives for connectivity, that is for possible ways of distributing contents:

distribution

- *telephone network* – from 0.5 - 2 Mbps to 60 Mbps (2.5km)
- *broadcast TV* – 6 MHz / 19 Mbps (4 channels MPEG HDTV)
- *cable TV* – hybrid fiber-optic coaxial cable 6 Mbps
- *fixed wireless* – 2 Mbps (radiotowers + rooftop antenna), phones/handhelds
- *satellite* – downloads to 100kbps, modem for uploads ...

Most probably, convergence with respect to distribution will not result in one single way of being connected, but rather a range of options from which one will be selected transparently, dependent on content and availability.

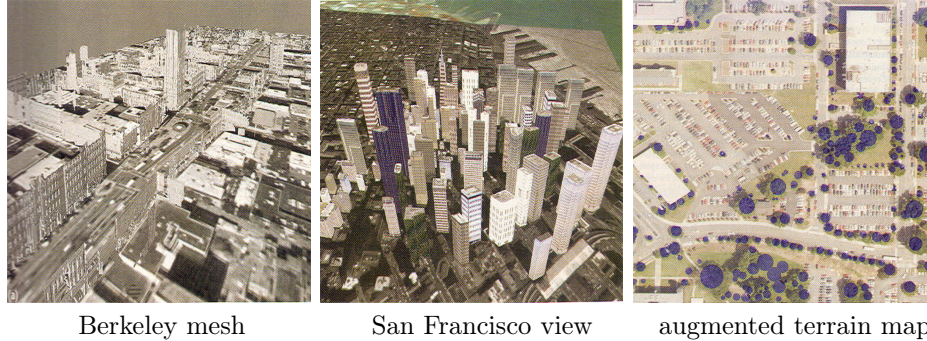
Let's stay optimistic, and ask ourselves the following question:

what will we do with convergence once we have it?

One possible scenario, not too unlikely after all, is to deploy it for installing computing devices everywhere, to allow for, to name a few, smart houses, smart clothes, or, in other words, to create a smart world. I wonder what a smart world will look like. In the end we will have to wait and see, but whatever will emerge

we will watch

That is to say, it is not likely that we will have a world without television. Television as we are used to it seems to be the dominant paradigm for d-entertainment, for both the near and distant future.



6

example(s) – *visible world*

Just imagine that every visible place on earth would be accessible in a virtual world. Researchers of the Georgia Institute of Technology⁷, Atlanta, have developed software for the semi-automated construction of detailed interactive urban environments, that takes data from multiple sources, including geo-corrected imagery from aerial photography and satellites and ground-based close-ups, Rosenblum and Macedonia (2002).

The challenge here is to collect data from multiple sources and convert this into models, and perhaps even more difficult, to make the models visible so that they can be navigated in an interactive fashion. Recently, the Georgia group teamed up with a group from Berkeley to develop more complex models (images on the left), and together they are working on automating the extraction of information from aerial pictures (image on the right), in particular the detection of groups of trees, and height estimation.

There are many applications for such technology, including urban planning, emergency response, tourism and entertainment, military operations, traffic management, construction and maintenance, mobile services, citizen-government relations, and (not in the least) games.

The next step might be to connect the cameras, that are already there in many of these places, to the model, to observe what happens there in real life. But, somehow, this vision becomes frightening.

However, if you want to give it a try yourself, and populate the virtual globe with your own creations, go download the viewer and editing tool from *Google Earth*:

Google Earth

- Earth – earth.google.com
- SketchUp – sketchup.google.com/download.html

and read the tutorials!

⁷www.gvu.gatech.edu/datavis/research

research directions– *technological determinism*

Although there are many technical issues involved in (digital) multimedia, as exemplified in the issues that play a role in digital convergence, a technical perspective alone does not suffice. Each technological innovation has its consequences on our social life. Conversely, each trend in society might result in the adoption or development of new technology. Looking at the history of the media, we may observe that media become *materials* in our social processes. Or, as phrased in Briggs and Burke (2001):

media as materials

each medium of communication tended to create a dangerous monopoly of knowledge

For example (Briggs and Burke (2001), p. 8) for Christians, images where both a means of conveying information and a means of persuasion, that is part of the rhetorics of institutionalized religion.

Looking at our age, and the media that have come into existence in the previous century (radio, television, ...), Briggs and Burke (2001) observe that:

technological determinism

technological determinism was not the answer, ... more attempts were to be made to provide answers about the social consequences of television than had ever been asked about radio.

In effect, underlying all developments in the media (including the computer) we may assume a basic need for information. A rather problematic need, for that matter:

information

Information became a major concern anywhere during the late 1960 and 1970s where there was simultaneous talk both of 'lack of information' and 'information saturation'.

Briggs and Burke (2001), p. 555

Nowadays, we regard information as a commodity. Train schedules, movies, roadmaps, touristic information, stock prices, we expect it all to be there, preferably online, at no cost. No information, no life. Information drives the economy. Upwards and downwards!

1.3 multimedia applications

In many stores there is a multimedia section. In some stores you will see B-movies being announced as *multimedia topper*. In other stores, the multimedia sections has a large offering of computer peripherals, ranging from DVD-RW drives to webcams and TV on PC hardware. Elsewhere you may buy authoring packages to organize your cell-phone photos, your family photo and video album, to create your personal archive on DVD. All this might make you wonder whether

multimedia is serious business. See figure (a) and (b), illustrating our personal *memex*, as explained below.

But more seriously, what is the commercial impact multimedia and in particular digital convergence may have? And, perhaps equally important, why should we be interested in this from, I must say, an academic perspective?

In this last section of the introductory chapter, we will look at some popular press items related to new media, mixed media (in particular the merge of TV and internet) and mobile multimedia. We will then briefly reflect on what significance these issues have from our academic perspective.



7

new media

As you may be read in the newspapers in the beginning of this century, large investments have been made (by both cable and telephone companies) to improve the technological infrastructure for the new media. Simultaneously, joint ventures have arisen between content developers and providers, as with the Dutch Endemol company.

Now, what does the popular press have to say about all these developments. Here is one comment, from a Dutch newspaper:

Peter Greven 23/3/2001 (Volkskrant)

new media sucks – people like new technology. they don't like new media.

The translation from Dutch is, admittedly, mine. It says, in other words, that people like to receive the old stuff on new gadgets, but that they are not willing to pay for any new sort of services. For example, when considering the smart video recorder, that uses a disk cache for storing MPEG coded versions of broadcasts, just think of other gadgets and services that didn't make it or that are encountering problems in being accepted. Some famous examples from the past are the videofoon, videotext, cd-i, and DCC.

Perhaps the reason for these failures is the *trial-and-error* method,, also referred to as the spaghetti method, that is being followed in developing new media. As characterized by Jan van Dijk, of a dutch university in the east of the Netherlands (Twente), the spaghetti method consists of throwing a plate against

the wall, and see what will stick to the wall. In other words, just throw your product on the market and see whether it will stick. Perhaps that is not the right method to be followed. But can you think of a better one?

In many cases 'the market', that is the people using a service, do not behave as expected. For example in Sweden, the upload of material far exceeded download, which is contrary to the assumptions underlying ADSL.

TV meets the Web

At first sight it seems promising to develop mixed media. As an example, a dutch agency announced services to support the integration of TV and the Web, promising the integration of

www.tvmeetstheweb.com

streaming media (audio and video), interactive gaming, virtual reality and 3D animation, interactive TV programming, interactive advertising, video on-demand, webcasting and multimedia

In 2000 they issued a report sketching the European broadband landscape. Quoting from this report: *The advent of broadband Internet access, which has been available in the US for some time but is only now beginning to make inroads into Europe, makes a whole range of new services possible. As download speeds have increased and more bandwidth has become available, the possibility of delivering screen-based content such as films, television programs and music has moved a step closer to mass market usage.* With respect to the adoption of cable or DSL in Europe, they observe that despite the fact that cable companies have gained firm ground, there is an even larger number of conventional telephone lines, around 180 million. In contrast, there are only 15 million cable subscribers, giving DSL a large potential audience. Matthijs Leendertse, co-author of the report, observes: *Gaining competitive advantage and future revenue in Europe's broadband landscape will depend heavily on a company's ability to offer integrated services: access (fixed and wireless) and content. It is virtually impossible at this point for one single company to offer these services on a pan-European level. This means that companies need to find partners to fill the gaps in their offerings.* Let me assure you, at the moment you will be reading this the battle is still going on!

mobile multimedia

Let's look at another potential hype. In 2000, Webnoize published a report (by Matt Bailey), entitled *Wireless Entertainment: What Is It Worth?*, which introduces the *wireless web*, and predicts that *young media junkies* will demand music videos and animations, and listen to wirelessly streamed music.

The intent of the report is to investigate whether investments in the mobile entertainment are justified. The report examines how providers of music and video services can benefit from the wireless delivery of multimedia. Using survey evidence, pricing information from new wireless networks and interviews with

industry visionaries, the report analyzes supply and demand to build an economic and business model for mobile multimedia.

Apart from the need to invent some business model, there are a number of strategic questions to be answered in order to estimate the risk of making investments in this direction. Following Bailey, we may list questions such as:

strategic questions

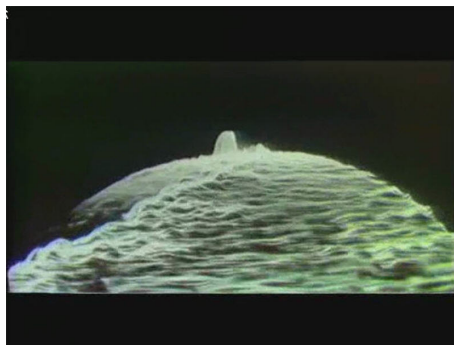
- how quickly will wireless connectivity speeds improve?
- what is the demand for services that deliver music and video to wireless devices?
- how can suppliers of multimedia services monetize demand for wireless access?
- how much will it cost to stream multimedia content to wireless devices now and in 2006?
- are consumers willing to compromise quality for lower cost?

And more. If you are interested whether anyone is willing to take such risks and invest in mobile multimedia, just look at what players were involved.

the players

Alltel, AT&T Wireless, AtomShockwave, Cingular Wireless, Clear Channel, HitHive, Ifilm, Infinity, KDDI, Liquid Audio, LMIV, Mannesmann, MP3.com, MTV, NetCom, Myplay, Nortel Networks, NTT DoCoMo, Omnitel, Sprint, Telefonica, Telstra, Vitaminic, Verizon Wireless, Virgin Megastores, Vodafone, Voicestream.

Now make up your mind, and ask yourself the question whether multimedia is worth your (intellectual) investment.



Vasulka Objects (1)



Vasulka Objects (2)

the academic perspective

Being sensitive to hype is only too human. So also academics may be fascinated by new trends. and get distracted by rumors on the market. Breaking loose from this fascination, we may ask ourselves what are the real issues, and what makes multimedia interesting. Let me start with answering the latter question first. As

an academic subject, multimedia is interesting because it offers such an intriguing mix of subjects, including multimedia technology, exploratory design and scientific validation. Commercially, it is safe to say that the volume of entertainment related multimedia content, including games, music and infotainment is substantial, and hence its economic interest is indisputable. But what are the real issues?

One of the examples of multimedia applications I will present in the last chapter is an application in the domain of cultural heritage. For this domain we have developed so-called *digital dossiers* containing a representation of the work(s) of a particular artist as well as information that characterizes the work in its historic and cultural context, needed for the re-exposition or installation of the work. Problems facing the developer of a digital dossier cover the interaction of the user with the dossier, the presentation of both textual and multimedia information, and facilities for search and navigation. And there are technical issues, such as which codecs to select for the videos and how to manage the content included in the dossier. Developing a dossier is not as one might naively think the creation of content only, but rather involves designing the functionality of the application as well.

Generalizing from the domain of cultural heritage to the area of infotainment and multimedia information systems, where an integrated presentation of textual and multimedia information must be achieved, we may boldly state that designing the functionality of the application is the most crucial issue, and as such of primary academic interest. All other topics, including multimedia technology, compression algorithms, software engineering, multimedia platform support and information retrieval techniques, may be regarded in some loose sense to be subservient to the issue of design.

digital art As the illustrations in the text testify, another personal motivation for being involved in multimedia comes from the area of digital art. And, with students I observe a similar interest in the potential digital content authoring offers as a vehicle for personal expression.

One of the artists of which I included material in this book is Woody Vasulka, who was a pioneer in the early days of video and computer art. In an interview, held in 1985 with Rene Coelho, the founder of Montevideo⁸, Vasulka explained his fascination with the scan processor and later the video computer by stating that it allowed him to *invent the image*. Still, however, as he said, in some sense traditional painting acted as a visual reference system by which to judge the images produced with the new technology. Later in the interview, he observed that after some time he became bored with the images produced this way, and he started to feel the need to include more narrative in his work. His wife, Steina Vasulka, with whom he founded the Kitchen, a gathering place for new media artists in New York in the 1970s, remarked that in the early phase she was struck by the fact that *the material was so friendly*, that is how easy it was to express your ideas.

These words suffice to emphasize the importance of the motivation you might

⁸www.montevideo.nl

get out of the material, to be susceptible to as Brancusi phrases it *the rethorics of the material*, even when you are an academic.



9

example(s) – *personal memex*

Just imagine that you would store all your photographs, SMS messages, emails, and in addition to that record your physiological condition, using body-wearable sensors. These data can then be uploaded to your PC, and later to a mass storage server, so that they can be used in your medical dossier, to improve your performance in your favorite sport, or to augment your memory when recollecting stories about your holidays or travels. Impossible? Not at all, Disk space will be cheap. Your body may act as a network to connect the body wearable devices, and, after all, most of the gadgets do already exist! Besides, the idea is not new. See section 2.2 for early visions of the *memex*.

research directions – *the information society*

There is no doubt about it, we live in an information society. But do we know what an information society is?

In Briggs and Burke (2001) (p. 187), the functions of the media are summarized as

functions of media

information, education, entertainment

So, perhaps, we could better state that we live in a *media society*. So far, in the latter part of the previous century, television has dominated our lives, and observe that (following Ernie Kovack, cited from Briggs and Burke (2001)):

medium

television is a medium 'because it is neither rare nor well done'

Back to the main issues, what is an *information society*? According to Briggs and Burke (2001):

information society

the new term 'information society' gave form to a cluster of hitherto more loosely related aspects of communication – knowledge, news, literature, entertainment, all exchanged through different media and different media materials – paper, ink, canvas, paint, celluloid, cinema, radio, television and computers. From the 1960s onwards, all messages, public and private, verbal and visual, began to be considered as 'data', information that could be transmitted, collected, recorded, whatever their point of origin, most effective through electronic technology.

So, from the varieties of perspectives we have discerned, including technological perspectives, societal perspectives and psychological perspectives, we must investigate the problem of communication:

communication

- *what* – content
- *who* – control
- *whom* – audience (how many)

That is, simply, who says what to whom in what channel with what effect?! The remainder of the book will, however, will treat these issues mainly from a technological perspective. In the chapters that follow, we will enquire after the technological assumptions that make an information society possible.

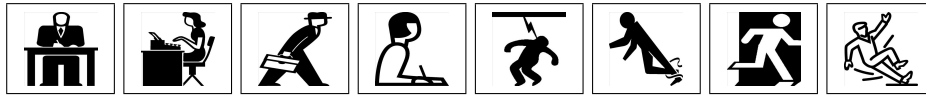
1.4 development(s) – convergence in second life

第二生命

Second Life seems to be overtaking the world. In the whole range of community-building platforms, Second Life stands out as an immersive 3D world with an almost stunning adoption, by both individuals, companies and institutions, followed attentively by the Press. Not entirely without an understanding of the value of press coverage, the VU University Amsterdam decided to create presence in Second Life, by creating a virtual campus, to realize a (virtual) community of learners, VUSL. And, indeed, we succeeded in being the first university in The Netherlands with presence in Second Life and, as hoped, this was covered in the 8 o'clock nation-wide TV news.

More substantial than getting into a nation-wide television broadcast, however, is our aim to communicate our institutional goals, *creating a community of learners*, by creating a virtual campus in Second Life, offering an *information portal* as well as a *meeting point*, in a media platform that is widely adopted by our target community. Virtual presence in Second Life, obviously, is not enough. The relatively long history of virtual worlds has shown that lack of interesting content and functionality easily leads to boredom, desinterest, and hence *churn*, users

dropping off. As a consequence, there is a need for sustainable functionality, that both motivates people to come back and participate, and, otherwise why choose Second Life, makes essential use of the 3D immersive environment offered by Second Life. In this paper, we will explore how to use web services in meaningful compositions or mashups to enhance our presence in Second Life, and create a community where visitors actively participate in both education and research,



10

questions

digital convergence

1. Sketch the developments in *multimedia*. What do you expect to be the commercial impact of multimedia in the (near) future?

concepts

2. Explain what is meant by *digital convergence*.
3. Which kinds of (*digital*) *convergence* do we have?
4. Discuss the relation between the *medium* and the *message*.

technology

5. Give a brief sketch of the development of *digital entertainment*.
6. Characterize: HDTV, SDTV, ITV.
7. Discuss convergence with respect to *platforms*.
8. Discuss convergence with respect to *delivery*.

projects & further reading As a project, consider the development of a Java-based mobile game using J2ME, see Morrison (2005), or a web-based game using Visual Basic .NET, see Santos Lobao and Hatton (2003).

You may further explore multiplatform game development, and find arguments to choose for either Java-based or managed code based implementations.

For further reading, I advice to have a look at the special issues of the Scientific American, American, and the CACM on the next 1000 years of computing, CACM (2001), and, for getting an idea where this all leads to, Schneidermann's *Leonardo's laptop*, Shneiderman (2003).

the artwork

1. photographs of art works by Marina Abramovic, *Art must be beautiful*, *Blue period*, *Dissolution*, *Dozing consciousness*, *In between*, with (pending) permission from Montevideo⁹. See also section 10.2.
2. *medium*, according to the Visual Thesaurus¹⁰.

⁹www.montevideo/nl

¹⁰www.visualthesaurus.com

3. fMRI Research on Virtual Reality Analgesia¹¹, see section 1.1.
4. *television* and *communication*, according to the Visual Thesaurus.
5. TV Today, exhibition at Montevideo, februari 2005.
6. visible world – taken from Rosenblum and Macedonia (2002), see section 1.2.
7. personal event database and personal gadgets, from Freeband¹² project.
8. *Thomas Lips 1975*, *Thomas Lips 1993*, from Marina Abramovic, with permission from Montevideo.
9. *scanlines* from Woody Vasulka¹³, 197x, with permission from the artist.
10. signs – people, van Rooijen (2003), p. 254, 256.

The work of Marina Abramovic has a strong *existential* flavor. It has also served as the material for a case study in developing a digital artist dossiers, the *abramovic dossier*, discussed in section 10.2. The work of Woody Vasulka is of a more *experimental* character, and shows the joy of discovering the possibilities of the, at the time, new electronic and digital tools and materials.

¹¹www.hitl.washington.edu/research/magnet

¹²www.freeband.nl

¹³www.vasulka.org

2. hypermedia information spaces

everything must be intertinkled <XX>

learning objectives

After reading this chapter you should be able to define information spaces in a precise manner, position the hypertextual capabilities of the web in a historical perspective, explain the difference between multimedia and hypermedia, and argue why computational support for narrative structure in multimedia applications is desirable.

However entertaining it might be presented to you, underlying every multimedia presentation there is an information space. That is to say, irrespective of the medium, there is a message. And being confronted with a message, we might want to inquire for more information. In this chapter, we will define the notion of information space more precisely. We will extend this definition to include information hyperspaces, by looking at the history of hypertext and hypermedia. Finally, we will discuss visualisation as a means to present (abstract) information in a more intuitive way, and we will reflect on what is involved in creating compelling multimedia.



2.1 information and data

Current day *multimedia information systems* distinguish themselves from older day information systems not only by what information they contain, that includes multimedia objects such as images and sounds, but also by a much more extensive repertoire of query mechanisms, visual interfaces and rich presentation facilities. See Chang and Costabile (1997).

Preceding the advent of multimedia information systems, which include networked multimedia systems as discussed in section 6.3, we have seen advances in

multimedia information systems

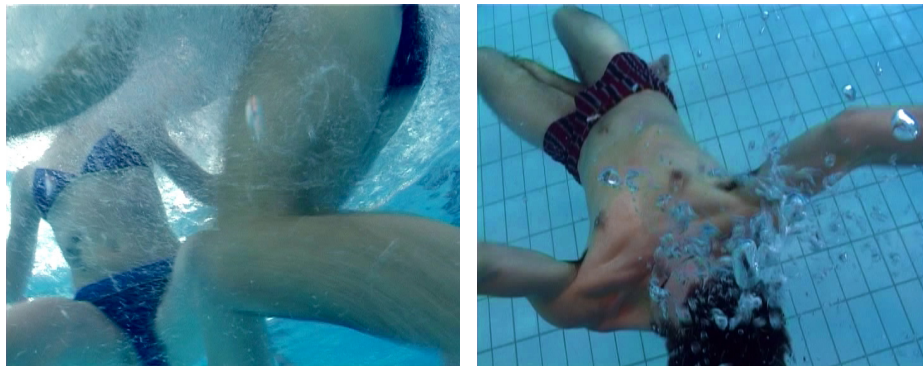
- *storage technology – multimedia databases*
- *wideband communication – distribution accross networks*
- *parallel computing – voice, image and video processing*
- *graphic co-processors – visual information with high image quality*

Now, the class of *multimedia information systems* is, admittedly, a large one and includes applications and application areas such as:

geographical information systems, office automation, distance learning, health care, computer aided design, scientific visualization, and information visualization.

Nevertheless, irrespective of what technology is used for storage and retrieval, multimedia information systems or multimedia databases impose specific requirements, with respect to: the size of data, synchronisation issues, query mechanisms and real-time processing.

Partly, these requirements concern the efficiency of storage and retrieval and partly they concern aspects of usability, that is the way information is presented to the user. In particular, we can think of a great number of query mechanisms that our multimedia information system of choice is expected to support: free text search, SQL-like querying, icon-based techniques, querying based on ER-diagrams, content-based querying, sound-based querying, query by example, and virtual reality techniques.



logical information spaces

But before thinking about the optimal architecture of multimedia information systems or the way the information is presented to the user, let's consider in what way a multimedia (information) system or presentation may be considered an *information space*.

As a tentative definition, let's assume that

an information space is a representation of the information stored in a system or database that is used to present that information to a user.

This may sound too abstract for most of you, so let's have a look at this definition in more detail.

First of all, observe that when we speak of representation, and when we choose for example a visual representation, then the representation chosen might be either the users conceptualization of the database, or a system generated visualization. In principle the same holds for a text-based representation, but this is far less interesting because the options in choosing a representation and presenting it to the user are much more limited.

Unfortunately, the phrase *representation* is also somewhat vague. To be more precise, we must distinguish between a *visual information space* (for presentation), a *logical information space* (in which we can reason about abstract information objects) and a *physical information space* (where our concrete multimedia objects are stored).

Summarizing we have:

- *physical information space* – images, animations, video, voice, ...
- *logical information space* – abstract database objects
- *presentational information space* – to present information to the user

Our visual information space, our presentation space, as you may prefer to call it, might reflect the logical information space in a symbolic manner by using diagrams, icons, text and possibly visualizations, or, going one step further, it may also mimic the logical information space by using virtual reality, as discussed in chapter 8.

Now we can give a more precise definition of the notion of information space, in particular *logical information spaces*:

a logical information space is a multidimensional space where each point represents an object from the physical information space (read database).

First of all, observe that when we speak of dimensions we might also speak of attributes that can take either continuous, numerical, discrete or logical values. So, concretely, these attributes may be directly or indirectly related to information stored in the database, and hence we can give a more precise definition of the notion of (multimedia) information objects, queries and *cues* (in the logical information space):

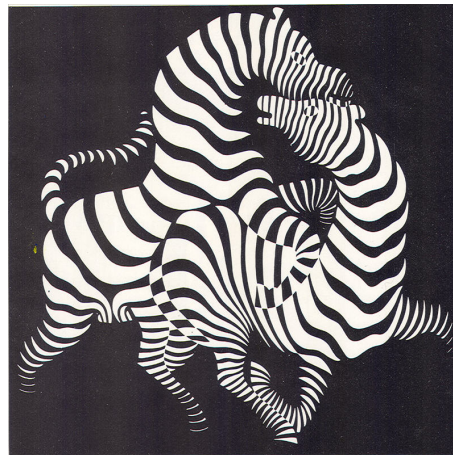
- *information object* – a point in the (logical) information space
- *query* – an arbitrary region in this information space
- *clue* – a region with *directional information*, to facilitate browsing

The notion of *clue* is actually quite interesting, since both examples and queries may be regarded as clues, that facilitate browsing through the contents of an information space. As an example, just think of the situation that, when looking for another notebook, you want something that is similar to the the thing you've previously seen, but that has an additional video output slot that may be connected to your TV.

Also, clues are needed to allow for *query by example*. In this case you need to help the user to define a query in the logical information space, so that the system can construct an *optimal query* to search for the desired object(s) in the physical information space.

When we regard *the information retrieval problem* to be the construction of the *optimal query* with respect to the examples and clues presented by the user, then we may characterize the *optimal query* as the one that will retrieve the largest number of relevant database objects within the smallest possible region in the (logical) information space.

extensions Given the stratification, that is levels or layers, of information systems discussed above, we can think of improvements or extensions on each level. At the physical layer, for example networked multimedia, in a client/server architecture, see 6.3. At the logical layer, as an information hyper space, consisting of chunks and hyperlinks, as explained in section 2.2. And at the presentation layer a virtual reality interface, representing for example the physical location of student records, somewhere at a virtual campus [x], as further explored in chapter 8. Each of these improvements or extensions can be regarded as a technological or scientific adventure in it's own right.



example(s) – *e-flux*

Do you recognize this?

When we visit a contemporary art exhibition, we find ourselves before the works, which are often quite difficult to interpret, and we observe them without understanding the process that generated them. Between a chopped-up cow immersed in formaldehyde and a replica of the Pope blindsided by a meteorite, it's legitimate to ask questions.

To provide a counter-force the exhibiton Project Room¹⁴ *challenges the usual exhibition routine and decides to not exhibit executed art works but rather offers ten self-interviewing videos by as many artists, who speak openly about a piece they are working on, or a visionary project they want to realize, or about their creative process.*

In other words, this is about works of art with no physical manifestation. It is an interesting issue whether this would still count as a *work of art*. And for multimedia, is there multimedia without a physical manifestation, with sensorily impressing the user/client. Do you remember the children story, the *New Clothes of the Emperor*?

research directions – *universal interchange*

Technology changes rapidly. Just think about the development of the PC in the last two decades of the previous century. And applications change rapidly too. At the time of writing the web does barely exist for ten years. Information spaces, on the other hand, from a sufficiently abstract perspective at least, should be rather stable over time. So the question is, *how can we encode information content in an application-independent way?* As a remark, application-independence implies technology-independence. The answer is, simply, XML. The next question then should be, what is XML and why is it more suitable for encoding information than any of the other formats, such as for example relational tables.

The first question is not so difficult. There are many sources from where an answer may be obtained. Perhaps too many. A good place to start is the XML FAQ (Frequently Asked Questions) at the Web Consortium site:

www.w3.org/XML/1999/XML-in-10-points

XML is a set of rules (you may also think of them as guidelines or conventions) for designing text formats that let you structure your data.

More specifically, XML may be characterized as follows:

XML in 10 points

1. XML is for structuring data
2. XML looks a bit like HTML
3. XML is text, but isn't meant to be read
4. XML is verbose by design

¹⁴www.e-flux.com

5. XML is a family of technologies
6. XML is new, but not that new
7. XML leads HTML to XHTML
8. XML is the basis for RDF and the Semantic Web
9. XML is license-free, platform-independent and well-supported

Perhaps not all of these points make sense to you at this stage. So let me first indicate that XML has in fact quite a long history. XML is the successor of SGML (the Structured Generalized Markup Language) that was developed in the 1980s to encode documents (such as airplane manuals) in an application-independent manner. SGML is not a language itself, but a description of how to create a content description language, using tags and attributes (as in HTML). In fact, HTML is an application of SGML, using tags with attributes both for formatting and hyperlinks. In other words, SGML is a meta language. And so is XML. Since everything got messy on the web, XML was proposed (as a subset of SGML) to make a clear distinction between content and presentation. Presentation aspects should be taken care of by stylesheets (see below) whereas the content was to be described using an XML-based language.

Now, why is XML a suitable format for encoding data? That question is a bit harder to answer. One of the reasons to use XML might be that it comes with a powerful set of related technologies (including facilities to write stylesheets):

related technologies

- Xlink – hyperlinks
- XPointer – anchors and fragments
- XSL – advanced stylesheets
- XSLT – transformation language
- DOM – object model for application programmer interface
- schemas – to specify the structure of XML documents

These technologies (that are, by the way, still in development) provide the support needed by applications to do something useful with the XML-encoded information. By itself, XML does not provide anything but a way to encode data in a meaningful manner. Meaning, however, comes by virtue of applications that make use of the (well-structured) data.

In summary, XML and its related technologies provide the means to

XML

- separate data from presentation
- transmit data between applications

Actually, the fact that XML was useful also for arbitrary data interchange became fully apparent when XML was available. To get an impression of what XML is used for nowadays, look at www.xml.org.

This leaves us with the question of why XML is to be preferred over other candidate technologies, such as relational databases and SQL. According to Kay (2001), the answer to that question is simply that XML provides a richer data

structure to encode information. In the multimedia domain we see that XML is widely adopted as an encoding format, see section ?? . For an example you might want to have a look at MusicXML, an interchange format for notation, analysis, retrieval, and performance applications, that is able to deal with common Western musical notation as used from the 17th century onwards. In appendix ?? we will explore how XML might be useful for your own multimedia application by treating some simple examples.



4

2.2 hypermedia

Given an information space we may turn it into an information hyperspace, that is, following Chang and Costabile (1997),

information hyperspace

the logical information space may further be structured in a *logical information hyperspace*, where the clues become hyperlinks that provide directional information, and the information space can be navigated by the user following directional clues.

In other words, information is chunked, and each chunk is illustrated or made accessible by an example (hypernode).

Now, what exactly does *information hyperspace* mean? To answer this question, let's briefly look at the history of hypertext and hypermedia.

history

- 1945 – Vannevar Bush (Memex) – as we may think, Bush (1995)
- 1963 – Douglas Engelbart (Augment) – boosting the human intellect Engelbart (1963)
- 1980 – Ted Nelson (Xanadu) – everything is intertwined, Nelson (1980)

Vannevar Bush' seminal paper *As we may think* may be regarded as the origin of what is known as *hypertext* with which, even if you don't know the phrase, every one of you is familiar, since it is (albeit in a rather simple way) realized in the web.

The phrase *hypertext* was invented by Ted Nelson (not patented, as far as I know), who looked for a less constraining way to organize information than was common in the educational system he grew up with. But before that, Douglas Engelbarth, who incidentally invented the mouse, developed the Augment system to, as he said, *boost the human intellect*. What for, you may ask. Let me quote the series of flashes that Engelbarth went through, according to *Dust or Magic* Hughes (2000):

- *flash 1*: we are in trouble (human mankind)
- *flash 2*: we need to boost mankind's ability to deal with complex urgent problems
- *flash 3*: aha, graphic vision surges forth of me ...
- *flash 4*: hypermedia – to augment the human intellect
- *flash 5*: augment (multimedia) workstation – portal into an information space

classification of hypermedia

Perhaps it is good to know that Vannevar Bush wrote his article when working for an information agency in the second world war period. From that perspective, we can easily see that hypermedia (combining hypertext and multimedia) were thought of as instruments of intelligence.

Basically, hypermedia systems must be able to deal with:

hypermedia systems

- components – *text, graphics, audio, video*
- links – *relations between components*
- presentation – *structured display*

Far from being a definition, this characterization gives some insight in what functionality hypermedia systems must support. Recall that dealing with complex information is what hypermedia is all about.

Is this a natural way to deal with information? Just think about how you are taught to deal with information and how you actually go about with it. Speaking about Ted Nelson, Hughes (2000) observed that *he realized that this intertwingularity was totally at odds with the education system he spent so long in and had been so uncomfortable with*. Quoting Ted Nelson himself from his book *Literary Machines*:

A curriculum promotes a false simplification of any subject, cutting the subject's many interconnections and leaving a skeleton of sequence which is only a caricature of its richness and intrinsic fascination.

Judge for yourself. Would you prefer to have an 'immersive' course in multimedia rather than a more or less ordered collection of abstractions?

True enough, the visions of the pioneers of hypermedia were overwhelming. Nevertheless, the concept of hypermedia, that is non-linear media with machine-supported links, or '*text*' as a *network*, found an application in a large variety of systems, see McKnight et al. (1991).

classification of hypermedia systems

- macro-literary systems – *publishing, reading, criticism*
- problem exploration tools – *authoring, outlining, programming*
- browsing systems – *teaching, references, information*
- general hypermedia technology – *authoring, browsing, collaboration*
- embedded hypermedia – *CASE, decision support, catalogs*

An example of a hypermedia system that has extensively been used in education, for example biology and chemistry classes, is the Brown University Intermedia system of which supports so-called *information webs*, consisting of *documents* and *links*, that could both be retrieved by specifying attribute, allowing in this way for respectively both filtered content and conditional navigation. An interesting aspect of this system is that the user may create *maps*, that is structures containing documents and links, which form a personalized version of the web of information for a specific user, superimposed on the information space offered by the system.



Dexter Hypertext Reference Model

After many years of developing ideas and exploring implementations, one group of experts in the field came together and developed what is commonly known as the *Dexter Hypertext Reference Model*, named after the location, actually a pub, where the meetings were held. The Dexter model offers an abstract description of *hypertext*. It made a distinction between *components*, *anchors* within components

and *links* between components, attached to anchors. The model was meant as a reference standard against which existing and future hypertext systems could be compared.

Components have the following attributes:

component

- content – *text, graphics, video, program*
- attributes – *semantic description*
- anchors – *(bi-directional) links to other documents*
- presentation – *display characteristics*

The Dexter Hypertext Model has been criticised from the beginning. Among others, because *compound documents*, that is documents having subcomponents, where not adequately dealt with. And also because it did not accomodate multimedia (such as video) content very well. In practice, however, the Dexter model has proven to be even somewhat overambitious in some respects. For example, the web does (currently) not support bi-directional links in a straightforward manner.

Amsterdam Hypermedia Model

When looking for alternatives, a Dutch multimedia research group at CWI proposed to extend the Dexter model with their own multimedia model (CMIF), an extension for which they coined the name *Amsterdam Hypermedia Model*.

Let's look at the (CMIF) multimedia model first:

(CMIF) multimedia model

- data block – *atomic component*
- channel – *abstract output device*
- synchronization arc – *specifying timing constraints*
- event – *actual presentation*

What strikes as an immediate difference with respect to the hypertext model is the availability of *channels*, that allow for presenting information simultaneously, and so-called *synchronization arcs*, that allow the author to specify timing constraints. Also, events are introduced in the model to deal with user interactions.

With respect to authoring, the model supports a declarative approach to specifying sequential and parallel compounds, that is in what order specific things must be presented and what may occur simultaneously. Again, channels may be employed to offer a choice in the presentation, for example a dutch or english account of a trip in Amsterdam, dependent on the preferences of the (human) viewer.

The Amsterdam Hypermedia Model (AHM) extends the Dexter Hypertext Reference Model in a rather straightforward way with channels and synchronization arcs.

Amsterdam Hypermedia Model

- contents – *data block*
- attributes – *semantic information*

- anchors – (*id*, *value*)
- presentation – *channel*, *duration*, ...

Obviously, the difference between Dexter and AHM is primarily the more precise definition of *presentation characteristics*, by introducing *channels* as in the (CMIF) multimedia model. Another (major) difference lies in the characterization of compounds. Each compound has one or more children, or subcomponents. Subcomponents may act as the source or destination of synchronization arcs. Each component obtains a start-time, that may result from parallel or sequential composition and synchronisation arcs.

Another interesting concept introduced by the Amsterdam Hypermedia Model is the notion of *context*. What happens when you click on a link? Does everything change or are only some parts affected? Then, when you return, does your video fragment start anew or does it take up where you left it? Such and other issues are clarified in the Amsterdam Hypermedia Model, of which we have omitted many details here.

It is perhaps interesting to know that the Amsterdam Hypermedia Model has served as a reference for the SMIL standard discussed in section 3.2. If you want to know more about the Amsterdam Hypermedia Model, you may consult Ossenbruggen (2001) or Hardman et al. (1994).



example(s) – *hush*

In the *hush*¹⁵ we explore a variety of hypermedia applications. In fact already in 1994 we developed a SGML-based browser with *applets* in Tcl/Tk, Applications and SGMLWEB. Somehow, we did a lot with music with optimistic titles such as *Bringing music to the We*, Ossenbruggen & Eliens (1994) and more pessimistic ones such as *Jamming (on) the Web*, Eliens et al. (1997). The acronym *hush* stands for *hyper utility shell*. Many of the projects with *hush* were student projects, in which we studied operational support for hypermedia applications. Although we used SGML for markup, we did not have any specific document model, as in CMIF. An overview and rationale of *hush* is given in Eliens (2000). A significant part of the *hush* software is being reused in the ViP system, that is discussed in section 4.3, albeit with an entirely different presentation technology.

¹⁵www.cs.vu.nl/~eliens/online/hush

research directions– *computational models*

Today, hypermedia functionality is to some extent embedded in almost all applications. However, to realize the full potential of hypermedia, and in effect the networked multimedia computer, there are still many (research) issues to be resolved. To get an impression of the issues involved, have a look at the famous seven hypermedia research issues formulated by Halasz.

research issues

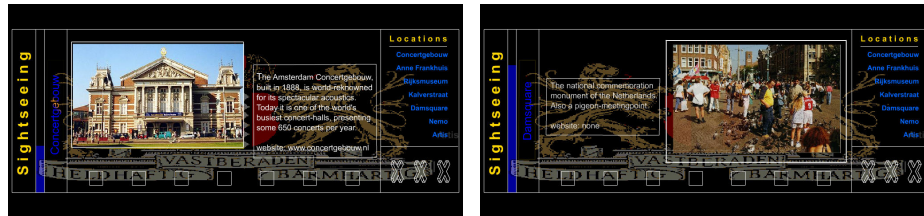
- *search and query* – for better access
- *composition* – for imposing structure
- *virtual structures* – on top of existing structures
- *computation* – for flexibility and interaction
- *versioning* – to store modification histories
- *collaborative work* – sharing objects with multiple users
- *extensibility and tailorability* – to adapt to individual preferences

See Ossensbruggen (2001), section 2.3 for a more extensive description. Although the research issues listed above were formulated quite early in the history of hypermedia, as a reflection on the requirements for second-generation hypermedia systems, they remain valid even today. Without going into any detail with respect to the individual research issues, I rather wish to pose the grand encompassing research issue for the networked multimedia computer: *What is the proper computational model underlying hypermedia or, more generally, for applications that exploit the networked multimedia computer in its full potential?* Some directions that are relevant to this issue will be given in section which deals with the multimedia semantic web.

2.3 multimedia authoring

It is tempting to identify a presentation with the information space it presents. This is what users often do, and perhaps should do. When that happens, the presentation is effective. But you must remember that the actual presentation is just one of the many possible ways to engage a user in exploring an information space. Making the choice of what to present to the user is what we understand by *(multimedia) authoring*.

Authoring is what we will discuss in this section. Not by giving detailed guidelines on how to produce a presentation (although you may look at the online assignment for some hints in this respect), but rather by collecting wisdom from a variety of sources.



7

visualization

Let's start with our explorations by looking at the problem of *visualisation* with a quote from David Gelernter, taken from Shneiderman (1997):

visualization

Grasping the whole is a gigantic theme, intellectual history's most important. Ant vision is humanity's usual fate; but seeing the whole is every thinking person's aspiration.

David Gelernter, Mirror Worlds 1992

Now, consider, there are many ways in which the underlying information space may be structured, or speaking as a computer scientist, what data types may be used to represent the (abstract) information.

data types

- *1-D linear data* – text, source code, word index
- *2-D map data* – floor plan, office layout
- *3-D world* – molecules, schematics, ...
- *temporal data* – 1 D (start, finish)
- *multi-dimensional data* – n-dimensional (information) space
- *tree data* – hierarchical
- *network data* – graph structure

The *visualisation problem* then is to find a suitable way to present these structures to the user. Basicall, following Shneiderman (1997), there are two paradigms to present this information:

- *interactive* – overview first, zoom and filter, then details on demand
- *storytelling* – as a paradigm for information presentation

Storytelling may be very compelling, and does not force the user to interact. On the other hand, storytelling may lead to information consumerism alike to television enslavement.

An interaction paradigm that combines 'storytelling' with opportunities for interaction, as for example in the *blendo* approach discussed in section 3.2, would seem to be most favorable. Interaction then may result in either changing the direction of the story, or in the display of additional information or even transactions with a third party (for example to buy some goodies).

persuasive technology

Whatever your target audience, whatever your medium, whatever your message, you have to be convincing if not compelling.

In the tradition of *rethorics*, which is the ancient craft of convincing others, a new line of research has arisen under the name of *persuasive technology*. In the words of my colleague, Claire Dormann, persuasion is:

persuasion

- a communication process in which the communicator seeks to elicit a desired response from his receiver
- a conscious attempt by one individual to change the attitudes, beliefs or behaviours of another individual or group individual through the transmission of some messages.

In other words, *the purpose of persuasion is to accomplish one of the following goals: to induce the audience to take some action, to educate the audience (persuade them to accept to accept information or data), or to provide the audience with an experience.* In the area of multimedia, one may think of many applications. Quoting Claire Dormann, *in interactive media, the field of application of persuasive technology ranges from E-commerce, social marketing (like an anti-AIDS campaign) to museum exhibits. Also E-commerce provides an obvious example. To convince people to buy more, more persuasive messages and technologies are developed through the use of humorous and emotional communication, agents (such as price finders) or 3D representations of products and shops. For health campaigns (or any campaign of your choice) one can imagine 3D information spaces with agents presenting different point of views and where users are given different roles to play. In a museum you might want to highlight key points through innovative and fun interactive exhibits.* Although the subject of *persuasive technology* is far less technology-oriented than the name suggests, multimedia (in a broad sense) form an excellent platform to explore *persuasion*. As concerns multimedia authoring, set yourself a goal, do the assignment, explore your capabilities, convey that message, and make the best of it.

(re)mediation

What can you hope to achieve when working with the new media? Think about it. Are the new media really new? Does anyone want to produce something that nobody has ever seen or heard before? Probably not. But it takes some philosophy to get that sufficiently clear.

In Bolter and Grusin (2000), the new media are analyzed from the perspective of remediation, that is the mutual influence of media on each other in a historical perspective. In any medium, according to Bolter and Grusin (2000), there are two forces at work:

(re)mediation

- *immediacy* – a tendency towards transparent immersion, and
- *hypermediacy* – the presence of referential context

Put in other words, immediacy occurs when the medium itself is forgotten, so to speak, as is (ideally) the case in realistic painting, dramatic movies, and (perhaps in its most extreme form) in virtual reality. Hypermediacy may be observed when either the medium itself becomes the subject of our attention as in some genres of modern painting, experimental literature and film making, or when there is an explicit reference to other related sources of information or areas of experience, as in conceptual art, many web sites, and also in CNN news, where apart from live reports of ongoing action, running banners with a variety of information keep the viewers up to date of other news facts.

Now, the notion of *remediation* comes into play when we observe that every medium draws on the history of other media, or even its own history, to achieve a proper level of immediacy, or 'natural immersion'. For example, Hollywood movies are only realistic to the extent that we understand the dramatic intent of cuts, close-ups and storylines, as they have been developed by the industry during the development of the medium. As another example, the realism of virtual reality can only be understood when we appreciate linear perspective (which arose out of realistic Renaissance painting) and dynamic scenes from a first person perspective (for which we have been prepared by action movies and TV).

Even if you may argue about the examples, let it be clear that each (new) medium refers, at least implicitly, to another medium, or to itself in a previous historic phase. So, what does this mean for new media, like TV or virtual reality?

Let's start with virtual reality. Bolter and Grusin (2000) comment on a statement of Arthur C. Clarke

Virtual Reality won't merely replace TV. It will eat it alive.

by saying that ... *he is right in the sense that virtual reality remediates television (and film) by the strategy of incorporation. This strategy does not mean that virtual reality can obliterate the earlier visual point-of-view technologies, rather it ensures that these technologies remain as least as reference points by which the immediacy of virtual reality is measured.*

So, they observe "paradoxically, then, remediation is as important for the logic of transparency as it is for hypermediacy". Following Bolter and Grusin (2000), we can characterize the notions of immediacy and hypermediacy somewhat more precisely.

immediacy

- epistemological: transparency, the absence of mediation
- psychological: the medium has disappeared, presence, immersion

hypermediacy

- epistemological: opacity, presence of the medium and mediation
- psychological: experience of the medium is an experience of the real

Now, sharpen your philosophical teeth at the following statement, taken from Bolter and Grusin (2000), p. 224:

Convergence is the mutual remediation of at least three important technologies – telephone, television and computer – each of which is a hybrid of technical, social and economic practice, and each of which offers its own path to immediacy.

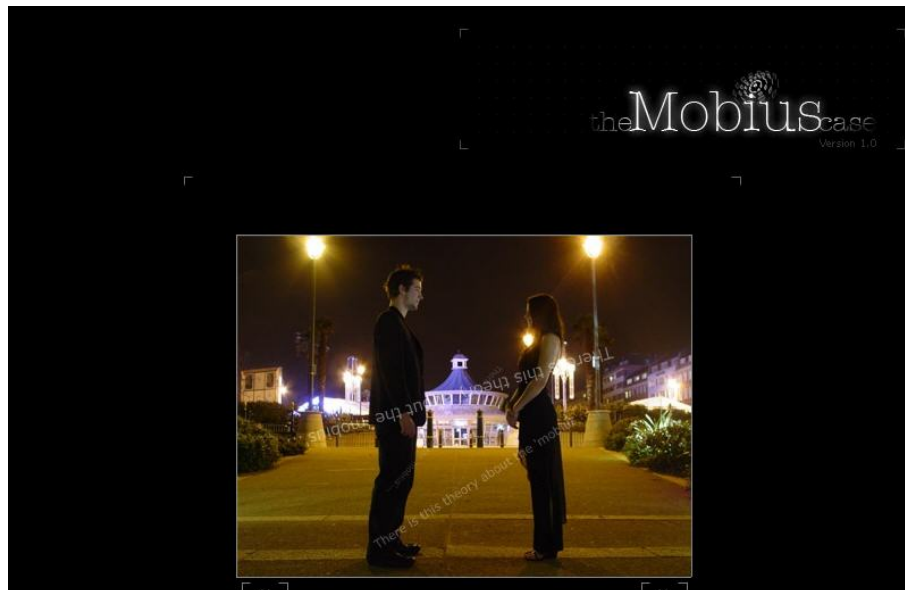
The telephone offers the immediacy of voice or the interchange of voices in real-time.

Television is a point-of-view technology that promises immediacy through its insistent real-time monitoring of the world.

The computer's promise of immediacy comes through the combination of three-dimensional graphics, automatic (programmed) action, and an interactivity that television can not match.

As they come together, each of these is trying to absorb the others and promote its own version of immediacy.

Let us once more come back to virtual reality and its possible relevance in our information age, Bolter and Grusin (2000), p. 225:: *in the claim that new media should not be merely archival but immersive, the rhetoric of virtual reality finally enters in, with its promise of the immediacy of experience through transparency.* . So, with respect to the new media, we may indeed conclude: *what is in fact new is the particular way in which each innovation rearranges and reconstitutes the meaning of earlier elements. What is new about media is therefore also old and familiar: that they promise the new by remediating what has gone before. The true novelty would be a new medium that did not refer to the other media at all. For our culture, such mediation without remediation seems to be impossible.*



example(s) – *mobius*

Rurger van Dijk, a former student of mine, has implemented an interactive story in *flash*. The story is a romance, told with images displaying scenes from the life of the players, a young man and a young women. The user can choose perspectives, either the man's or woman's, and watch the story from that point of view. The story is both non-linear and circular. The scenes can be connected in various way, and order is not compulsory.

research directions– *narrative structure*

Where do we go from here? What is the multimedia computer, if not a new medium? To close this section on multimedia authoring, let us reconsider in what way the networked multimedia computer differs from other media, by taking up the theme of convergence again. The networked multimedia computer seems to remediate all other media. Or, in the words of Murray (1997):

convergence

... merging previously disparate technologies of communication and representation into a single medium.

The networked computer acts like a telephone in offering one-to-one real-time communication, like a television in broadcasting moving pictures, like an auditorium in bringing groups together for lectures and discussion, like a library in offering vast amounts of textual information for reference, like a museum in its ordered presentation of visual information, like a billboard, a radio, a gameboard and even like a manuscript in its revival of scrolling text.

In Murray (1997), an analysis is given of a great variety of computer entertainment applications, varying from shoot-em-up games to collaborative interactive role playing. Murray (1997) identifies four essential properties that make these applications stand out against the entertainment offered by other media, which include books and TV. Two key properties determine the interactive nature of computer entertainment applications:

interactive

- *procedural* – 'programmed media' ...
- *participatory* – offering agency

All applications examined in Murray (1997) may be regarded as 'programmed media', for which interactivity is determined by 'procedural rules'. With *agency* is meant that the user can make active choices and thus influence the course of affairs, or at least determine the sequence in which the material is experienced.



9

Another common characteristic of the applications examined is what Murray (1997) calls *immersiveness*. Immersiveness is determined by two other key properties:

immersive

- *spatial* – explorable in (state) space
- *encyclopedic* – with (partial) information closure

All applications are based on some spatial metaphor. Actually, many games operate in 'levels' that can be accessed only after demonstrating a certain degree of mastery. Networked computer applications allow for incorporating an almost unlimited amount of information. Some of the information might be open-ended, with storylines that remain unfinished. Closure, then, is achieved simply by exhaustive exploration or diminishing attention.

multimedia authoring Coming back to the question what the 'new medium', that is the networked multimedia computer, has to offer from the perspective of multimedia authoring, two aspects come to the foreground:

multimedia authoring

- narrative format
- procedural authorship

The narrative format is incredibly rich, offering all possibilities of the multimedia computer, including 3D graphics, real-time sound, text. In short, everything up to virtual reality. But perhaps the most distinguishing feature of the new medium is that true authorship requires both artistic capabilities as well as an awareness of the computational power of the medium. That is to say, authorship also means to formulate generic computational rules for telling a story while allowing for interactive interventions by the user. Or, as phrased in Murray (1997), the new *cyberbard* must create prototypical stories and formulaic characters that, in some way, lead their own life and tell their stories following their innate (read: programmed) rules. In section ?? and appendix ??, we will present a framework that may be used as a testbed for developing programmed narrative structures with embodied agents as the main characters.

2.4 development(s) – semantic mashups

mashup(s)

- substituting a single pragmatism for ideal design
- light weight programming models

web 2.0 design pattern(s)

- web 1.0 – the web as platform
- web 2.0 – architecture of participation
- web 3.0 – data is the (intel) inside

blogosphere / perma link / track back / page ran

daft punk – technologic (cn / jp)

Buy it, use it, break it, fix it.
 Trash it, change it, melt – upgrade it.
 Change it, point it, zoom it, press it.
 Snap it, work it, quick – erase it.
 Write it, out it, paste it, save it.
 Load it, check it, quick – rewrite it.
 Plug it, play it, burn it, rip it.
 Drag and drop it, zip – unzip it.
 Look it, fill it, curl it, find it.
 View it, coat it, jam – unlock it.
 Surf it, scroll it, pose it, click it.
 Cross it, crack it, twitch – update it.
 Name it, rate it, tune it, print it.
 Scan it, send it, fax – rename it.
 Touch it, bring it. Pay it, watch it.
 Turn it, leave it, stop – format it.

web services

By now the phrase *Web 2.0* as well as applications representing it, such as Flickr and YouTube, are well established, and enjoyed by a wide community. Each day new items are added to the growing list of mashups¹⁶, and the number of web services that constitute the building blocks of mashups also shows a steady growth. Mashups seem to be the easy way to start up a company, since the technology is relatively easy and, making use of appropriate services, initial investment costs can be low. Cf. Amazon.

What Web 2.0 stands for, from a technical perspective, is succinctly expressed in Dorai's Learnlog¹⁷ *XML Is The Fabric Of Web 2.0 Applications*:

- The client side is AJAX (Asynchronous Javascript and XML)
- The server application typically exposes data through XML

¹⁶www.programmableweb.com/mashuplist/

¹⁷dorai.wordpress.com/tag/mashups/

- The interaction model is web services
- Mashups combine multiple webservices to create new types of applications

And eventhough many alternative representations, such as JSON¹⁸ (Javascript Object Notation) are increasingly being used, all in all XML may be regarded as the *interlingua* of the Web 2.0.

Before taking a closer look at the communication protocol(s) underlying Web 2.0 and de-construct the tight link of AJAX to HTML in-page formatting, it is worthwhile, following Amazon, to give an overview of a selected number of services, that may be used to create mashups:

- google – code.google.com/
- yahoo – developer.yahoo.com/
- del.icio.us – del.icio.us/help/api/
- flickr – www.flickr.com/services/
- bbc – www0.rdtthdo.bbc.co.uk/services/
- youtube – www.youtube.com/dev

Although mashups featuring google maps seem to be the dominant mashup type, other services such as offered by del.ici.us, Flickr and BBC might prove to be more worthwhile for 'serious' applications. For example, for developing e-commerce applications Amazon¹⁹ offers services for *product operations*, such as item search and similarity lookup, *remote shopping carts*, to create and manage purchase collections, *customer content*, to access information contributed by customers, and *third party listings*, to find related resellers. It is important to note that many of these services, as for example the *shoppong cart* services, may be used independently of the commercial offerings of Amazon!

Most of the service providers and services mentioned above are accessible using a choice of protocols, including WSDL, SOAP, XML-RPC and the REST protocol. The REST protocol seems to be most widespread and as we will discuss in the next section, it seems to be the most appropriate protocol in Second Life.

REST stands for *Representational State Transfer*. In essence, the REST protocol uses the url as a command-line for stateless RPC invocations, which allows for services to be executed by typing in the address box of a web browser. A great tutorial about the REST protocol can be found in Joe Gregorio's column²⁰: *The Restful Web*. As fully explained in Visser and Eliens (2000), the phrases *representation*, *state* and *transfer*, respectively, stand for:

- representation – encoding in a particular format
- state – data encapsulated in an object
- transfer – using HTTP methods

In practice, the use of REST means that the state associated with a resource or service must be managed by the client. Together with mechanisms such as

¹⁸www.json.org/

¹⁹aws.amazon.com

²⁰www.xml.com/pub/a/2004/12/01/restful-web.html

content-negotiation and URL-rewriting, REST provides a simple, yet powerful method to invoke services using HTTP requests.

A common misunderstanding is that AJAX is intimately tied to web browsers and in-page HTML formatting. This misunderstanding is due to the fact that AJAX is often used to improve the user experience of web pages by emulating RIA (Rich Internet Applications) using DHTML and CSS. However, the real meaning of AJAX in our view is that AJAX allows for asynchronous client-controlled server requests, that are executed without an immediate visible effect for the user. We explored the use of AJAX and web services in an X3D/VRML implementation of PANORAMA, a system meant to support social awareness in a work environment, Panorama, Munro et al. (1999). PANORAMA represents *casual encounters* in the work environment and displays *self reflections*, that is postcards and other items contributed by employees, at a large display in a central public space. In section 6 we will discuss how to realize PANORAMA in Second Life, from both technical and aesthetic perspectives. Fig. 2, taken from PanoramaWeb, illustrates the architecture of an Ajax-based web implementation of PANORAMA, which includes facilities for game playing as *occasional battle(s)*, using the google GWT toolkit to allow users to contribute their image material, and a PHP server for storage and retrieval of images.

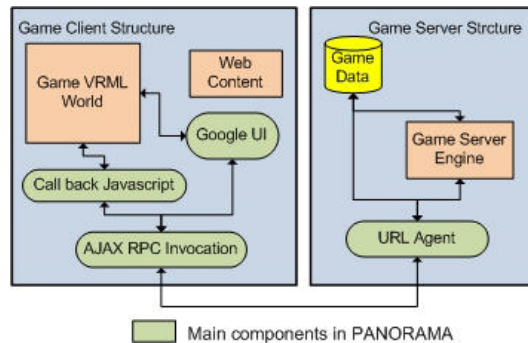


Fig. 2: Architecture PANORAMA (Web)

The Web 2.0 offers a lively arena for consumers and developers alike, with a multitude of blogs discussing the future of the web. For example, in Dion Hinchcliffe rebuttal²¹ of Jeffrey Zeldman's *Web 3.0 Û Web 1.0 = Web 2.0* blog, entitled *Is Web 2.0 Entering "The Trough of Disillusionment"?* it is suggested that *our services could even be more powerful by creating semantic mashups*²². Although the notion of *semantic web technology* is widely known and accepted, we include for reference a characterization of Nova Spivack quoted from Dan Farber and Larry Dignan's blog²³ *Web 2.0 isn't dead, but Web 3.0 is bubbling up*:

The Semantic Web is a set of technologies which are designed to enable a particular vision for the future of the Web Û a future in which all knowledge

²¹web2.sys-con.com/read/172417.htm

²²www.web2journal.com/read/361294.htm

²³blogs.zdnet.com/BTL/?p=3934

exists on the Web in a format that software applications can understand and reason about. By making knowledge more accessible to software, software will essentially become able to understand knowledge, think about knowledge, and create new knowledge. In other words, software will be able to be more intelligent Û not as intelligent as humans perhaps, but more intelligent than say, your word processor is today.

But even in the semantic web community the discussion whether to go for *folksonomies* or *formal ontologies* rages, Folk, and it is not clear at this stage what will prove to be more powerful, HTML-scraping, tags, microformats, or full ontologies. Anyway, from our perspective, incorporating web services in Second Life, we are not in a position to enter this discussion, but simply explore what is useful in realizing meaningful mashups in 3D virtual environments. To conclude this brief overview of web services and mashups we wish to give another quote from Dorai's Learnlog, this time from Jon Udell, in his blog on his move to Microsoft:

The most powerful mashups don't just mix code and data, they mix cultures.

which provides a challenge that transcends all issues of mere technological correctness.

technological infrastructure

Second Life offers an advanced scripting language with a C-like syntax and an extensive library of built-in functionality. Although it has support for objects, LSL (the Linden Scripting Language) is not object-oriented. Cf. Eliens (2000). Scripts in Second Life are server-based, that is all scripts are executed at the server, to allow sharing between visitors. Characteristic for LSL are the notions of *state* and *eventhandler*, which react to events in the environments.

Among the built-in functions there are functions to connect to a (web) server, and obtain a response, in particular (with reference to their wiki page):

- request – wiki.secondlife.com/wiki/LIHTTPRequest
- escape – wiki.secondlife.com/wiki/LIEscapeURL
- response – wiki.secondlife.com/wiki/Http_response

Other functions to connect to the world include *sensors*, for example to detect the presence of (visitors') avatars, and chat and instant messaging functions to communicate with other avatars using scripts. In addition, LSL offers functions to control the behavior and appearance of objects, including functions to make objects react to physical laws, to apply force to objects, to activate objects attached to an avatar (as for example the phantom Mario sprites mentioned earlier), and functions to animate textures, that can be used to present slide shows in Second Life.

On the Mashable²⁴ *Social Networking News* site a brief overview is given of the use of web services in Second Life, entitled *Second Life + Web 2.0 = Virtual World Mashups*. To access Second Life from outside-in (that is from a web browser),

²⁴mashable.com/2006/05/30/second-life-web-20-virtual-world-mashups/

so-called *slurls* may be used, for example to reach VU²⁵ @ Second Life, and all slurls listed in del.icio.us under *slurlmarker*²⁶ may be used, also to activate in-world teleporting using scraping techniques.

As remarked in the *hackdiary*²⁷ by Matt Biddulph, Second Life (currently) lacks the ability to parse XML or JSON, so the best way to incorporate web services is to set up a web server with adequate resources. As Matt Biddulph indicates, to access *flickr* photographs for a particular user (avatar), a web server may contain the following resources:

- /seen?user=SomeAvatar – records the presence of SomeAvatar
- /touched?user=SomeAvatar – invokes flickr API with users tag
- /set_tag?user=SomeAvatar&tag=FavoriteTag – records SomeAvatar’s favourite tag

For example, in response to a ‘touch’ event, invoking *touch* results in consulting the database for the user’s tag and asking the Flickr API for a random photo with that tag. It then returns a string containing the url for a particular photograph. LSL functions used in this application include *sensors*, to check for presence, *listen* functions, to respond to spoken commands, and *touch* events, for the physical interface. In addition to supporting strings and lists, LSL provides a perl-like split function to convert a string into a list of strings, thus allowing for processing multiple items in response to a server request.

Another example of using web services in Second Life is writing blogs²⁸ from within Second Life using the BlogHUD²⁹ developed by Koz Farina who also is reported to have found a flash hack that allows for reading RSS feeds. As explained by Koz Farina

Quicktime supports Flash, but only up to Flash version 5. We’re up to version 9 on that now! Luckily, I have been dabbling with Flash since the early days, so already knew how to do this ‘the old way’... So, Flash is doing all the work. No LSL at all... I heart feeds. Did I say ‘I heart feeds?’

The RSS display uses the ability to stream Quicktime video in Second Life, and again the mashup is not created in Second Life but by appropriate server support.

In a similar vein we may incorporate live streaming video³⁰, for example by using WireCast³¹ to capture and organize live camera input, possibly together the screen output of other applications such as *powerpoint*, which must then be sent to a streaming server supporting Quicktime, such as Apple’s Darwin³², which may then be accessed from Second Life to texture a display object.

Finally, as another *Web 2.0 to Web 3D* phenomenon, announced in New World Notes³³, we may mention the use of Twitter³⁴ messages, that allow residents to

²⁵slurl.com/secondlife/VU%20University%20NL/29/151

²⁶del.icio.us/tag/slurlmarker

²⁷www.hackdiary.com/archives/000085.html

²⁸nwn.blogs.com/nwn/2006/10/really_simple_s.html

²⁹bloghud.com/

³⁰blogs.electricsheepcompany.com/chris/?p=206

³¹www.varasoftware.com/products/wirecast/

³²developer.apple.com/opensource/server/streaming/

³³nwn.blogs.com/nwn/2007/03/post_1.html

³⁴twitter.com/

send and receive message about ongoing activities. A similar service is reported to exist for *jaiku*³⁵ messages.

models of Behavior

tracking user interaction

Our virtual campus in Second Life already allows for performing simple statistics, by recording the presence of users at particular spots in the virtual world, using sensors and listeners installed in 3D objects. Since the LSL script-based counters appear to be rather volatile, tracking data are sent to a web server and stored in a database. This mechanism can easily be extended to a more encompassing form of user tracking, recording for a particular user not only presence at particular spots, but also the duration of presence, the actual proximity to objects, and the proximity to other users, as well as explicitly spoken comments or actions such as the donation of (Linden) money.

In Recommend, we introduced the notion of *serial recommender*, which generates guided tours in 3D digital dossier, Dossier, based on (expert) user tracking. To incrementally refine such tours for individual users, we used a behavioral model originally developed in Privacy to cope with privacy issues in distributed recommender systems. In this section we will briefly sketch this model and discuss how it applies to generating annotations for objects in Second Life, and how these annotations may be used to provide recommendations to other users.



10

questions

information spaces

1. (*) What factors play a role in the development of *multimedia information systems*? What research issues are there? When do you expect the major problems to be solved?

concepts

2. Define the notion of *information spaces*?
3. Indicate how multimedia objects may be placed (and queried for) in an *information (hyper) space*?
4. Characterize the notion of *hypermedia*.

technology

5. Discuss which developments make a large scale application of multimedia information systems possible.
6. Give a characterization of an object, a query and a clue in an *information space*.

³⁵devku.org/docs

7. Describe the *Dexter Hypertext Reference Model*.
8. Give a description of the *Amsterdam Hypermedia Model*.

projects & further reading As a project, I suggest the development of a virtual tour in a city, museum or other interesting locatoion.

You may further explore the implementation of traversal within a context, taking into account the history of navigation when backtracking to a particular point, issues in hyperlinking and interaction in multimedia applications, and computational support for narratives.

For further reading I advice you to take a look at the history of hypermedia and the web, using online material from the W3C³⁶, or the history of media as accounted for in Briggs and Burke (2001) and Bolter and Grusin (2000).

the artwork

1. book covers – Desing, Eco (1994), Avantgarde, Kunst, Betsky (2004)
2. Federico Campanale³⁷ – Oxygen, fragments from video installation, 2004
3. Vasarely – Diehl 1973.
4. Vasarely – Diehl 1973.
5. Vasarely – Diehl 1973.
6. Federico Campanale – Oxygen, more fragments.
7. student work – from *introduction multimedia* 2000.
8. Rutger van Dijk – *mobius*, interactive story, opening screen, see section 2.3.
9. edgecodes – screenshots, see section 2.3
10. signs – people, van Rooijen (2003), p. 244, 245.

The work of Vasarely has served as an example for many contemporary digital artists. It is playful, mat may be characterized also as *formalist*. The highly aesthetic video work of Federico Campanale who, as he told me was strongly influenced by vasarely in his early years, shows a similar combination of formalism and playfulness. The interactive story by Rutger van Dijk has a rather different atmosphere, it is highly romantic, with slick graphics. The musea sites are included to point to the existence of (an increasing number) of virtual tours.

³⁶www.w3c.org

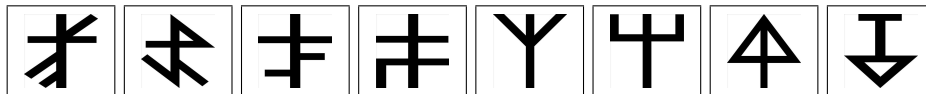
³⁷www.blue-frame.com

part ii. delivery & presentation

if you linger for a long time in one place you'd almost think there must be something there
wittgenstein

chapters:

- 3. codecs and standards
- 4. multimedia platforms



2

reading directives In this part we will look at the issues involved in delivery and presentation, primarily from a technical perspective. We will argue the importance of codecs (read compression), and we will discuss the criteria for selecting a particular codec, as well as the standards that have been developed for packaging multimedia content in an effective way. In chapter 4, we will discuss multimedia presentation platforms, and we will look at the Microsoft DirectX 9 platform in somewhat greater detail.

Essential sections are section 3.1, which introduces codecs, 3.2, which discusses the MPEG-4 and SMIL standards and section 4.1, which puts the development of high-end multimedia platforms in a historical perspective. Sections 3.3 and 4.3 can safely be skipped on first reading.

perspectives As you can see below, the topics introduced in this part are not only relevant from a technical perspective. Other perspectives are equally valid:

perspectives – delivery & presentation

- technical – codec selection
- political – market vs. consortium
- sociological – digital services
- legal – copyright protection
- scientific – experience design
- computer science – computational support
- futuristic – global & personal information
- commercial – WMV, Quicktime, RealONE

For example, the issues of copyrights and copyright protection are hot topics, since the rise of the internet is obviously a threat to the traditional industries of music and film distribution.

essay topics Since many of the interesting topics will only be hinted, you may select on or more topics for further investigation and study. As essay titles I would suggest:

- multimedia standards – MPEG4
- XML-based multimedia – SMIL
- multimedia technology – the DirectX 9 toolbox

When you write the essay, then assess first from which perspective you will tackle the subject. When you approach the material from a technical perspective, then make sure that you do understand the technical issues in sufficient detail.



3

the artwork

1. logo – a drawing by Soutine, it is (almost) my personal logo, and also decorates the cover of *Eliens* (2000).
2. signs – property marks, van Rooijen (2003), p. 76, 77.
3. photographs – Jaap Stahlie³⁸, commissioned work.

³⁸www.jaapstahlie.com

3. codecs and standards

without compression delivery is virtually impossible

learning objectives

After reading this chapter you should be able to demonstrate the necessity of compression, to discuss criteria for the selection of codecs and mention some of the alternatives, to characterize the MPEG-4 and SMIL standards, to explain the difference between MPEG-4 and MPEG-2, and to speculate about the feasibility of a semantic multimedia web.

Without compression and decompression, digital information delivery would be virtually impossible. In this chapter we will take a more detailed look at compression and decompression. It contains the information that you may possibly need to decide on a suitable compression and decompression scheme (codec) for your future multimedia productions. We will also discuss the standards that may govern the future (multimedia) Web, including MPEG-4, SMIL and RM3D. We will explore to what extent these standards allow us to realize the optimal multimedia platform, that is one that embodies digital convergence in its full potential. Finally, we will investigate how these ideas may ultimately lead to a (multimedia) semantic web.



1

3.1 codecs

Back to the everyday reality of the technology that surrounds us. What can we expect to become of networked multimedia? Let one thing be clear

compression is the key to effective delivery

There can be no misunderstanding about this, although you may wonder why you need to bother with compression (and decompression). The answer is simple. You need to be aware of the size of what you put on the web and the demands that imposes on the network. Consider the table, taken from Vasudev and Li (1997), below.

<i>media</i>	uncompressed	compressed
voice 8k samples/sec, 8 bits/sample	64 kbps	2-4 kbps
slow motion video 10fps 176x120 8 bits	5.07 Mbps	8-16 kbps
audio conference 8k samples/sec 8bits	64 kbps	16-64 kbps
video conference 15 fps 352x240 8bits	30.4 Mbps	64-768 kbps
audio (stereo) 44.1 k samples/s 16 bits	1.5 Mbps	128k-1.5Mbps
video 15 fps 352x240 15 fps 8 bits	30.4 Mbps	384 kbps
video (CDROM) 30 fps 352x240 8 bits	60.8 Mbps	1.5-4 Mbps
video (broadcast) 30 fps 720x480 8 bits	248.8 Mbps	3-8 Mbps
HDTV 59.9 fps 1280x720 8 bits	1.3 Gbps	20 Mbps

You'll see that, taking the various types of connection in mind

(phone: 56 Kb/s, ISDN: 64-128 Kb/s, cable: 0.5-1 Mb/s, DSL: 0.5-2 Mb/s)

you must be careful to select a media type that is suitable for your target audience. And then again, choosing the right compression scheme might make the difference between being able to deliver or not being able to do so. Fortunately,

images, video and audio are amenable to compression

Why this is so is explained in Vasudev and Li (1997). Compression is feasible because of, on the one hand, the statistical redundancy in the signal, and the irrelevance of particular information from a perceptual perspective on the other hand. Redundancy comes about by both spatial correlation, between neighboring pixels, and temporal correlation, between successive frames.

The actual process of encoding and decoding may be depicted as follows:

$$\text{codec} = (\text{en})\text{coder} + \text{decoder}$$

signal \rightarrow source coder \rightarrow channel coder (encoding)

signal \leftarrow source decoder \leftarrow channel decoder (decoding)

Of course, the coded signal must be transmitted accross some channel, but this is outside the scope of the coding and decoding issue. With this diagram in mind we can specify the *codec design problem*:

From a systems design viewpoint, one can restate the codec design problem as a bit rate minimization problem, meeting (among others) constraints concerning: specified levels of signal quality, implementation complexity, and communication delay (start coding – end decoding).



2

compression methods

As explained in Vasudev and Li (1997), there is a large variety of compression (and corresponding decompression) methods, including model-based methods, as for example the object-based MPEG-4 method that will be discussed later, and waveform-based methods, for which we generally make a distinction between lossless and lossy methods. Huffman coding is an example of a lossless method, and methods based on Fourier transforms are generally lossy. Lossy means that actual data is lost, so that after decompression there may be a loss of (perceptual) quality.

Leaving a more detailed description of compression methods to the diligent students' own research, it should come as no surprise that when selecting a compression method, there are a number of tradeoffs, with respect to, for example, coding efficiency, the complexity of the coder and decoder, and the signal quality. In summary, the following issues should be considered:

tradeoffs

- *resilience to transmission errors*
- *degradations in decoder output – lossless or lossy*
- *data representation – browsing & inspection*
- *data modalities – audio & video.*
- *transcoding to other formats – interoperability*
- *coding efficiency – compression ratio*
- *coder complexity – processor and memory requirements*
- *signal quality – bit error probability, signal/noise ratio*

For example, when we select a particular coder-decoder scheme we must consider whether we can guarantee resilience to transmission errors and how these will affect the users' experience. And to what extent we are willing to accept degradations in decoder output, that is lossy output. Another issue in selecting a method of compression is whether the (compressed) data representation allows for browsing & inspection. And, for particular applications, such as conferencing, we should be worried about the interplay of data modalities, in particular, audio & video. With regard to the many existing codecs and the variety of platforms we may desire the possibility of transcoding to other formats to achieve, for example, exchange of media objects between tools, as is already common for image processing tools.

compression standards

Given the importance of codecs it should come as no surprise that much effort has been put in developing standards, such as JPEG for images and MPEG for audio and video.

Most of you have heard of MP3 (the audio format), and at least some of you should be familiar with MPEG-2 video encoding (which is used for DVDs).

Now, from a somewhat more abstract perspective, we can, again following Vasudev and Li (1997), make a distinction between a *pixel-based approach* (coding the raw signal so to speak) and an *object-based approach*, that uses segmentation and a more advanced scheme of description.

- *pixel-based* – MPEG-1, MPEG-2, H3.20, H3.24
- *object-based* – MPEG-4

As will be explained in more detail when discussing the MPEG-4 standard in section 3.2, there are a number of advantages with an object-based approach. There is, however, also a price to pay. Usually (object) segmentation does not come for free, but requires additional effort in the phase of authoring and coding.

MPEG-1 To conclude this section on codecs, let's look in somewhat more detail at what is involved in coding and decoding a video signal according to the MPEG-1 standard.

MPEG-1 video compression uses both *intra-frame analysis*, for the compression of individual frames (which are like images), as well as *inter-frame analysis*, to detect redundant blocks or invariants between frames.

The MPEG-1 encoded signal itself is a sequence of so-called I, P and B frames.

frames

- I: intra-frames – independent images
- P: computed from closest frame using DCT (or from P frame)
- B: computed from two closest P or I frames

Decoding takes place by first selecting I-frames, then P-frames, and finally B-frames. When an error occurs, a safeguard is provided by the I-frames, which stand on themselves.

Subsequent standards were developed to accommodate for more complex signals and greater functionality. These include MPEG-2, for higher pixel resolution and data rate, MPEG-3, to support HDTV, MPEG-4, to allow for object-based compression, and MPEG-7, which supports content description. We will elaborate on MPEG-4 in the next section, and briefly discuss MPEG-7 at the end of this chapter.

example(s) – *gigaport*

GigaPort³⁹ is a project focussing on the development and use of advanced and innovative Internet technology. The project, as can be read on the website,

³⁹www.gigaport.nl/info/en/about/home.jsp

focuses on research on next-generation networks and the implementation of a next-generation network for the research community.

Topics for research include:

GigaPort

- optical network technologies - models for network architecture, optical network components and light path provisioning.
- high performance routing and switching - new routing technologies and transport protocols, with a focus on scalability and stability robustness when using data-intensive applications with a high bandwidth demand.
- management and monitoring - incident response in hybrid networks (IP and optical combined) and technologies for network performance monitoring, measuring and reporting.
- grids and access - models, interfaces and protocols for user access to network and grid facilities.
- test methodology - effective testing methods and designing tests for new technologies and network components.

As one of the contributions, internationally, the development of optical technology is claimed, in particular *lambda* networking, networking on a specific wavelength. Locally, the projects has contributed to the introduction of fibre-optic networks in some major cities in the Netherlands.

research directions– *digital video formats*

In the online version you will find a brief overview of *digital video technology*, written by Andy Tanenbaum, as well as some examples of videos of our university, encoded at various bitrates for different viewers.

What is the situation? For traditional television, there are three standards. The american (US) standard, NTSC, is adopted in North-America, South-America and Japan. The european standard, PAL, which seems to be technically superior, is adopted by the rest of the world, except France and the eastern-european countries, which have adopted the other european standard, SECAM. An overview of the technical properties of these standards, with permission taken from Tanenbaum's account, is given below.

system	spatial resolution	frame rate	mbps
NTSC	704 x 480	30	243 mbps
PAL/SECAM	720 x 576	25	249 mbps

Obviously real-time distribution of a more than 200 mbps signal is not possible, using the nowadays available internet connections. Even with compression on the fly, the signal would require 25 mbps, or 36 mbps with audio. Storing the signal on disk is hardly an alternative, considering that one hour would require 12 gigabytes.

When looking at the differences between streaming video (that is transmitted real-time) and storing video on disk, we may observe the following tradeoffs:

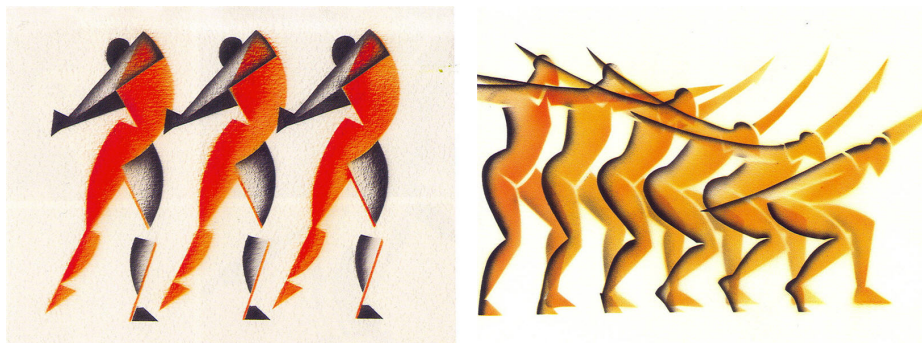
item	streaming	downloaded
bandwidth	equal to the display rate	may be arbitrarily small
disk storage	none	the entire file must be stored
startup delay	almost none	equal to the download time
resolution	depends on available bandwidth	depends on available disk storage

So, what are our options? Apart from the quite successful MPEG encodings, which have found their way in the DVD, there are a number of proprietary formats used for transmitting video over the internet: Quicktime, introduced by Apple, early 1990s, for local viewing; RealVideo, streaming video from RealNetworks; and Windows Media, a proprietary encoding scheme from Microsoft. Examples of these formats, encoded for various bitrates are available at Video at VU.

Apparently, there is some need for digital video on the internet, for example as propaganda for attracting students, for looking at news items at a time that suits you, and (now that digital video cameras become affordable) for sharing details of your family life.

Is digital video all there is? Certainly not! In the next section, we will deal with standards that allow for incorporating (streaming) digital video as an element in a compound multimedia presentation, possibly synchronized with other items, including synthetic graphics. Online, you will find some examples of digital video that are used as texture maps in 3D space. These examples are based on the technology presented in section ??, and use the streaming video codec from Real Networks that is integrated as a rich media extension in the *blaxxun* Contact 3D VRML plugin.

comparison of codecs A review of codecs⁴⁰, including Envivio MPEG-4, QuickTime 6, RealNetworks 9 en Windows Media 9 was published januari 2005 by the European Broadcast Union⁴¹. It appeared that The Real Networks codecs came out best, closely followed by the Windows Media 9 result. Ckeck it out!



⁴⁰www.ebu.ch/trev_301-samviq.pdf

⁴¹www.ebu.ch/trev_home.html

3.2 standards

Imagine what it would be like to live in a world without standards. You may get the experience when you travel around and find that there is a totally different socket for electricity in every place that you visit.

Now before we continue, you must realize that there are two types of standards: *de facto* market standards (enforced by sales politics) and committee standards (that are approved by some official organization). For the latter type of standards to become effective, they need consent of the majority of market players.

For multimedia on the web, we will discuss three standards and RM3D which was once proposed as a standard and is now only of historical significance.

standards

- XML – eXtensible Markup Language (SGML)
- MPEG-4 – coding audio-visual information
- SMIL – Synchronized Multimedia Integration Language
- RM3D – (Web3D) Rich Media 3D (extensions of X3D/VRML)

XML, the *eXtensible Markup Language*, is becoming widely accepted. It is being used to replace HTML, as well as a data exchange format for, for example, business-to-business transactions. XML is derived from SGML (Structured Generalized Markup Language) that has found many applications in document processing. As SGML, XML is a generic language, in that it allows for the specification of actual markup languages. Each of the other three standards mentioned allows for a syntactic encoding using XML.

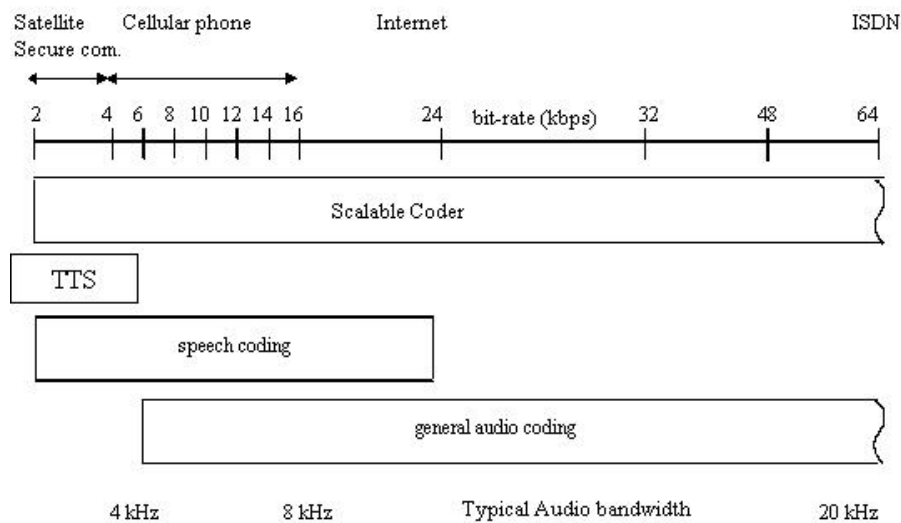
MPEG-4 aims at providing "the standardized technological elements enabling the integration of production, distribution and content access paradigms of digital television, interactive graphics and multimedia", Koenen (2000). A preliminary version of the standard has been approved in 1999. Extensions in specific domains are still in progress.

SMIL, the *Synchronized Multimedia Integration Language*, has been proposed by the W3C "to enable the authoring of TV-like multimedia presentations, on the Web". The SMIL language is an easy to learn HTML-like language. SMIL presentations can be composed of streaming audio, streaming video, images, text or any other media type, W3C (2001). SMIL-1 has become a W3C recommendation in 1998. SMIL-2 is at the moment of writing still in a draft stage.

RM3D, *Rich Media 3D*, is not a standard as MPEG-4 and SMIL, since it does currently not have any formal status. The RM3D working group arose out of the X3D working group, that addressed the encoding of VRML97 in XML. Since there were many disagreements on what should be the core of X3D and how extensions accomodating VRML97 and more should be dealt with, the RM3D working group was founded in 2000 to address the topics of extensibility and the integration with rich media, in particular video and digital television.

remarks Now, from this description it may seem as if these groups work in total isolation from eachother. Fortunately, that is not true. MPEG-4, which is the

most encompassing of these standards, allows for an encoding both in SMIL and X3D. The X3D and RM3D working groups, moreover, have advised the MPEG-4 committee on how to integrate 3D scene description and human avatar animation in MPEG-4. And finally, there have been rather intense discussions between the SMIL and RM3D working groups on the timing model needed to control animation and dynamic properties of media objects.



4

MPEG-4

The MPEG standards (in particular 1,2 and 3) have been a great success, as testified by the popularity of mp3 and DVD video.

Now, what can we expect from MPEG-4? Will MPEG-4 provide *multimedia for our time*, as claimed in Koenen (1999). The author, Rob Koenen, is senior consultant at the dutch KPN telecom research lab, active member of the MPEG-4 working group and editor of the MPEG-4 standard document.

"Perhaps the most immediate need for MPEG-4 is defensive. It supplies tools with which to create uniform (and top-quality) audio and video encoders on the Internet, preempting what may become an unmanageable tangle of proprietary formats."

Indeed, if we are looking for a general characterization it would be that MPEG-4 is primarily

MPEG-4

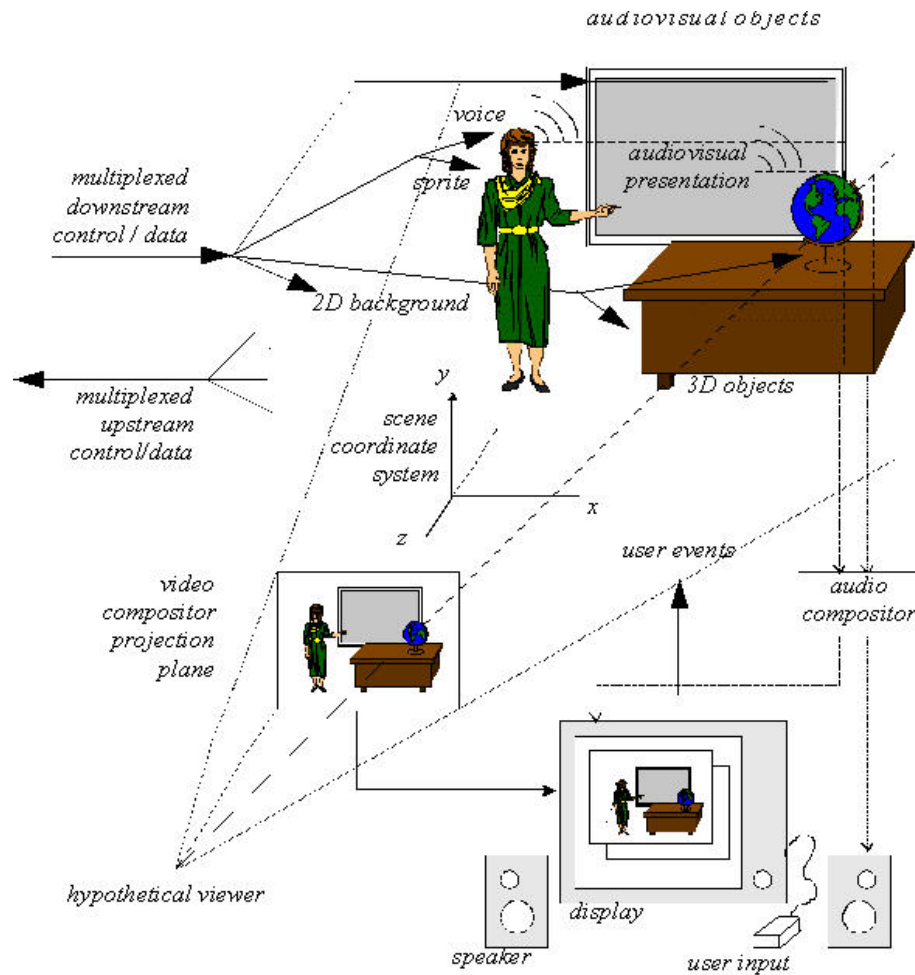
a toolbox of advanced compression algorithms for audiovisual information

and, moreover, one that is suitable for a variety of display devices and networks, including low bitrate mobile networks. MPEG-4 supports scalability on a variety of levels:

scalability

- *bitrate* – switching to lower bitrates
- *bandwidth* – dynamically discard data
- *encoder and decoder complexity* – signal quality

Dependent on network resources and platform capabilities, the 'right' level of signal quality can be determined by selecting the optimal codec, dynamically.



media objects It is fair to say that MPEG-4 is a rather ambitious standard. It aims at offering support for a great variety of audiovisual information, including still images, video, audio, text, (synthetic) talking heads and synthesized speech, synthetic graphics and 3D scenes, streamed data applied to media objects, and user interaction – e.g. changes of viewpoint.

Let's give an example, taken from the MPEG-4 standard document.

example

Imagine, a talking figure standing next to a desk and a projection screen, explaining the contents of a video that is being projected on the screen, pointing at a globe that stands on the desk. The user that is watching that scene decides to change from viewpoint to get a better look at the globe ...

How would you describe such a scene? How would you encode it? And how would you approach decoding and user interaction?

The solution lies in defining *media objects* and a suitable notion of composition of media objects.

media objects

- *media objects* – units of aural, visual or audiovisual content
- *composition* – to create compound media objects (audiovisual scene)
- *transport* – multiplex and synchronize data associated with media objects
- *interaction* – feedback from users' interaction with audiovisual scene

For 3D-scene description, MPEG-4 builds on concepts taken from VRML (Virtual Reality Modeling Language, discussed in chapter 7).

Composition, basically, amounts to building a *scene graph*, that is a tree-like structure that specifies the relationship between the various simple and compound media objects. Composition allows for placing media objects anywhere in a given coordinate system, applying transforms to change the appearance of a media object, applying streamed data to media objects, and modifying the users viewpoint.

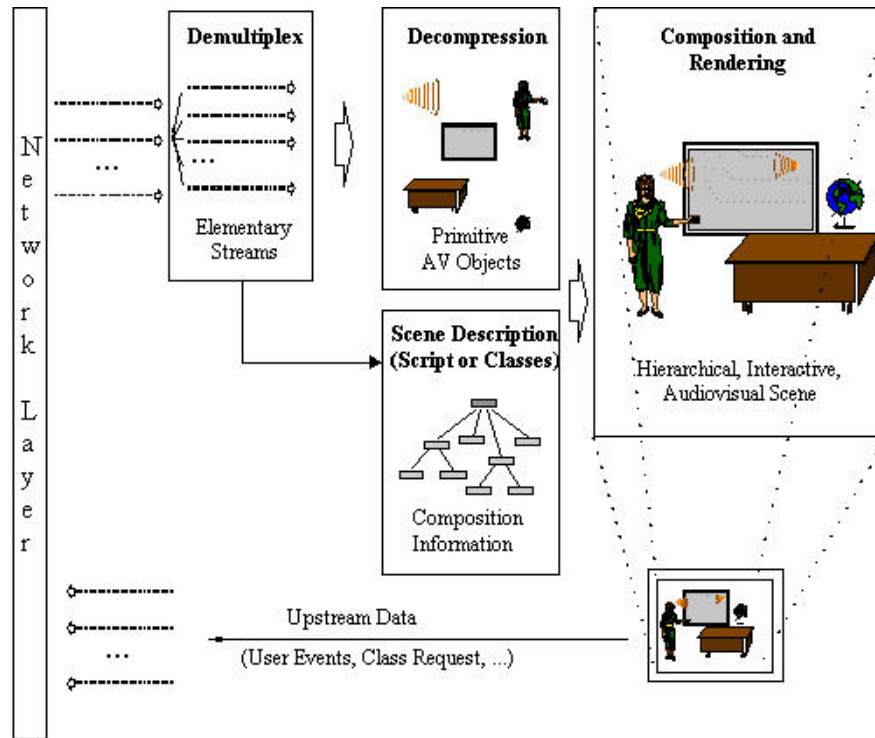
So, when we have a multimedia presentation or audiovisual scene, we need to get it across some network and deliver it to the end-user, or as phrased in Koenen (2000):

transport

The data stream (Elementary Streams) that result from the coding process can be transmitted or stored separately and need to be composed so as to create the actual multimedia presentation at the receivers side.

At a system level, MPEG-4 offers the following functionalities to achieve this:

- BIFS (Binary Format for Scenes) – describes spatio-temporal arrangements of (media) objects in the scene
- OD (Object Descriptor) – defines the relationship between the elementary streams associated with an object
- *event routing* – to handle user interaction

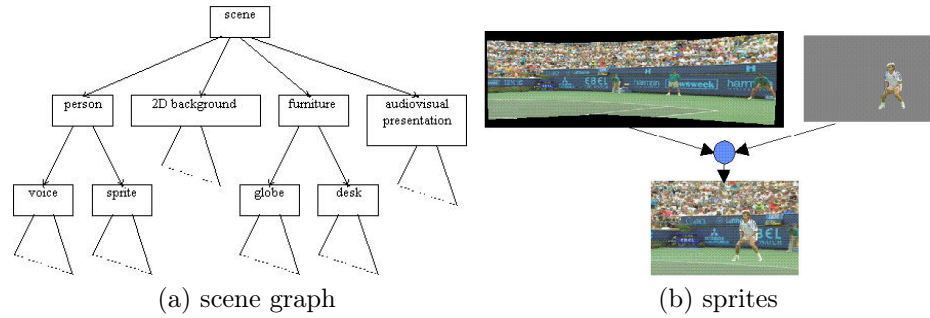


6

In addition, MPEG-4 defines a set of functionalities For the delivery of streamed data, DMIF, which stands for

Delivery Multimedia Integration Framework

that allows for transparent interaction with resources, irrespective of whether these are available from local storage, come from broadcast, or must be obtained from some remote site. Also transparency with respect to network type is supported. *Quality of Service* is only supported to the extent that it is possible to indicate needs for bandwidth and transmission rate. It is however the responsibility of the network provider to realize any of this.



7

authoring What MPEG-4 offers may be summarized as follows

benefits

- *end-users* – interactive media across all platforms and networks
- *providers* – transparent information for transport optimization
- *authors* – reusable content, protection and flexibility

In effect, although MPEG-4 is primarily concerned with efficient encoding and scalable transport and delivery, the *object-based* approach has also clear advantages from an authoring perspective.

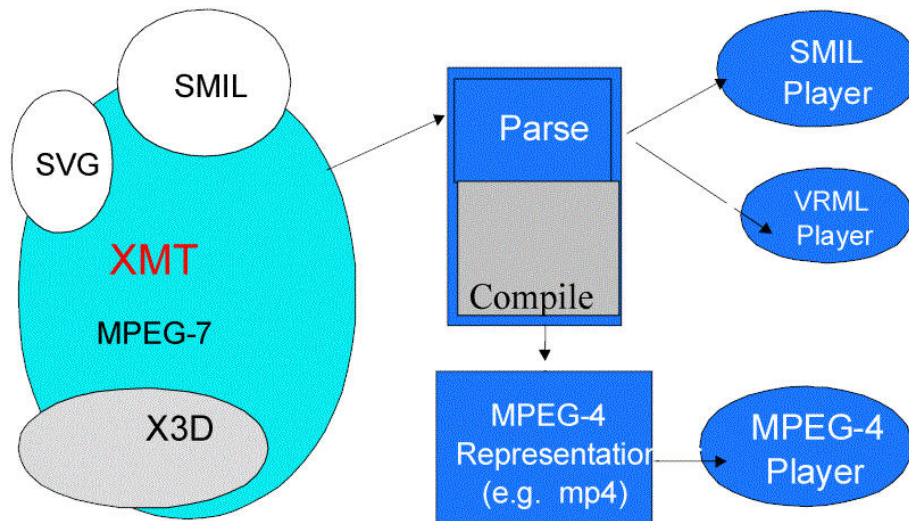
One advantage is the possibility of reuse. For example, one and the same background can be reused for multiple presentations or plays, so you could imagine that even an amateur game might be 'located' at the centre-court of Roland Garros or Wimbledon.

Another, perhaps not so obvious, advantage is that provisions have been made for

managing intellectual property

of media objects.

And finally, media objects may potentially be annotated with meta-information to facilitate information retrieval.



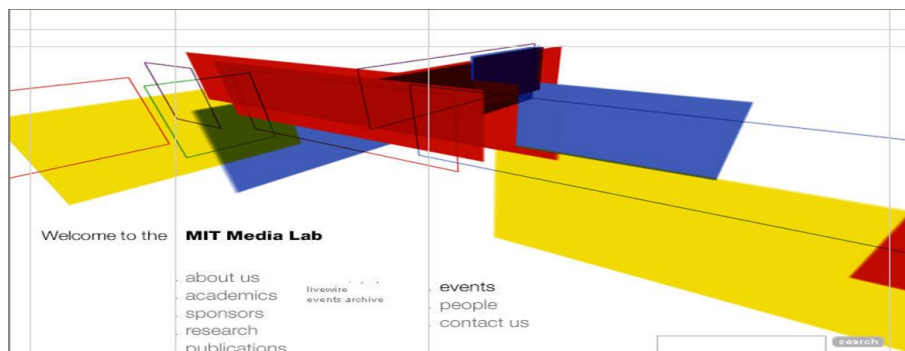
8

syntax In addition to the binary formats, MPEG-4 also specifies a syntactical format, called XMT, which stands for *eXtensible MPEG-4 Textual format*.

XMT

- XMT contains a subset of X3D
- SMIL is mapped (incompletely) to XMT

when discussing RM3D which is of interest from a historic perspective, we will further establish what the relations between, respectively MPEG-4, SMIL and RM3D are, and in particular where there is disagreement, for example with respect to the timing model underlying animations and the temporal control of media objects.



9

example(s) – *structured audio*

The Machine Listening Group⁴² of the MIT Media Lab⁴³ is developing a suite of tools for *structured audio*, which means *transmitting sound by describing it rather than compressing it*. It is claimed that tools based on the MPEG-4 standard will be the future platform for computer music, audio for gaming, streaming Internet radio, and other multimedia applications.

The structured audio project is part of a more encompassing research effort of the Music, Mind and Machine Group⁴⁴ of the MIT Media Lab, which *envisages a new future of audio technologies and interactive applications that will change the way music is conceived, created, transmitted and experienced*,

SMIL

SMIL is pronounced as *smile*. SMIL, the Synchronized Multimedia Integration Language, has been inspired by the Amsterdam Hypermedia Model (AHM). In fact, the dutch research group at CWI that developed the AHM actively participated in the SMIL 1.0 committee. Moreover, they have started a commercial spinoff to create an editor for SMIL, based on the editor they developed for CMIF. The name of the editor is GRINS. Get it?

As indicated before SMIL is intended to be used for

TV-like multimedia presentations

The SMIL language is an XML application, resembling HTML. SMIL presentations can be written using a simple text-editor or any of the more advanced tools, such as GRINS. There is a variety of SMIL players. The most wellknown perhaps is the RealNetworks G8 players, that allows for incorporating RealAudio and RealVideo in SMIL presentations.

parallel and sequential

Authoring a SMIL presentation comes down, basically, to name media components for text, images, audio and video with URLs, and to schedule their presentation either in parallel or in sequence.

Quoting the SMIL 2.0 working draft, we can characterize the SMIL presentation characteristics as follows:

presentation characteristics

- The presentation is composed from several components that are accessible via URL's, e.g. files stored on a Web server.
- The components have different media types, such as audio, video, image or text. The begin and end times of different components are specified relative to events in other media components. For example, in a slide show, a particular slide is displayed when the narrator in the audio starts talking about it.

⁴²sound.media.mit.edu/mpeg4

⁴³www.media.mit.edu

⁴⁴sound.media.mit.edu

- Familiar looking control buttons such as stop, fast-forward and rewind allow the user to interrupt the presentation and to move forwards or backwards to another point in the presentation.
- Additional functions are "random access", i.e. the presentation can be started anywhere, and "slow motion", i.e. the presentation is played slower than at its original speed.
- The user can follow hyperlinks embedded in the presentation.

Where HTML has become successful as a means to write simple hypertext content, the SMIL language is meant to become a vehicle of choice for writing *synchronized hypermedia*. The working draft mentions a number of possible applications, for example a photoalbum with spoken comments, multimedia training courses, product demos with explanatory text, timed slide presentations, online music with controls.

As an example, let's consider an interactive news bulletin, where you have a choice between viewing a weather report or listening to some story about, for example, the decline of another technology stock. Here is how that could be written in SMIL:

example

```
<par>
  <a href="#Story">  </a>
  <a href="#Weather"> </a>
  <excl>
    <par id="Story" begin="0s">
      <video src="video1.mpg"/>
      <text src="captions.html"/>
    </par>

    <par id="Weather">
      
      <audio src="weather-rpt.mp3"/>
    </par>
  </excl>
</par>
```

Notice that there are two *parallel* (PAR) tags, and one *exclusive* (EXCL) tag. The *exclusive* tag has been introduced in SMIL 2.0 to allow for making an exclusive choice, so that only one of the items can be selected at a particular time. The SMIL 2.0 working draft defines a number of elements and attributes to control presentation, synchronization and interactivity, extending the functionality of SMIL 1.0.

Before discussing how the functionality proposed in the SMIL 2.0 working draft may be realized, we might reflect on how to position SMIL with respect to the many other approaches to provide multimedia on the web. As other approaches we may think of *flash*, dynamic HTML (using javascript), or java applets. In the SMIL 2.0 working draft we read the following comment:

history

Experience from both the CD-ROM community and from the Web multimedia community suggested that it would be beneficial to adopt a declarative format for expressing media synchronization on the Web as an alternative and complementary approach to scripting languages.

Following a workshop in October 1996, W3C established a first working group on synchronized multimedia in March 1997. This group focused on the design of a declarative language and the work gave rise to SMIL 1.0 becoming a W3C Recommendation in June 1998.

In summary, SMIL 2.0 proposes a *declarative format* to describe the temporal behavior of a multimedia presentation, associate hyperlinks with media objects, describe the form of the presentation on a screen, and specify interactivity in multimedia presentations. Now, why such a fuzz about "declarative format"? Isn't scripting more exciting? And aren't the tools more powerful? Ok, ok. I don't want to go into that right now. Let's just consider a *declarative format* to be more elegant. Ok?

To support the functionality proposed for SMIL 2.0 the working draft lists a number of modules that specify the interfaces for accessing the attributes of the various elements. SMIL 2.0 offers modules for animation, content control, layout, linking, media objects, meta information, timing and synchronization, and transition effects.

This modular approach allows to reuse SMIL syntax and semantics in other XML-based languages, in particular those that need to represent timing and synchronization. For example:

module-based reuse

- SMIL modules could be used to provide lightweight multimedia functionality on mobile phones, and to integrate timing into profiles such as the WAP forum's WML language, or XHTML Basic.
- SMIL timing, content control, and media objects could be used to coordinate broadcast and Web content in an enhanced-TV application.
- SMIL Animation is being used to integrate animation into W3C's Scalable Vector Graphics language (SVG).
- Several SMIL modules are being considered as part of a textual representation for MPEG4.

The SMIL 2.0 working draft is at the moment of writing being finalized. It specifies a number of language profiles to promote the reuse of SMIL modules. It also improves on the accessibility features of SMIL 1.0, which allows for, for example, replacing captions by audio descriptions.

In conclusion, SMIL 2.0 is an interesting standard, for a number of reasons. For one, SMIL 2.0 has solid theoretical underpinnings in a well-understood, partly formalized, hypermedia model (AHM). Secondly, it proposes interesting functionality, with which authors can make nice applications. In the third place, it specifies a high level declarative format, which is both expressive and flexible. And finally, it is an open standard (as opposed to proprietary standard). So everybody can join in and produce players for it!



10

RM3D – not a standard

The web started with simple HTML hypertext pages. After some time static images were allowed. Now, there is support for all kinds of user interaction, embedded multimedia and even synchronized hypermedia. But despite all the graphics and fancy animations, everything remains flat. Perhaps surprisingly, the need for a 3D web standard arose in the early days of the web. In 1994, the acronym VRML was coined by Tim Berners-Lee, to stand for *Virtual Reality Markup Language*. But, since 3D on the web is not about text but more about worlds, VRML came to stand for *Virtual Reality Modeling Language*. Since 1994, a lot of progress has been made.

www.web3d.org

- VRML 1.0 – *static 3D worlds*
- VRML 2.0 or VRML97 – *dynamic behaviors*
- VRML200x – *extensions*
- X3D – *XML syntax*
- RM3D – *Rich Media in 3D*

In 1997, VRML2 was accepted as a standard, offering rich means to create 3D worlds with dynamic behavior and user interaction. VRML97 (which is the same as VRML2) was, however, not the success it was expected to be, due to (among others) incompatibility between browsers, incomplete implementations of the standards, and high performance requirements.

As a consequence, the Web3D Consortium (formerly the VRML Consortium) broadened its focus, and started thinking about extensions or modifications of VRML97 and an XML version of VRML (X3D). Some among the X3D working group felt the need to rethink the premisses underlying VRML and started the Rich Media Working Group:

groups.yahoo.com/group/rm3d/

The Web3D Rich Media Working Group was formed to develop a Rich Media standard format (RM3D) for use in next-generation media devices. It is a highly active group with participants from a broad range of companies including 3Dlabs, ATI, Eyematic, OpenWorlds, Out of the Blue Design, Shout Interactive, Sony, Uma, and others.

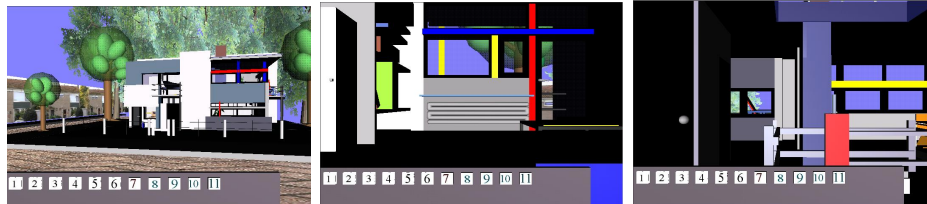
In particular:

RM3D

The Web3D Consortium initiative is fueled by a clear need for a standard high performance Rich Media format. Bringing together content creators with successful graphics hardware and software experts to define RM3D will ensure that the new standard addresses authoring and delivery of a new breed of interactive applications.

The working group is active in a number of areas including, for example, multi-texturing and the integration of video and other streaming media in 3D worlds.

Among the driving forces in the RM3D group are Chris Marrin and Richter Rafey, both from Sony, that proposed *Blendo*, a rich media extension of VRML. Blendo has a strongly typed object model, which is much more strictly defined than the VRML object model, to support both declarative and programmatic extensions. It is interesting to note that the premisses underlying the Blendo proposal confirms (again) the primacy of the TV metaphor. That is to say, what Blendo intends to support are TV-like presentations which allow for user interaction such as the selection of items or playing a game. Target platforms for Blendo include graphic PCs, set-top boxes, and the Sony Playstation!



11

requirements The focus of the RM3D working group is not *syntax* (as it is primarily for the X3D working group) but *semantics*, that is to enhance the VRML97 standard to effectively incorporate rich media. Let's look in more detail at the requirements as specified in the RM3Ddraft proposal.

requirements

- *rich media* – audio, video, images, 2D & 3D graphics (with support for temporal behavior, streaming and synchronisation)
- *applicability* – specific application areas, as determined by commercial needs and experience of working group members

The RM3D group aims at interoperability with other standards.

- *interoperability* – VRML97, X3D, MPEG-4, XML (DOM access)

In particular, an XML syntax is being defined in parallel (including interfaces for the DOM). And, there is mutual interest and exchange of ideas between the MPEG-4 and RM3D working group.

As mentioned before, the RM3D working group has a strong focus on defining an object model (that acts as a common model for the representation of objects and their capabilities) and suitable mechanisms for extensibility (allowing for the integration of new objects defined in Java or C++, and associated scripting primitives and declarative constructs).

Notice that extensibility also requires the definition of a declarative format, so that the content author need not bother with programmatic issues.

The RM3D proposal should result in effective 3D media presentations. So as additional requirements we may, following the working draft, mention: high-quality realtime rendering, for realtime interactive media experiences; platform adaptability, with query functions for programmatic behavior selection; predictable behavior, that is a well-defined order of execution; a high precision number systems, greater than single-precision IEEE floating point numbers; and minimal size, that is both download size and memory footprint.

Now, one may be tempted to ask how the RM3D proposals is related to the other standard proposals such as MPEG-4 and SMIL, discussed previously. Briefly put, paraphrased from one of Chris Marrin's messages on the RM3D mailing list

SMIL is closer to the author and RM3D is closer to the implementer.

MPEG-4, in this respect is even further away from the author since its chief focus is on compression and delivery across a network.

RM3D takes 3D scene description as a starting point and looks at pragmatic ways to integrate rich media. Since 3D is itself already computationally intensive, there are many issues that arise in finding efficient implementations for the proposed solutions.



12

timing model RM3D provides a declarative format for many interesting features, such as for example texturing objects with video. In comparison to VRML, RM3D is meant to provide more temporal control over time-based media objects and animations. However, there is strong disagreement among the working group members as to what time model the dynamic capabilities of RM3D should be based on. As we read in the working draft:

working draft

Since there are three vastly different proposals for this section (time model), the original <RM3D> 97 text is kept. Once the issues concerning time-dependent nodes are resolved, this section can be modified appropriately.

Now, what are the options? Each of the standards discussed to far provides us with a particular solution to timing. Summarizing, we have a time model based on a spring metaphor in MPEG-4, the notion of cascading time in SMIL (inspired by cascading stylesheets for HTML) and timing based on the routing of events in RM3D/VRML.

The MPEG-4 standard introduces the *spring metaphor* for dealing with temporal layout.

MPEG-4 – spring metaphor

- duration – minimal, maximal, optimal

The spring metaphor amounts to the ability to shrink or stretch a media object within given bounds (minimum, maximum) to cope with, for example, network delays.

The SMIL standard is based on a model that allows for propagating durations and time manipulations in a hierarchy of media elements. Therefore it may be referred to as a cascading model of time.

SMIL – cascading time

- time container – speed, accelerate, decelerate, reverse, synchronize

Media objects, in SMIL, are stored in some sort of container of which the timing properties can be manipulated.

```
<seq speed="2.0">
  <video src="movie1.mpg" dur="10s"/>
  <video src="movie2.mpg" dur="10s"/>
  
    <animateMotion from="-100,0" to="0,0" dur="10s"/>
  </img>
  <video src="movie4.mpg" dur="10s"/>
</seq>
```

In the example above, we see that the speed is set to *2.0*, which will affect the pacing of each of the individual media elements belonging to that (sequential) group. The duration of each of the elements is specified in relation to the parent container. In addition, SMIL offers the possibility to synchronize media objects to control, for example, the end time of parallel media objects.

VRML97's capabilities for timing rely primarily on the existence of a *TimeSensor* that sends out time events that may be routed to other objects.

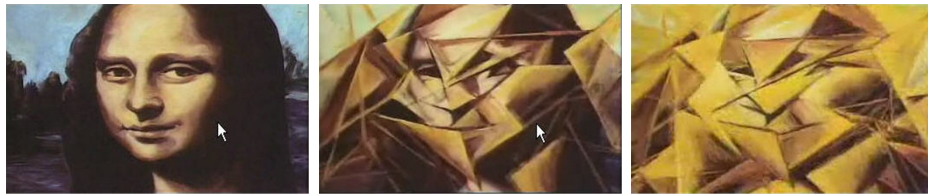
RM3D/VRML – event routing

- *TimeSensor* – isActive, start, end, cycleTime, fraction, loop

When a *TimeSensor* starts to emit time events, it also sends out an event notifying other objects that it has become active. Dependent on its so-called *cycleTime*, it sends out the fraction it covered since it started. This fraction may be sent to one of the standard interpolators or a script so that some value can be set, such as for

example the orientation, dependent on the fraction of the time intercal that has passed. When the *TimeSensor* is made to loop, this is done repeatedly. Although time in VRML is absolute, the frequency with which fraction events are emitted depends on the implementation and processor speed.

Lacking consensus about a better model, this model has provisionally been adopted, with some modifications, for RM3D. Nevertheless, the SMIL cascading time model has raised an interest in the RM3D working group, to the extent that Chris Marrin remarked (in the mailing list) "*we could go to school here*". One possibility for RM3D would be to introduce *time containers* that allow for a temporal transform of their children nodes, in a similar way as grouping containers allow for spatial transforms of their children nodes. However, that would amount to a dual hierarchy, one to control (spatial) rendering and one to control temporal characteristics. Merging the two hierarchies, as is (implicitly) the case in SMIL, might not be such a good idea, since the rendering and timing semantics of the objects involved might be radically different. An interesting problem, indeed, but there seems to be no easy solution.



example(s) – *rich internet applications*

In a seminar held by *Lost Boys*, which is a dutch subdivision of Icon Media Lab⁴⁵, *rich internet applications* (RIA), were presented as the new solutions to present applications on the web. As indicated by Macromedia⁴⁶, who is one of the leading companies in this field, *experience matters*, and so plain html pages do not suffice since they require the user to move from one page to another in a quite unintuitive fashion. Macromedia presents its new line of *flash*-based products to create such *rich internet applications*. An alternative solution, based on general W3C recommendations, is proposed by BackBase⁴⁷. Interestingly enough, using either technology, many of the participants of the seminar indicated a strong preference for a backbutton, having similar functionality as the often used backbutton in general internet browsers.

⁴⁵www.iconmedialab.com

⁴⁶www.macromedia.com/resources/business/rich_internet_apps/whitepapers.html

⁴⁷www.backbase.com

research directions— *meta standards*

All these standards! Wouldn't it be nice to have one single standard that encompasses them all? No, it would not! Simply, because such a standard is inconceivable, unless you take some proprietary standard or a particular platform as the defacto standard (which is the way some people look at the Microsoft win32 platform, ignoring the differences between 95/98/NT/2000/XP/...). In fact, there is a standard that acts as a glue between the various standards for multimedia, namely XML. XML allows for the interchange of data between various multimedia applications, that is the transformation of one encoding into another one. But this is only syntax. What about the semantics?

Both with regard to delivery and presentation the MPEG-4 proposal makes an attempt to delineate chunks of core functionality that may be shared between applications. With regard to presentation, SMIL may serve as an example. SMIL applications themselves already (re)use functionality from the basic set of XML-related technologies, for example to access the document structure through the DOM (Document Object Model). In addition, SMIL defines components that it may potentially share with other applications. For example, SMIL shares its animation facilities with SVG (the Scalable Vector Graphics format recommended by the Web Consortium).

The issue in sharing is, obviously, how to relate constructs in the syntax to their operational support. When it is possible to define a common base of operational support for a variety of multimedia applications we would approach our desired meta standard, it seems. A partial solution to this problem has been proposed in the now almost forgotten HyTime standard for time-based hypermedia. HyTime introduces the notion of *architectural forms* as a means to express the operational support needed for the interpretation of particular encodings, such as for example synchronization or navigation over bi-directional links. Apart from a base module, HyTime compliant architectures may include a units measurement module, a module for dealing with location addresses, a module to support hyperlinks, a scheduling module and a rendition module.

To conclude, wouldn't it be wonderful if, for example, animation support could be shared between rich media X3D and SMIL? Yes, it would! But as you may remember from the discussion on the timing models used by the various standards, there is still too much divergence to make this a realistic option.



3.3 a multimedia semantic web?

To finish this chapter, let's reflect on where we are now with 'multimedia' on the web. Due to refined compression schemes and standards for authoring and delivery, we seemed to have made great progress in realizing *networked multimedia*. But does this progress match what has been achieved for the dominant media type of the web, that is text or more precisely textual documents with markup?

web content

- *1st generation* – hand-coded HTML pages
- *2nd generation* – templates with content and style
- *3rd generation* – rich markup with metadata (XML)

Commonly, a distinction is made between successive generations of web content, with the first generation being simple hand-coded HTML pages. The second generation may be characterized as HTML pages that are generated on demand, for example by filling in templates with contents retrieved from a database. The third generation is envisaged to make use of rich markup, using XML, that reflects the (semantic) content of the document more directly, possibly augmented with (semantic) meta-data that describe the content in a way that allows machines, for example search engines, to process it. The great vision underlying the third generation of web content is commonly referred to as the *the semantic web*, which enhances the functionality of the current web by deploying knowledge representation and inference technology from Artificial Intelligence, using a technology known as the *Resource Description Framework* (RDF). As phrased in Ossenbruggen et. al. (2001), the semantic web will bring

structure to the meaningful content of web pages,

thus allowing computer programs, such as search engines and intelligent agents, to do their job more effectively. For search engines this means more effective information retrieval, and for agents better opportunities to provide meaningful services.

A great vision indeed. So where are we with multimedia? As an example, take a *shockwave* or *flash* presentation showing the various musea in Amsterdam. How would you attach meaning to it, so that it might become an element of a semantic structure? Perhaps you wonder what meaning could be attached to it? That should not be too difficult to think of. The (meta) information attached to such a presentation should state (minimally) that the location is Amsterdam, that the sites of interest are musea, and (possibly) that the perspective is touristic. In that way, when you search for touristic information about musea in Amsterdam, your search engine should have no trouble in selecting that presentation. Now, the answer to the question how meaning can be attached to a presentation is already given, namely by specifying meta-information in some format (of which the only requirement is that it is machine-processable). For our *shockwave* or *flash* presentation we cannot do this in a straightforward manner. But for MPEG-4 encoded material, as well as for SMIL content, such facilities are readily available.

Should we then always duplicate our authoring effort by providing (meta) information, on top of the information that is already contained in the presentation? No, in some cases, we can also rely to some extent on content-based search or feature extraction, as will be discussed in the following chapters.



15

Resource Description Framework – the Dublin Core

The Resource Description Framework, as the W3C/RDF⁴⁸ site informs us *integrates a variety of applications from library catalogs and world-wide directories to syndication and aggregation of news, software, and content to personal collections of music, photos, and events using XML as an interchange syntax*. The RDF specifications provide, in addition *a lightweight ontology system to support the exchange of knowledge on the Web*.

The Dublin Core Metadata Initiative⁴⁹ is an open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models.

What exactly is meta-data? As phrased in the RDF Primer⁵⁰

meta data

Metadata is data about data. Specifically, the term refers to data used to identify, describe, or locate information resources, whether these resources are physical or electronic. While structured metadata processed by computers is relatively new, the basic concept of metadata has been used for many years in helping manage and use large collections of information. Library card catalogs are a familiar example of such metadata.

The Dublin Core proposes a small number of elements, to be used to give information about a resource, such as an electronic document on the Web. Consider the following example:

Dublin Core example

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns #"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dcterms="http://purl.org/dc/terms/">
```

⁴⁸www.w3.org/RDF

⁴⁹dublincore.org

⁵⁰www.w3.org/TR/rdf-primer

```

<rdf:Description rdf:about="http://www.dlib.org/dlib/may98/miller/05miller.html">
  <dc:title>An Introduction to the Resource Description Framework</dc:title>
  <dc:creator>Eric J. Miller</dc:creator>
  <dc:description>The Resource Description Framework (RDF) is an
    infrastructure that enables the encoding, exchange and reuse of
    structured metadata. rdf is an application of xml that imposes needed
    structural constraints to provide unambiguous methods of expressing
    semantics. rdf additionally provides a means for publishing both
    human-readable and machine-processable vocabularies designed to
    encourage the reuse and extension of metadata semantics among
    disparate information communities. the structural constraints rdf
    imposes to support the consistent encoding and exchange of
    standardized metadata provides for the interchangeability of separate
    packages of metadata defined by different resource description
    communities. </dc:description>
  <dc:publisher>Corporation for National Research Initiatives</dc:publisher>
  <dc:subject>
    <rdf:Bag>
      <rdf:li>machine-readable catalog record formats</rdf:li>
      <rdf:li>applications of computer file organization and
        access methods</rdf:li>
    </rdf:Bag>
  </dc:subject>
  <dc:rights>Copyright © 1998 Eric Miller</dc:rights>
  <dc:type>Electronic Document</dc:type>
  <dc:format>text/html</dc:format>
  <dc:language>en</dc:language>
  <dcterms:isPartOf rdf:resource="http://www.dlib.org/dlib/may98/05contents.html"/>
</rdf:Description>
</rdf:RDF>

```

Items such as *title*, *creator*, *subject* and *description*, actually all tags with the prefix *dc*, belong to the Dublin Core and are used to give information about the document, which incidentally concerns an introduction to the Resource Description Framework. The example also shows how *rdf* constructs can be used together with the Dublin Core elements. The prefixes *rdf* and *dc* are used to distinguish between the distinct namespaces of respectively RDF and the Dublin Core.

The Dublin Core contains the following elements:

Dublin Core⁵¹

- *title* – name given to the resource
- *creator* – entity primarily responsible for making the content of the resource
- *subject* – topic of the content of the resource
- *description* – an account of the content of the resource
- *publisher* – entity responsible for making the resource available

⁵¹dublincore.org/documents/dces

- *contributor* – entity responsible for making contributions to the content of the resource
- *date* – date of an event in the lifecycle of the resource
- *type* – nature or genre of the content of the resource
- *format* – physical or digital manifestation of the resource
- *identifier* – unambiguous reference to the resource within a given context
- *source* – reference to a resource from which the present resource is derived
- *language* – language of the intellectual content of the resource
- *relation* – reference to a related resource
- *coverage* – extent or scope of the content of the resource
- *rights* – information about rights held in and over the resource

In section 10.3 we discuss an application in the domain of cultural heritage, where the Dublin Core elements are used to provide meta information about the information available for the conservation of contemporary artworks.



16

research directions– *agents everywhere*

The web is an incredibly rich resource of information. Or, as phrased in Baeza-Yates and Ribeiro-Neto (1999):

information repository

The Web is becoming a universal repository of human knowledge and culture, which has allowed unprecedented sharing of ideas and information in a scale never seen before.

Now, the problem (as many of you can acknowledge) is to get the information out of it. Of course, part of the problem is that we often do not know what we are looking for. But even if we do know, it is generally not so easy to find our way. Again using the phrasing of Baeza-Yates and Ribeiro-Neto (1999):

browsing & navigation

To satisfy his information need, the user might navigate the hyperspace of web links searching for information of interest. However, since the hyperspace is vast and almost unknown, such a navigation task is usually inefficient.

The solution of the problem of *getting lost in hyperspace* proposed in Baeza-Yates and Ribeiro-Neto (1999) is information retrieval, in other words *query & search*. However, this may not so easily be accomplished. As observed in Baeza-Yates and Ribeiro-Neto (1999), The main obstacle is the absence of a well-defined data model for the Web, which implies that information definition and structure is frequently of low quality. Well, that is exactly the focus of the semantics web initiative, and in particular of the Resource Description Framework discussed above.

Standardizing knowledge representation and reasoning about web resources is certainly one (important) step. Another issue, however, is how to support the user in finding the proper resources and provide the user with assistance in accomplishing his task (even if this task is merely finding suitable entertainment).

What we need, in other words, is a unifying model (encompassing both a data model and a model of computation) that allows us to deal effectively with web resources, including multimedia objects. For such a model, we may look at another area of research and development, namely *intelligent agents*, which provides us not only with a model but also with a suitable metaphor and the technology, based on and extending object-oriented technology, to realize intelligent assistance, Eliens (2000).

For convenience, we make a distinction between two kinds of agents, *information agents* and *presentation agents*.

information agent

- gather information
- filter and select

Information agents are used to gather information. In addition, they filter the information and select those items that are relevant for the user. A key problem in developing information agents, however, is to find a proper representation of what the user considers to be relevant.

presentation agent

- access information
- find suitable mode of presentation

Complementary to the information agent is a *presentation agent* (having access to the information gathered) that displays the relevant information in a suitable way. Such a presentation agent can have many forms. To appease your phantasy, you may look at the vision of *angelic guidance* presented in Broll et. al (2001). More concretely, my advice is to experiment with embodied agents that may present information in rich media 3D. In section ??, we will present a framework for doing such experiments.



17

navigating information spaces Having *agents everywhere* might change our perspective on computing. But, it may also become quite annoying to be bothered by an agent each time that you try to interact with your computer (you know what I mean!). However, as reported by Kristina Höök, even annoyance can be instrumental in keeping your attention to a particular task. In one of her projects, the *PERSONAS* project, which stands for

PERSONal and SOcial NAVigation through information spaceS

the use of agents commenting on people navigating information space(s) is explored. As a note, the plural form of *spaces* is mine, to do justice to the plurality of information spaces.

As explained on the *PERSONAS* web site, which is listed with the acronyms, the *PERSONAS* project aims at:

PERSONAS

investigating a new approach to navigation through information spaces, based on a personalised and social navigational paradigm.

The novel idea pursued in this project is to have agents (*Agneta* and *Frieda*) that are not helpful, but instead just give comments, sometimes with humor, but sometimes ironic or even sarcastic comments on the user's activities, in particular navigating an information space or (plain) web browsing. As can be read on the *PERSONAS* web site:

Agneta & Frieda

The AGNETA & FRIDA system seeks to integrate web-browsing and narrative into a joint mode. Below the browser window (on the desktop) are placed two female characters, sitting in their livingroom chairs, watching the browser during the session (more or less like watching television). Agneta and Frida (mother and daughter) physically react, comment, make ironic remarks about and develop stories around the information presented in the browser (primarily to each other), but are also sensitive to what the navigator is doing and possible malfunctions of the browser or server.

In one of her talks, Kristina Höök observed that some users get really fed up with the comments delivered by *Agneta* and *Frieda*. So, as a compromise, the level of interference can be adjusted by the user, dependent on the task at hand.

Agneta & Frieda

In this way they seek to attach emotional, comical or anecdotal connotations to the information and happenings in the browsing session. Through an activity slider, the navigator can decide on how active she wants the characters to be, depending on the purpose of the browsing session (serious information seeking, wayfinding, exploration or entertainment browsing).

As you may gather, looking at the presentations accompanying this *introduction to multimedia* and *Dialogs*, I found the *PERSONAS* approach rather intriguing. Actually, the *PERSONAS* approach is related to the area of *affective computing*, see Picard (1998), which is an altogether different story.

The *Agneta* and *Frieda* software is available for download at the *PERSONAS* web site.

3.4 rethorics of change

using web services in Second Life

Second Life offers an advanced scripting language with a C-like syntax and an extensive library of built-in functionality. Although it has support for objects, LSL (the Linden Scripting Language) is not object-oriented. Cf. Eliens (2000). Scripts in Second Life are server-based, that is all scripts are executed at the server, to allow sharing between visitors. Characteristic for LSL are the notions of *state* and *eventhandler*, which react to events in the environments.

Among the built-in functions there are functions to connect to a (web) server, and obtain a response, in particular (with reference to their wiki page):

- request – wiki.secondlife.com/wiki/LHTTPRequest
- escape – wiki.secondlife.com/wiki/LlEscapeURL
- response – wiki.secondlife.com/wiki/Http_response

Other functions to connect to the world include *sensors*, for example to detect the presence of (visitors') avatars, and chat and instant messaging functions to communicate with other avatars using scripts. In addition, LSL offers functions

to control the behavior and appearance of objects, including functions to make objects react to physical laws, to apply force to objects, to activate objects attached to an avatar (as for example the phantom Mario sprites mentioned earlier), and functions to animate textures, that can be used to present slide shows in Second Life.

On the Mashable⁵² *Social Networking News* site a brief overview is given of the use of web services in Second Life, entitled *Second Life + Web 2.0 = Virtual World Mashups*. To access Second Life from outside-in (that is from a web browser), so-called *slurls* may be used, for example to reach VU⁵³ @ Second Life, and all slurls listed in delicio.us under *slurlmarker*⁵⁴ may be used, also to activate in-world teleporting using scraping techniques.

As remarked in the *hackdiary*⁵⁵ by Matt Biddulph, Second Life (currently) lacks the ability to parse XML or JSON, so the best way to incorporate web services is to set up a web server with adequate resources. As Matt Biddulph indicates, to access *flickr* photographs for a particular user (avatar), a web server may contain the following resources:

- /seen?user=SomeAvatar – records the presence of SomeAvatar
- /touched?user=SomeAvatar – invokes flickr API with users tag
- /set_tag?user=SomeAvatar&tag=FavoriteTag – records SomeAvatar's favourite tag

For example, in response to a 'touch' event, invoking *touch* results in consulting the database for the user's tag and asking the Flickr API for a random photo with that tag. It then returns a string containing the url for a particular photograph. LSL functions used in this application include *sensors*, to check for presence, *listen* functions, to respond to spoken commands, and *touch* events, for the physical interface. In addition to supporting strings and lists, LSL provides a perl-like split function to convert a string into a list of strings, thus allowing for processing multiple items in response to a server request.

Another example of using web services in Second Life is writing blogs⁵⁶ from within Second Life using the BlogHUD⁵⁷ developed by Koz Farina who also is reported to have found a flash hack that allows for reading RSS feeds. As explained by Koz Farina

Quicktime supports Flash, but only up to Flash version 5. We're up to version 9 on that now! Luckily, I have been dabbling with Flash since the early days, so already knew how to do this 'the old way'... So, Flash is doing all the work. No LSL at all... I heart feeds. Did I say 'I heart feeds'?

The RSS display uses the ability to stream Quicktime video in Second Life, and again the mashup is not created in Second Life but by appropriate server support.

⁵² mashable.com/2006/05/30/second-life-web-20-virtual-world-mashups/

⁵³ slurl.com/secondlife/VU%20University%20NL/29/151

⁵⁴ del.icio.us/tag/slurlmarker

⁵⁵ www.hackdiary.com/archives/000085.html

⁵⁶ nwn.blogs.com/nwn/2006/10/really_simple_s.html

⁵⁷ bloghud.com/

In a similar vein we may incorporate live streaming video⁵⁸, for example by using WireCast⁵⁹ to capture and organize live camera input, possibly together the screen output of other applications such as *powerpoint*, which must then be sent to a streaming server supporting Quicktime, such as Apple's Darwin⁶⁰, which may then be accessed from Second Life to texture a display object.

Finally, as another *Web 2.0 to Web 3D* phenomenon, announced in New World Notes⁶¹, we may mention the used of Twitter⁶² messages, that allow residents to send and receive message about ongoing activities. A similar service is reported to exist for *jaiku*⁶³ messages.

rhetorics

In addition, we would like to develop a facility that allows players not only submit their own video material, but also to build or modify their own minigames, which might then be included in the collection of minigames provided by Clima Futura.

For the actual production, we will use additional components, including *game physics*⁶⁴, a *relation browser*⁶⁵, and an *earch*⁶⁶ component. In particular, both physics and in-game building facilities seemed to have contributed to a great extent to the popularity of Second Life, VUSL. In creating *digital dossiers*⁶⁷ for contemporary art, we have deployed concept graphs, that is a relation browser, to give access to highly-related rich media information about art in an immersive manner. Finally, given the topic of Clima Futura, being able to visual models of the surface of the earth seems to be more than appropriate. It is interesting to note that our technology also allows for the use of *flash* movies directly by invoking the *youtube* API⁶⁸ as a web service, which means that we could, in principle, build minigames around the evergrowing collection of *youtube*, or similar providers.

Providing flexible access to collections of video(s) to support arguments concerning controversial issues has been explored in Vox Populi⁶⁹, Vox. The Vox Populi system distinguishes between the following types of argument(s):

- topic-centered – common beliefs, use of logic, examples
- viewer-centered – patriotisms, religious or romantic sentimentality
- speaker-centered – the makers are intelligent, well-informed, sincere and trustworthy

⁵⁸blogs.electricsheepcompany.com/chris/?p=206

⁵⁹www.varasoftware.com/products/wirecast/

⁶⁰developer.apple.com/opensource/server/streaming/

⁶¹nwn.blogs.com/nwn/2007/03/post_1.html

⁶²twitter.com/

⁶³devku.org/docs

⁶⁴www.fisixengine.com

⁶⁵<http://der-mo.net/relationBrowser>

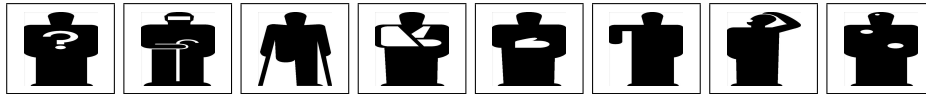
⁶⁶www.flashearth.com

⁶⁷www.few.vu.nl/~dossier05

⁶⁸www.youtube.com/dev

⁶⁹homepages.cwi.nl/~media/demo/IWA/

These argument types are related to what we have previously characterized as, respectively, *logos*, arguments based on logic, reason and factual data, *pathos*, arguments that appeal to the emotion(s) of the audience, and *ethos*, which in essence does an appeal on the belief in the trustworthiness of the speaker. In Vox Populi, video fragments are annotated with meta-information to allow for searching relevant material, supporting or opposing a particular viewpoint. based on the users' preference, either a *propagandist* presentation can be chosen, epressing a single point of view (POV), a *binary commentator*, which shows arguments pro and con, or an *omniscient presenter* (mind opener), which displays all viewpoints. Although a research topic in itself, we would like to develop a *video content module* (3), that provides flexible access to the collection of video(s), and is media driven to the extent that video-material can be added later, with proper annotation. Together with in-game minigame building facilities, it would be in the spirit of a participatory culture, to provide annotation facilities to the player(s) of Clima Futura as well, to comment on the relevance and status of the video material, Participatory.



18

questions

codecs and standards

1. (*) What role do standards play in *multimedia*? Why are standards necessary for compression and delivery. Discuss the MPEG-4 standard and indicate how it is related to other (possible) standards.

concepts

2. What is a *codec*?
3. Give a brief overview of current multimedia standards.
4. What criteria must a (*multimedia*) *semantic web* satisfy?

technology

5. What is the *data rate* for respectively (*compressed*) voice, audio and video?
6. Explain how a *codec* functions.
7. Which considerations can you mention for choosing a compression method?
8. Give a brief description of: XML, MPEG-4, SMIL, RM3D.

projects & further reading As a project, you may think of implementing for example JPEG compression, following Li and Drew (2004), or a SMIL-based application for cultural heritage.

You may further explore the technical issues on authoring DV material, using any of the Adobe⁷⁰, mentioned in appendix E. or compare

For further reading I advice you to take a look at the respective specifications of MPEG-4 and SMIL⁷¹, and compare the functionality of MPEG-4 and SMIL-based presentation environments. An invaluable book dealing with the many technical aspects of compression and standards in Li and Drew (2004).

the artwork

1. costume designs – photographed from *Die Russische Avantgarde und die Buhne 1890-1930*
2. theatre scene design, also from (above)
3. dance Erica Russel, Wiedermann (2004)
4. MPEG-4 – bits rates, from Koenen (2000).
5. MPEG-4 – scene positioning, from Koenen (2000).
6. MPEG-4 – up and downstream data, from Koenen (2000).
7. MPEG-4 – left: scene graph; right: sprites, from Koenen (2000).
8. MPEG-4 – syntax, from Koenen (2000).
9. MIT Media Lab⁷² web site.
10. student work – *multimedia authoring I*, dutch windmill.
11. student work – *multimedia authoring I*, Schröder house.
12. student work – *multimedia authoring I*, train station.
13. animation – Joan Gratch, from Wiedermann (2004).
14. animation – Joan Gratch, from Wiedermann (2004).
15. animation – Joan Gratch, from Wiedermann (2004).
16. animation – Joan Gratch, from Wiedermann (2004).
17. Agneta and Frieda example.
18. signs – people, van Rooijen (2003), p. 246, 247.

Both the costume designs and theatre scene designs of the russian avantgarde movement are *expressionist* in nature. Yet, they show humanity and are in their own way very humorous. The dance animation by Erica Russell, using basic shapes and rhythms to express the movement of dance, is to some extent both solemn and equally humorous. The animations by Joan Gratch use *morphing*, to transform wellknown artworks into other equally wellknown artworks.

⁷⁰www.adobe.com/tutorials

⁷¹www.w3c.org/AudioVideo

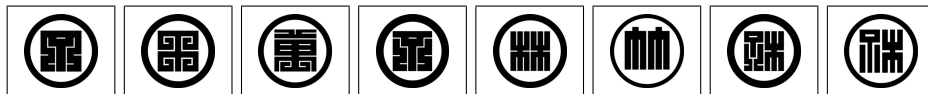
⁷²medai.mit.edu

part iii. multimedia information retrieval

.. my history might well be your future ...
ted nelson

chapters:

- 5. information retrieval
- 6. content annotation
- 7. information system architecture



2

reading directives In the following chapters we will discuss how we can make the various media formats, including text, images, audio and video amenable to search, either by analyzing content or by providing explicit meta information. For video, in particular, we develop a simple annotation logic that captures both the story line and the actors, that is persons and objects, that figure in it.

Essential sections are section 5.1, that characterizes scenarios for information retrieval, section 5.3, that introduces standard information retrieval concepts stemming from text search, section 6.4, that defines the aforementioned annotation logic, and section 7.2, that gives an outline of an abstract multimedia data format.

Section 6.3 is rather technical and may safely be skipped. Also sections 5.2, 6.1 and 7.3 may be skipped on first reading.

perspectives Apart from the many technical issues in information retrieval, perhaps the human interaction issues are the most urgent. As possible perspectives to look at these issues, consider:

perspectives – multimedia information retrieval

- application(s) – digital dossier
- psychological – focus
- experimental – user interaction
- algorithmic – (information) access
- system – unified presentation space
- presentation – embodied agents
- search – semantic annotation
- commercial – future systems

As you will see in the *research directions* given for each section, there are many proposals to improve interaction, for example the use of 3D virtual environments as an alternative way of presenting information.

essay topics For further study you may want to look at algorithms for analyzing content, annotation schemes for particular application domains, or the presentation issues mentioned before. Possible essay titles are:

- searching the web – searching for images, video and sound
- finding a tune – mobile music search services

Since the retrieval problem seems to be rather intractable in a general fashion, you should limit your discussion to a specific domain, for example retrieval in the domain of cultural heritage, and relate technical issues to the requirements of users in that particular domain.



3

the artwork

1. kata – japanese martial arts picture.
2. signs – japanese coats of arms, van Rooijen (2003), p. 140, 141.
3. photographs – Jaap Stahlie⁷³, two early experiments (left, and right)

⁷³www.jaapstahlie.com

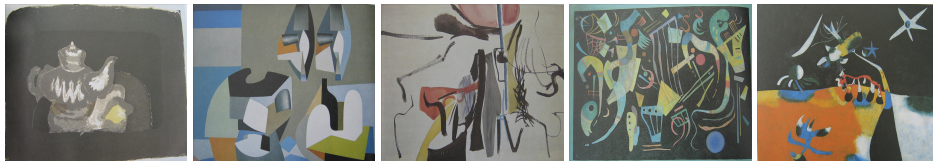
5. information retrieval

information retrieval is usually an afterthought

learning objectives

After reading this chapter you should be able to describe scenarios for information retrieval, to explain how content analysis for images can be done, to characterize similarity metrics, to define the notions of recall and precision, and to give an example of frequency tables, as used in text search.

Searching for information on the web is cumbersome. Given our experiences today, we may not even want to think about searching for multimedia information on the (multimedia) web. Nevertheless, in this chapter we will briefly sketch one of the possible scenarios indicating the need for multimedia search. In fact, once we have the ability to search for multimedia information, many scenarios could be thought of. As a start, we will look at two media types, images and documents. We will study search for images, because it teaches us important lessons about content analysis of media objects and what we may consider as *being similar*. Perhaps surprisingly, we will study text documents because, due to our familiarity with this media type, text documents allow us to determine what we may understand by effective search.



1

5.1 scenarios

Multimedia is not only for entertainment. Many human activities, for example medical diagnosis or scientific research, make use of multimedia information. To get an idea about what is involved in multimedia information retrieval look at the following scenario, adapted from Subrahmanian (1998),

Amsterdam Drugport

Amsterdam is an international centre of traffic and trade. It is renowned for its culture and liberal attitude, and attracts tourists from various ages, including young tourists that are attracted by the availability of soft drugs. Soft drugs may be obtained at so-called coffeeshops, and the possession of limited amounts of soft drugs is being tolerated by the authorities.

The European Community, however, has expressed their concern that Amsterdam is the centre of an international criminal drug operation. Combining national and international police units, a team is formed to start an exhaustive investigation, under the code name Amsterdam Drugport.

Now, without bothering ourselves with all the logistics of such an operation, we may establish what sorts of information will be gathered during the investigation, and what support for (multimedia) storage and (multimedia) information retrieval must be available.

Information can come from a variety of sources. Some types of information may be gathered continuously, for example by video cameras monitoring parking lots, or banks. Some information is already available, for example photographs in a (legacy database) police archive. Also of relevance may be information about financial transactions, as stored in the database of a bank, or geographic information, to get insight in possible drug traffic routes.

From a perspective of information storage our information (data) include the following media types: images, from photos; video, from surveillance; audio, from interviews and phone tracks; documents, from forensic research and reports; handwriting, from notes and sketches; and structured data, from for example bank transactions.

We have to find a way to store all these data by developing a suitable multimedia information system architecture, as discussed in chapter 6. More importantly, however, we must provide access to the data (or the information space, if you will) so that the actual police investigation is effectively supported. So, what kind of queries can we expect? For example, to find out more about a murder which seems to be related to the drugs operation.

retrieval

- *image query* – all images with this person
- *audio query* – identity of speaker
- *text query* – all transactions with BANK Inc.
- *video query* – all segments with victim
- *complex queries* – convicted murderers with BANK transactions
- *heterogeneous queries* – photograph + murderer + transaction
- *complex heterogeneous queries* – in contact with + murderer + transaction

Apparently, we might have simple queries on each of the media types, for example to detect the identity of a voice on a telephone wiretap. But we may also have more complex queries, establishing for example the likelihood that a murderer known by the police is involved, or even *heterogeneous* queries (as they are called in Subrahmanian (1998)), that establish a relation between information coming

from multiple information sources. An example of the latter could be, *did the person on this photo have any transactions with that bank in the last three months*, or more complex, *give me all the persons that have been in contact with the victim (as recorded on audio phonetaps, photographs, and video surveillance tapes) that have had transactions with that particular bank*.

I believe you'll have the picture by now. So what we are about to do is to investigate how querying on this variety of media types, that is images, text, audio and video, might be realized.



research directions– *information retrieval models*

Information retrieval research has quite a long history, with a focus on indexing text and developing efficient search algorithms. Nowadays, partly due to the wide-spread use of the web, research in information retrieval includes modeling, classification and clustering, system architectures, user interfaces, information visualisation, filtering, descriptive languages, etcetera. See Baeza-Yates and Ribeiro-Neto (1999).

Information retrieval, according to Baeza-Yates and Ribeiro-Neto (1999), deals with the representation, storage, organisation of, and access to information items. To see what is involved, imagine that we have a (user) query like:

find me the pages containing information on ...

Then the goal of the information retrieval system is to retrieve information that is useful or relevant to the user, in other words: *information that satisfies the user's information need*.

Given an information repository, which may consist of web pages but also multimedia objects, the information retrieval system must extract syntactic and semantic information from these (information) items and use this to match the user's information need.

Effective information retrieval is determined by, on the one hand, the *user task* and, on the other hand, the *logical view* of the documents or media objects that constitute the information repository. As user tasks, we may distinguish between *retrieval* (by query) and *browsing* (by navigation). To obtain the relevant information in retrieval we generally apply *filtering*, which may also be regarded as a ranking based on the attributes considered most relevant.

The logical view of text documents generally amounts to a set of index terms characterizing the document. To find relevant index terms, we may apply operations to the document, such as the elimination of stop words or text stemming. As you may easily see, full text provides the most complete logical view, whereas a small set of categories provides the most concise logical view. Generally, the user task will determine whether semantic richness or efficiency of search will be considered as more important when deciding on the obvious tradeoffs involved.

information retrieval models In Baeza-Yates and Ribeiro-Neto (1999), a great variety of information retrieval models is described. For your understanding, an information retrieval model makes explicit how index terms are represented and how the index terms characterizing an information item are matched with a query.

When we limit ourselves to the classic models for search and filtering, we may distinguish between:

information retrieval models

- boolean or set-theoretic models
- vector or algebraic models
- probabilistic models

Boolean models typically allow for *yes/no* answers only. They have a set-theoretic basis, and include models based on fuzzy logic, which allow for somewhat more refined answers.

Vector models use algebraic operations on vectors of attribute terms to determine possible matches. The attributes that make up a vector must in principle be orthogonal. Attributes may be given a weight, or even be ignored. Much research has been done on how to find an optimal selection of attributes for a given information repository.

Probabilistic models include general inference networks, and belief networks based on Bayesian logic.

Although it is somewhat premature to compare these models with respect to their effectiveness in actual information retrieval tasks, there is, according to Baeza-Yates and Ribeiro-Neto (1999), a general consensus that vector models

will outperform the probabilistic models on general collections of text documents. How they will perform for arbitrary collections of multimedia objects might be an altogether different question!

Nevertheless, in the sections to follow we will focus primarily on generalized vector representations of multimedia objects. So, let's conclude with listing the advantages of vector models.

vector models

- attribute term weighting scheme improves performance
- partial matching strategy allows retrieval of approximate material
- metric distance allows for sorting according to degree of similarity

Reading the following sections, you will come to understand how to adopt an attribute weighting scheme, how to apply partial matching and how to define a suitable distance metric.

So, let me finish with posing a research issue: *How can you improve a particular information retrieval model or matching scheme by using a suitable method of knowledge representation and reasoning?* To give you a point of departure, look at the logic-based multimedia information retrieval system proposed in Fuhr et al. (1998).

5.2 images

An image may tell you more than 1000 words. Well, whether images are indeed a more powerful medium of expression is an issue I'd rather leave aside. The problem how to get information out of an image, or more generally how to query image databases is, in the context of our *Amsterdam Drugport* operation more relevant. There are two issues here

- obtaining descriptive information
- establishing similarity

These issues are quite distinct, although descriptive information may be used to establish similarity.

descriptive information

When we want to find, for example, all images that contain a person with say sunglasses, we need to have of the images in our database that includes this information one way or another. One way would be to annotate all images with (meta) information and describe the objects in the picture to some degree of detail. More challenging would be to extract image content by image analysis, and produce the description (semi) automatically.

According to Subrahmanian (1998), content-based description of images involves the identification of objects, as well as an indication of where these objects are located in the image, by using a *shape descriptor* and possibly *property descriptors* indicating the pictorial properties of a particular region of the object or image.

Shape and property descriptors may take a form as indicated below.

shape

- bounding box – (XLB,XUB,YLB,YUB)

property

- property – name=value

As an example of applying these descriptors.

example

shape descriptor: XLB=10; XUB=60; YLB=3; YUB=50
 property descriptor: pixel(14,7): R=5; G=1; B=3

Now, instead of taking raw pixels as the unit of analysis, we may subdivide an image in a grid of cells and establish properties of cells, by some suitable algorithm.

definitions

- image grid: $(m * n)$ cells of equal size
- cell property: (Name, Value, Method)

As an example, we can define a property that indicates whether a particular cell is black or white.

example

property: (bwcolor,{b,w},bwalgo)

The actual algorithm used to establish such a property might be a matter of choice. So, in the example it is given as an explicit parameter.

From here to automatic content description is, admittedly, still a long way. We will indicate some research directions at the end of this section.



similarity-based retrieval

We need not necessarily know what an image (or segment of it) depicts to establish whether there are other images that contain that same thing, or something similar to it. We may, following Subrahmanian (1998), formulate the problem of similarity-based retrieval as follows:

How do we determine whether the content of a segment (of a segmented image) is similar to another image (or set of images)?

Think of, for example, the problem of finding all photos that match a particular face.

According to Subrahmanian (1998), there are two solutions:

- *metric approach* – distance between two image objects
- *transformation approach* – relative to specification

As we will see later, the transformation approach in some way subsumes the metric approach, since we can formulate a distance measure for the transformation approach as well.

metric approach What does it mean when we say, the distance between two images is less than the distance between this image and that one. What we want to express is that the first two images (or faces) are more alike, or maybe even identical.

Abstractly, something is a distance measure if it satisfies certain criteria.

metric approach

distance $d : X \rightarrow [0, 1]$ is distance measure if:

$$\begin{aligned} d(x,y) &= d(y,x) \\ d(x,y) &\leq d(x,z) + d(z,y) \\ d(x,x) &= 0 \end{aligned}$$

For your intuition, it is enough when you limit yourself to what you are familiar with, that is measuring distance in ordinary (Euclidian) space.

Now, in measuring the distance between two images, or segments of images, we may go back to the level of pixels, and establish a distance metric on pixel properties, by comparing all properties pixel-wise and establishing a distance.

pixel properties

- objects with pixel properties p_1, \dots, p_n
- pixels: (x, y, v_1, \dots, v_n)
- object contains $w \times h$ $(n+2)$ -tuples

Leaving the details for your further research, it is not hard to see that even if the absolute value of a distance has no meaning, relative distances do. So, when an image contains a face with dark sunglasses, it will be closer to (an image of) a face with dark sunglasses than a face without sunglasses, other things being

equal. It is also not hard to see that a pixel-wise approach is, computationally, quite complex. An object is considered as

complexity

a set of points in k -dimensional space for $k = n + 2$

In other words, to establish similarity between two images (that is, calculate the distance) requires $n+2$ times the number of pixels comparisons.

feature extraction Obviously, we can do better than that by restricting ourselves to a pre-defined set of properties or features.

feature extraction

- maps object into s -dimensional space

For example, one of the features could indicate whether or not it was a face with dark sunglasses. So, instead of calculating the distance by establishing color differences of between regions of the images where sunglasses may be found, we may limit ourselves to considering a binary value, yes or no, to see whether the face has sunglasses.

Once we have determined a suitable set of features that allow us to establish similarity between images, we no longer need to store the images themselves, and can build an index based on feature vectors only, that is the combined value on the selected properties.

Feature vectors and extensive comparison are not exclusive, and may be combined to get more precise results. Whatever way we choose, when we present an image we may search in our image database and present all those objects that fall within a suitable *similarity range*, that is the images (or segments of images) that are close enough according to the distance metric we have chosen.



4

transformation approach Instead of measuring the distance between two images (objects) directly, we can take one image and start modifying that until it exactly equals the target image. In other words, as phrased in Subrahmanian (1998), the principle underlying the transformation approach is:

transformation approach

Given two objects o1 and o2, the level of dissimilarity is proportional to the (minimum) cost of transforming object o1 into object o2 or vice versa

Now, this principle might be applied to any representation of an object or image, including feature vectors. Yet, on the level of images, we may think of the following operations:

to_1, \dots, to_r – translation, rotation, scaling

Moreover, we can attach a cost to each of these operations and calculate the cost of a transformation sequence TS by summing the costs of the individual operations. Based on the cost function we can define a distance metric, which we call for obvious reasons the *edit distance*, to establish similarity between objects.

cost

- $cost(TS) = \sum_{i=1}^r cost(to_i)$

distance

- $d(o, o') = \min \{ cost(TS) \mid TS \text{ in } TSeq(o, o') \}$

An obvious advantage of the *edit distance* over the pixel-wise distance metric is that we may have a rich choice of transformation operators that we can attach (user-defined) cost to at will.

For example, we could define low costs for normalization operations, such as scaling and rotation, and attach more weight to operations that modify color values or add shapes. For face recognition, for example, we could attribute low cost to adding sunglasses but high cost to changing the sex.

To support the *transformation approach* at the image level, our image database needs to include suitable operations. See Subrahmanian (1998).

operations

```
rotate(image-id, dir, angle)
segment(image-id, predicate)
edit(image-id, edit-op)
```

We might even think of storing images, not as a collection of pixels, but as a sequence of operations on any one of a given set of base images. This is not such a strange idea as it may seem. For example, to store information about faces we may take a base collection of prototype faces and define an individual face by selecting a suitable prototype and a limited number of operations or additional properties.



5

example(s) – *match of the day*

The images in this section present a *match of the day*, which is part of the project *split representation* by the Dutch media artist Geert Mul. As explain in the email sending the images, about once a week, *Television images are recorded at random from satellite television and compared with each other. Some 1000.000.000 (one billion) equations are done every day.*

. The *split representation* project uses the image analyses and image composition software *NOTATION*⁷⁴, which was developed by Geert Mul (concept) and Carlo Preize (programming & software design).

research directions – *multimedia repositories*

What would be the proper format to store multimedia information? In other words, what is the shape multimedia repositories should take? Some of the issues involved are discussed in chapter , which deals with information system architectures. With respect to image repositories, we may rephrase the question into *what support must an image repository provide, minimally, to allow for efficient access and search?*. In Subrahmanian (1998), we find the following answer:

image repository

- *storage* – unsegmented images
- *description* – limited set of features
- *index* – feature-based index
- *retrieval* – distance between feature vectors

And, indeed, this seems to be what most image databases provide. Note that the actual encoding is not of importance. The same type of information can be encoded using either XML, relational tables or object databases. What is of importance is the functionality that is offered to the user, in terms of storage and retrieval as well as presentation facilities.

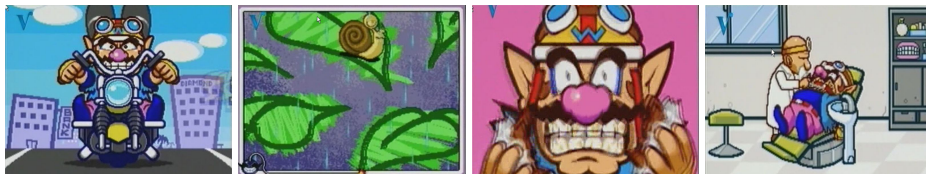
⁷⁴homepage.mac.com/geertmul2

What is the relation between presentation facilities and the functionality of multimedia repositories? Consider the following mission statement, which is taken from my research and projects page.

mission

Our goal is to study aspects of the deployment and architecture of virtual environments as an interface to (intelligent) multimedia information systems <black>...

Obviously, the underlying multimedia repository must provide adequate retrieval facilities and must also be able to deliver the desired objects in a format suitable for the representation and possibly incorporation in such an environment. Actually, at this stage, I have only some vague ideas about how to make this vision come through. Look, however, at chapter and appendix ?? for some initial ideas.



6

5.3 documents

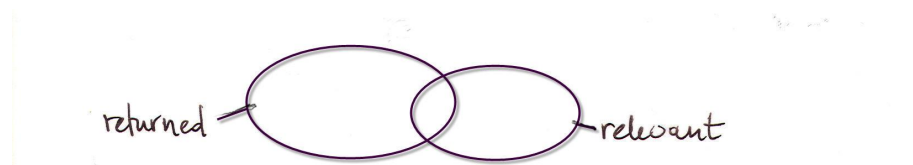
Even in the presence of audiovisual media, text will remain an important vehicle for human communication. In this section, we will look at the issues that arise in querying a text or document database. First we will characterize more precisely what we mean by effective search, and then we will study techniques to realize effective search for document databases.

Basically, answering a query to a document database comes down to string matching. However, some problems may occur such as synonymy and polysemy.

problems

- synonymy – topic T does not occur literally in document D
- polysemy – some words may have many meanings

As an example, *church* and *house of prayer* have more or less the same meaning. So documents about churches and cathedrals should be returned when you ask for information about 'houses of prayer'. As an example of polysemy, think of the word *drum*, which has quite a different meaning when taken from a musical perspective than from a transport logistics perspective.



precision and recall

Suppose that, when you pose a query, everything that is in the database is returned. You would probably not be satisfied, although every relevant document will be included, that is for sure. On the other hand, when nothing is returned, at least you cannot complain about non-relevant documents that are returned, or can you?

In Subrahmanian (1998), the notions of *precision* and *recall* are proposed to measure the effectiveness of search over a document database. In general, precision and recall can be defined as follows.

effective search

- precision – how many answers are correct
- recall – how many of the right documents are returned

For your intuition, just imagine that you have a database of documents. With full knowledge of the database you can delineate a set of documents that are of relevance to a particular query. Also, you can delineate a set that will be returned by some given search algorithm. Then, *precision* is the intersection of the two sets in relation to what the search algorithm returns, and *recall* that same intersection in relation to what is relevant. In pseudo-formulas, we can express this as follows:

precision and recall

$$\begin{aligned}\text{precision} &= (\text{returned and relevant}) / \text{returned} \\ \text{recall} &= (\text{returned and relevant}) / \text{relevant}\end{aligned}$$

Now, as indicated in the beginning, it is not too difficult to get either perfect recall (by returning all documents) or perfect precision (by returning almost nothing). But these must be considered anomalies (that is, sick cases), and so the problem is to find an algorithm that performs optimally with respect to both precision and recall.

For the total database we can extend these measures by taking the averages of precision and recall for all topics that the database may be queried about.

Can these measures only be applied to document databases? Of course not, these are general measures that can be applied to search over any media type!

frequency tables

A *frequency table* is an example of a way to improve search. Frequency tables, as discussed in Subrahmanian (1998), are useful for documents only. Let's look at an example first.

example

term/document	d0	d1	d2
snacks	1	0	0
drinks	1	0	3
rock-roll	0	1	1

Basically, what a frequency table does is, as the name implies, give a frequency count for particular words or phrases for a number of documents. In effect, a

complete document database may be summarized in a frequency table. In other words, the frequency table may be considered as an index to facilitate the search for similar documents.

To find a similar document, we can simply make a word frequency count for the query, and compare that with the columns in the table. As with images, we can apply a simple distance metric to find the nearest (matching) documents. (In effect, we may take the square root for the sum of the squared differences between the entries in the frequency count as our distance measure.)

The complexity of this algorithm may be characterized as follows:

complexity

compare term frequencies per document – $O(M*N)$

where M is the number of terms and N is the number of documents. Since both M and N can become very large we need to make an effort to reduce the size of the frequency table.

reduction

- stop list – irrelevant words
- word stems – reduce different words to relevant part

We can, for example, introduce a *stop list* to prevent irrelevant words to enter the table, and we may restrict ourselves to including *word stems* only, to bring back multiple entries to one canonical form. With some additional effort we could even deal with synonymy and polysemy by introducing, respectively equivalence classes, and alternatives (although we then need a suitable way for ambiguation). By the way, did you notice that frequency tables may be regarded as feature vectors for documents?



research directions– *user-oriented measures*

Even though the reductions proposed may result in limiting the size of the frequency tables, we may still be faced with frequency tables of considerable size. One way to reduce the size further, as discussed in Subrahmanian (1998), is

to apply *latent semantic indexing* which comes down to clustering the document database, and limiting ourselves to the most relevant words only, where relevance is determined by the ratio of occurrence over the total number of words. In effect, the less the word occurs, the more discriminating it might be. Alternatively, the choice of what words are considered relevant may be determined by taking into account the area of application or the interest of a particular group of users.



8

user-oriented measures Observe that, when evaluating a particular information retrieval system, the notions of precision and recall as introduced before are rather system-oriented measures, based on the assumption of a user-independent notion of relevance. However, as stated in Baeza-Yates and Ribeiro-Neto (1999), different users might have a different interpretation on which document is relevant. In Baeza-Yates and Ribeiro-Neto (1999), some user-oriented measures are briefly discussed, that to some extent cope with this problem.

user-oriented measures

- *coverage ratio* – fraction of known documents
- *novelty ratio* – fraction of new (relevant) documents
- *relative recall* – fraction of expected documents
- *recall effort* – fraction of examined documents

Consider a reference collection, an example information request and a retrieval strategy to be evaluated. Then the *coverage ratio* may be defined as the fraction of the documents known to be relevant, or more precisely the number of (known) relevant documents retrieved divided by the total number of documents known to be relevant by the user.

The *novelty ratio* may then be defined as the fraction of the documents retrieved which were not known to be relevant by the user, or more precisely the number of relevant documents that were not known by the user divided by the total number of relevant documents retrieved.

The *relative recall* is obtained by dividing the number of relevant documents found by the number of relevant documents the user expected to be found.

Finally, *recall effort* may be characterized as the ratio of the number of relevant documents expected and the total number of documents that has to be examined to retrieve these documents.

Notice that these measures all have a clearly 'subjective' element, in that, although they may be generalized to a particular group of users, they will very likely not generalize to all groups of users. In effect, this may lead to different retrieval strategies for different categories of users, taking into account level of expertise and familiarity with the information repository.

5.4 development(s) – inconvenient truth(s)

Over the last couple of years, climate change has come into the focus of public attention. Moved by television images of dislocated people in far-away countries, ice bears threatened by the corruption of their native environment, tsunami waves flooding the third world, and hurricanes destroying urban areas, the general public is becoming worried by what Al Gore has so aptly characterized as *an inconvenient truth*: the climate is changing and human affluence may be the prime cause.

In response to the *pathos* of the media, many civil groups do an appeal on the responsibility of individual citizens and start campaigns for an *ethos* of climate-correct behavior, by saving on energy-consumption or driving CO2-friendly cars. In the media, such campaigns are either advocated or criticized by authorities from public government, and experts from a multitude of sciences, with conflicting opinions. As a result, the general audience, initially with genuine concern about the state of our world, gets confused and loses interest. And more worrisome, the adolescents, looking at the serious way adults express their confusion and ignorance, take distance and may decide that the *climate issue* is not of their concern.

At the Climate Centre of the VU University Amsterdam, we are not happy to observe that *pathos* and *ethos* overtake the public debate, and we actively wish to participate in the public debate bringing our multi-disciplinary scientific background into play. Moreover, since we *borrow the earth from our children*, as the old Indian saying goes, which Al Gore again brought to our attention, we feel that we must take an active interest in bringing the *climate issue* to the attention of the youth, in a form that is appropriate. From this background, we engaged in developing Clima Futura, a multi-disciplinary undertaking, bringing together climate experts from a variety of backgrounds with multimedia/game development researchers. The Clima Futura game addresses the issues of climate change, not altogether without *pathos* nor *ethos*, but nevertheless primarily focussed on bringing the *logos* of climate change into the foreground, in other words the scientific issues that are at play, and the science-based insights and uncertainties that may govern our decisions in the political debate. Given the state of our knowledge, the science of climate change itself may be characterized as an inconvenient science, and as such an interesting challenge to present by means of a game.



questions

information retrieval

1. (*) What is meant by the *complementarity of authoring and retrieval*? Sketch a possible scenario of (multimedia) information retrieval and indicate how this may be implemented. Discuss the issues that arise in accessing multimedia information and how content annotation may be deployed.

concepts

2. How would you approach *content-based description* of images?
3. What is the difference between a *metric* approach and the *transformational* approach to establishing similarity between images?
4. What problems may occur when searching in text or document databases?

technology

5. Give a definition of: *shape descriptor* and *property descriptor*. Give an example of each.
6. How would you define *edit distance*?
7. Characterize the notions *precision* and *recall*.
8. Give an example (with explanation) of a *frequency table*.

projects & further reading As a project, you may implement simple image analysis algorithms that, for example, extract a color histogram, or detect the presence of a horizon-like edge.

You may further explore scenarios for information retrieval in the cultural heritage domain, and compare this with other applications of multimedia information retrieval, for example monitoring in hospitals.

For further reading I suggest to make yourself familiar with common techniques in information retrieval as described in Baeza-Yates and Ribeiro-Neto (1999), and perhaps devote some time to studying image analysis, Gonzales and Wintz (1987).

the artwork

1. artworks – ..., Miro, Dali, photographed from Kunstsammlung Nordrhein-Westfalen, see artwork 2.
2. left Miro from Kunst, right: Karel Appel
3. *match of the day* (1) – Geert Mul
4. *match of the day* (2) – Geert Mul
5. *match of the day* (3) – Geert Mul
6. *mario ware* – taken from gammo/veronica⁷⁵.
7. *baten kaitos – eternal ways and the lost ocean*, taken from gammo/veronica.
8. idem.
9. signs – people, van Rooijen (2003), p. 252, 253.

⁷⁵www.gammo.nl

The art opening this chapter belongs to the tradition of 20th century art. It is playful, experimental, with strong existential implications, and it shows an amazing *variety of styles*.

The examples of *match of the day* by Geert Mul serve to illustrate the interplay between *technology and art*, and may also start you to think about what *similarity* is. Some illustrations from games are added to show the difference in styles.

6. content annotation

video annotation requires a logical approach to story telling

learning objectives

After reading this chapter you should be able to explain the difference between content and meta information, to mention relevant content parameters for audio, to characterize the requirements for video libraries, to define an annotation logic for video, and to discuss feature extraction in samples of musical material.

Current technology does not allow us to extract information automatically from arbitrary media objects. In these cases, at least for the time being, we need to assist search by annotating content with what is commonly referred to as meta-information. In this chapter, we will look at two more media types, in particular audio and video. Studying audio, we will learn how we may combine feature extraction and meta-information to define a data model that allows for search. Studying video, on the other hand, will indicate the complexity of devising a knowledge representation scheme that captures the content of video fragments. Concluding this chapter, we will discuss an architecture for feature extraction for arbitrary media objects.



1

6.1 audio

The audio media type covers both spoken voice and musical material. In this section we will discuss audio signal, stored in a raw or compressed (digital) format, as well as similarity-based retrieval for musical patterns.

In general, for providing search access to audio material we need, following Subrahmanian (1998), a data model that allows for both meta-data (that is information about the media object) and additional attributes of features, that we in principle obtain from the media object itself, using feature extraction.

audio data model

- *meta-data* – describing content
- *features* – using feature extraction

As an example of audi meta-data, consider the (meta-data) characterization that may be given for opera librettos.

example

singers – (Opera,Role,Person)
score – ...
transcript – ...

For signal-based audio content, we have to perform an analysis of the audio signal for which we may take parameters such as frequency, velocity and amplitude. For the actual analysis we may have to break up the signal in small windows, along the time-axis. Using feature extraction, we may characterize (signal-based) properties such as indicated below.

feature extraction

- *intensity* – watts/ m^2
- *loudness* – in decibels
- *pitch* – from frequency and amplitude
- *brightness* – amount of distortion

For a more detailed treatment of signal-based audio content description, consult Subrahmanian (1998).

In the following we will first give an overview of musical search facilities on the web and then we will discuss similarity-based retrieval of musical patterns in somewhat more depth in the section on *research directions*. In section 6.3, we will have a closer look at feature extraction for arbitrary media types.



musical similarity

Although intuitively obvious, how can we characterize musical similarity? And perhaps more importantly, how can we compute the extent to which one piece of music or a melody line is similar to another piece of music or melody line. As concerns musical content, at least for most genres, it appears that

According to Selfridge (1998), we should focus primarily on *melody*, since

"It is melody that makes music memorable: we are likely to recall a tune long after we have forgotten its text."

Other features, content-based as well as descriptive, may however be used as additional filters in the process of retrieval.

Melodic searching and matching has been explored mainly in the context of bibliographic tools and for the analysis of (monophonic) repertoires Hewlett and Selfridge-Field (1998). As described in section , many of these efforts have been made available to the general public through the Web. Challenges for the near future are, however, to provide for melodic similarity matching on polyphonic works, and retrieval over very large databases of musical fragments.

In this section we will look in somewhat more detail at the problem of melodic similarity matching. In particular, we will discuss representational issues, matching algorithms and additional analysis tools that may be used for musical information retrieval.

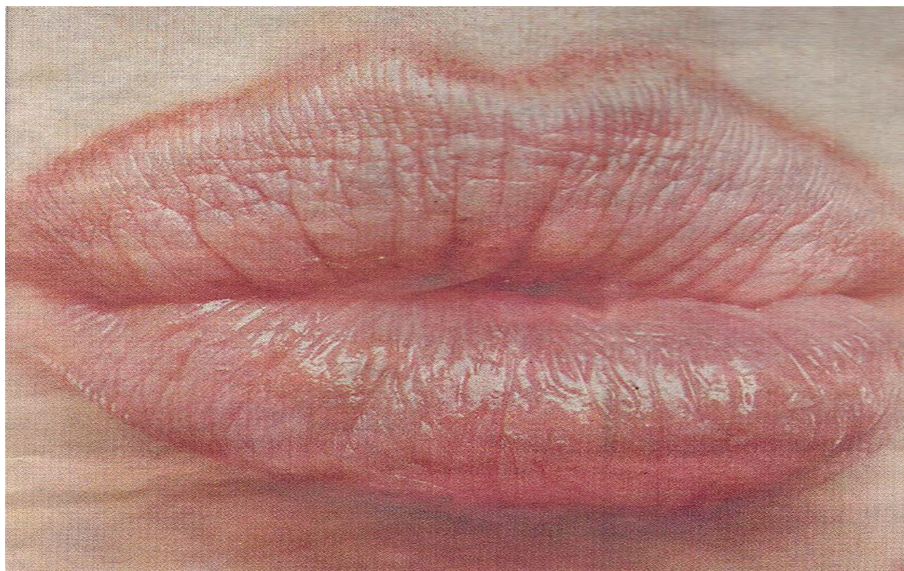
melodic similarity Consider the musical fragment *Twinkle, twinkle little star* (known in the Dutch tradition as "*Altijd is Kortjakje ziek*"), which has been used by Mozart for a series of variations Mozart (1787). Now, imagine how you would approach establishing the similarity between the original theme and these variations. As a matter of fact, we discovered that exactly this problem had been tackled in the study reported in Mongeau and Sankoff (1990), which we will discuss later. Before that, we may reflect on what we mean by the concept of a *melody*. In the aforementioned variations the original melody is disguised by, for example, decorations and accompaniments. In some variations, the melody is distributed among the various parts (the left and right hand). In other variations, the melody is only implied by the harmonic structure. Nevertheless, for the human ear there seems to be, as it is called in Selfridge (1998), a '*prototypical*' melody that is present in each of the variations.

When we restrict ourselves to pitch-based comparisons, melodic similarity may be established by comparing profiles of pitch-direction (up, down, repeat) or pitch contours (which may be depicted graphically). Also, given a suitable representation, we may compare pitch-event strings (assuming a normalized pitch representation such as position within a scale) or intervallic contours (which gives the distance between notes in for example semitones). Following Selfridge (1998), we may observe however that the more general the system of representation, the longer the (query) *string* will need to be to produce meaningful discriminations. As further discussed in Selfridge (1998), recent studies in musical perception indicate that pitch-information without durational values does not suffice.

representational issues Given a set of musical fragments, we may envisage several reductions to arrive at the (hypothetical) prototypical melody. Such reductions must provide for the elimination of confounds such as rests, repeated notes and grace notes, and result in, for example, a pitch-string (in a suitable representation), a duration profile, and (possibly) accented note profiles and harmonic reinforcement profiles (which capture notes that are emphasized by harmonic changes). Unfortunately, as observed in Selfridge (1998), the problem of which reductions to apply is rather elusive, since it depends to a great extent on the goals of the query and the repertory at hand.

As concerns the representation of pitch information, there is a choice between a base-7 representation, which corresponds with the position relative to the tonic in the major or minor scales, a base-12 representation, which corresponds with a division in twelve semitones as in the chromatic scale, and more elaborate encodings, which also reflect notational differences in identical notes that arise through the use of accidentals. For MIDI applications, a base-12 notation is most suitable, since the MIDI note information is given in semitone steps. In addition to relative pitch information, octave information is also important, to establish the rising and falling of melodic contour.

When we restrict ourselves to directional profiles (up, down, repeat), we may include information concerning the slope, or degree of change, the relation of the current pitch to the original pitch, possible repetitions, recurrence of pitches after intervening pitches, and possible segmentations in the melody. In addition, however, to support relevant comparisons it seems important to have information on the rhythmic and harmonic structure as well.



example(s) – *napster*

Wasn't it always your dream to have all your music free? Napster⁷⁶ was the answer. (But not for long.) Napster is, as we learn in the Wikipedia⁷⁷, *an online music service which was originally a file sharing service created by Shawn Fanning. Napster was the first widely-used peer-to-peer music sharing service, and it made a major impact on how people, especially college students, used the Internet. Its technology allowed music fans to easily share MP3 format song files with each other, thus leading to the music industry's accusations of massive copyright violations. The service was named Napster after Fanning's nickname.* However, Napster has been forced to become commercial. So the question is: is there life after napster? Well, there is at least open source!

research directions – *musical similarity matching*

An altogether different approach at establishing melodic similarity is proposed in Mongeau and Sankoff (1990). This approach has been followed in the Meldex system McNab et al. (1997), discussed in section . This is a rather technical section, that may be skipped on first reading. The approach is different in that it relies on a (computer science) theory of finite sequence comparison, instead of musical considerations. The general approach is, as explained in Mongeau and Sankoff (1990), to search for an optimal correspondence between elements of two sequences, based on a distance metric or measure of dissimilarity, also known more informally as the *edit-distance*, which amounts to the (minimal) number of transformations that need to be applied to the first sequence in order to obtain the second one. Typical transformations include *deletion*, *insertion* and *replacement*. In the musical domain, we may also apply transformations such as *consolidation* (the replacement of several elements by one element) and *fragmentation* (which is the reverse of consolidation). The metric is even more generally applicable by associating a weight with each of the transformations. Elements of the musical sequences used in Mongeau and Sankoff (1990) are pitch-duration pairs, encoded in base-12 pitch information and durations as multiples of 1/16th notes.

The matching algorithm can be summarized by the following recurrence relation for the dissimilarity metric. Given two sequences $A = a_1, \dots, a_m$ and $B = b_1, \dots, b_n$ and $d_{ij} = d(a_i, b_j)$, we define the distance as

$$d_{ij} = \min \begin{cases} d_{i-1,j} + w(a_i, 0) & \text{deletion} \\ d_{i,j-1} + w(0, b_j) & \text{insertion} \\ d_{i-1,j-1} + w(a_i, b_j) & \text{replacement} \\ d_{i-k,j-1} + w(a_{i-k+1}, \dots, a_i, b_j). \quad 2 \leq k \leq i & \text{consolidation} \\ d_{i-1,j-k+1} + w(a_{-i}, b_{-j-k+1}, \dots, b_{-j}) \quad 2 \leq k \leq j & \text{fragmentation} \end{cases}$$

with

$$d_{i0} = d_{i-1,0} + w(a_i, 0), \quad i \geq 1 \quad \text{deletion}$$

⁷⁶www.napster.com

⁷⁷en.wikipedia.org/wiki/Napster

$$d_{0j} = d_{0,j-1} + w(0, b_i), j \geq 1$$

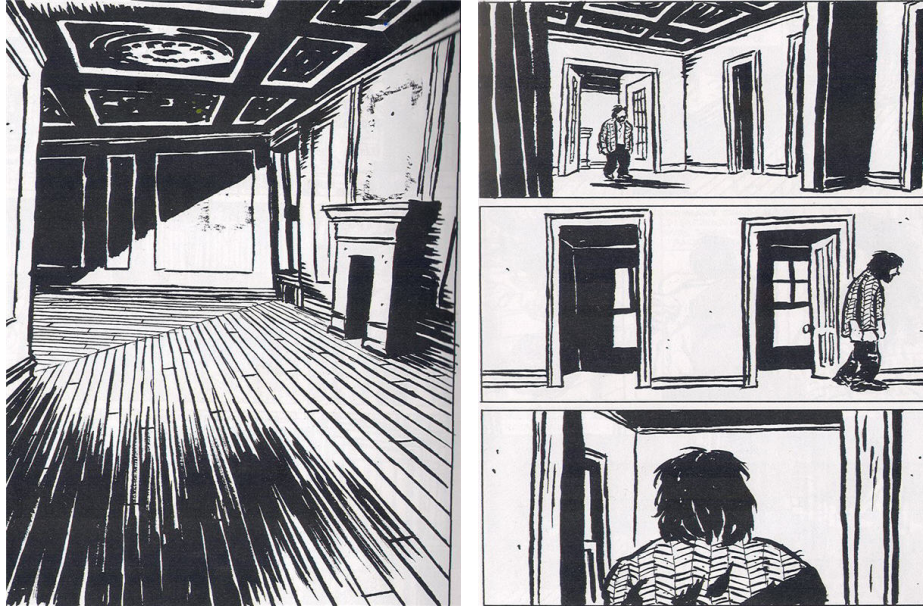
insertion

and $d_{00} = 0$. The weights $w(-, -)$ are determined by the degree of dissonance and the length of the notes involved.

The actual algorithms for determining the dissimilarity between two sequences uses dynamic programming techniques. The algorithm has been generalized to look for matching phrases, or subsequences, within a sequence. The complexity of the algorithm is $O(mn)$, provided that a limit is imposed on the number of notes involved in consolidation and fragmentation.

Nevertheless, as indicated in experiments for the Meldex database, the resulting complexity is still forbidding when large databases are involved. The Meldex system offers apart from the (approximate) dynamic programming algorithm also a state matching algorithm that is less flexible, but significantly faster. The Meldex experiments involved a database of 9400 songs, that were used to investigate six musical search criteria: (1) exact interval and rhythm, (2) exact contour and rhythm, (3) exact interval, (4) exact contour, (5) approximate interval and rhythm, and (6) approximate contour and rhythm. Their results indicate that the number of notes needed to return a reasonable number of songs scales logarithmically with database size McNab et al. (1997). It must be noted that the Meldex database contained a full (monophonic) transcription of the songs. An obvious solution to manage the complexity of searching over a large database would seem to be the storage of prototypical themes or melodies instead of complete songs.

indexing and analysis There are several tools available that may assist us in creating a proper index of musical information. One of these tools is the Humdrum system, which offers facilities for metric and harmonic analysis, that have proven their worth in several musicological investigations Huron (1997). Another tool that seems to be suitable for our purposes, moreover since it uses a simple pitch-duration, or *piano-roll*, encoding of musical material, is the system for metric and harmonic analysis described in Temperley and Sleator (1999). Their system derives a metrical structure, encoded as hierarchical levels of equally spaced beats, based on preference-rules which determine the overall likelihood of the resulting metrical structure. Harmonic analysis further results in (another level of) *chord spans* labelled with roots, which is also determined by preference rules that take into account the previously derived metrical structure. As we have observed before, metrical and harmonic analysis may be used to eliminate confounding information with regard to the 'prototypical' melodic structure.



4

6.2 video

Automatic content description is no doubt much harder for video than for any other media type. Given the current state of the art, it is not realistic to expect content description by feature extraction for video to be feasible. Therefore, to realize content-based search for video, we have to rely on some knowledge representation schema that may adequately describe the (dynamic) properties of video fragments.

In fact, the description of video content may reflect the story-board, that after all is intended to capture both time-independent and dynamically changing properties of the objects (and persons) that play a role in the video.

In developing a suitable annotation for a particular video fragment, two questions need to be answered:

video annotation

- what are the interesting aspects?
- how do we represent this information?

Which aspects are of interest is something you have to decide for yourself. Let's see whether we can define a suitable knowledge representation scheme.

One possible knowledge representation scheme for annotating video content is proposed in Subrahmanian (1998). The scheme proposed has been inspired by knowledge representation techniques in Artificial Intelligence. It captures both static and dynamic properties.

video content

video v , frame f
 f has associated objects and activities
 objects and activities have properties

First of all, we must be able to talk about a particular video fragment v , and frame f that occurs in it. Each frame may contain objects that play a role in some activity. Both objects and activities may have properties, that is attributes that have some value.

property

property: name = value

As we will see in the examples, properties may also be characterized using predicates.

Some properties depend on the actual frame the object is in. Other properties (for example sex and age) are not likely to change and may be considered to be frame-independent.

object schema

(fd,fi) – frame-dependent and frame-independent properties

Finally, in order to identify objects we need an object identifier for each object. Summing up, for each object in a video fragment we can define an *object instance*, that characterizes both frame-independent and frame-dependent properties of the object.

object instance: (oid,os,ip)

- *object-id* – oid
- *object-schema* – os = (fd,fi)
- *set of statements* – ip: name = v and name = v IN f

Now, with a collection of object instances we can characterize the contents of an entire video fragment, by identifying the frame-dependent and frame-independent properties of the objects.

Look at the following example, borrowed from Subrahmanian (1998) for the *Amsterdam Drugport* scenario.

frame	objects	<i>frame-dependent properties</i>
1	Jane	has(briefcase), at(path)
-	house	door(closed)
-	briefcase	
2	Jane	has(briefcase), at(door)
-	Dennis	at(door)
-	house	door(open)
-	briefcase	

In the first frame Jane is near the house, at the path that leads to the door. The door is closed. In the next frame, the door is open. Jane is at the door, holding a briefcase. Dennis is also at the door. What will happen next?

Observe that we are using predicates to represent the state of affairs. We do this, simply because the predicate form *has(briefcase)* looks more natural than the other form, which would be *has = briefcase*. There is no essential difference between the two forms.

Now, to complete our description we can simply list the frame-independent properties, as illustrated below.

object	frame-independent properties	value
Jane	age	35
	height	170cm
house	address	...
	color	brown
briefcase	color	black
	size	40 x 31

How to go from the tabular format to sets of statements that comprise the object schemas is left as an (easy) exercise for the student.

Let's go back to our *Amsterdam Drugport* scenario and see what this information might do for us, in finding possible suspects. Based on the information given in the example, we can determine that there is a person with a briefcase, and another person to which that briefcase may possibly be handed. Whether this is the case or not should be disclosed in frame 3. Now, what we are actually looking for is the possible exchange of a briefcase, which may indicate a drug transaction. So why not, following Subrahmanian (1998), introduce another somewhat more abstract level of description that deals with *activities*.

activity

- activity name – id
- statements – *role* = *v*

An activity has a name, and consists further simply of a set of statements describing the *roles* that take part in the activity.

example

```
{ giver : Person, receiver : Person, item : Object }
giver = Jane, receiver = Dennis, object = briefcase
```

For example, an *exchange* activity may be characterized by identifying the *giver*, *receiver* and *object* roles. So, instead of looking for persons and objects in a video fragment, you'd better look for activities that may have taken place, by finding a matching set of objects for the particular roles of an activity. Consult Subrahmanian (1998) if you are interested in a further formalization of these notions.



5

video libraries

Assuming a knowledge representation scheme as the one treated above, how can we support search over a collection of videos or video fragments in a video library.

What we are interested in may roughly be summarized as

video libraries

- which videos are in the library
- what constitutes the content of each video
- what is the location of a particular video

Take note that all the information about the videos or video fragments must be provided as meta-information by a (human) librarian. Just imagine for a moment how laborious and painstaking this must be, and what a relief video feature extraction would be for an operation like *Amsterdam Drugport*.

To query the collection of video fragments, we need a query language with access to our knowledge representation. It must support a variety of retrieval operations, including the retrieval of segments, objects and activities, and also property-based retrievals as indicated below.

query language for video libraries

- *segment retrievals* – exchange of briefcase
- *object retrievals* – all people in v:[s,e]
- *activity retrieval* – all activities in v:[s,e]
- *property-based* – find all videos with object oid

Subrahmanian (1998) lists a collection of video functions that may be used to extend SQL into what we may call VideoSQL. Abstractly, VideoSQL may be characterized by the following schema:

VideoSQL

```

SELECT - v:[s,e]
FROM - video:<source><V>
WHERE - term IN funcall

```

where $v:[s,e]$ denotes the fragment of video v , starting at frame s and ending at frame e , and *term IN funcall* one of the video functions giving access to the information about that particular video. As an example, look at the following VideoSQL snippet:

example

```

SELECT vid:[s,e]
FROM video:VidLib
WHERE (vid,s,e) IN VideoWithObject(Dennis) AND
      object IN ObjectsInVideo(vid,s,e) AND
      object != Dennis AND
      typeof(object) = Person

```

Notice that apart from calling video functions also constraints can be added with respect to the identity and type of the objects involved.



example(s) – *video retrieval evaluation*

The goal of the TREC⁷⁸ conference series is to encourage research in information retrieval by providing a large test collection, uniform scoring procedures, and a forum for organizations interested in comparing their results. Since 2003 there is an independent *video* track devoted to research in automatic segmentation,

⁷⁸trec.nist.gov

indexing, and content-based retrieval of digital video. In the TRECVID⁷⁹ 2004 workshop, thirty-three teams from Europe, the Americas, Asia, and Australia participated. Check it out!



7

research directions— *presentation and context*

Let's consider an example. Suppose you have a database with (video) fragments of news and documentary items. How would you give access to that database? And, how would you present its contents? Naturally, to answer the first question, you need to provide search facilities. Now, with regard to the second question, for a small database, of say 100 items, you could present a list of videos that matches the query. But with a database of over 10,000 items this will become problematic, not to speak about databases with over a million of video fragments. For large databases, obviously, you need some way of visualizing the results, so that the user can quickly browse through the candidate set(s) of items.

Pesce (2003) provide an interesting account on how *interactive maps* may be used to improve search and discovery in a (digital) video library. As they explain in the abstract:

To improve library access, the Informedia Digital Video Library uses automatic processing to derive descriptors for video. A new extension to the video processing extracts geographic references from these descriptors.

The operational library interface shows the geographic entities addressed in a story, highlighting the regions discussed in the video through a map display synchronized with the video display.

⁷⁹www-nlpir.nist.gov/projects/trecvid

So, the idea is to use geographical information (that is somehow available in the video fragments themselves) as an additional descriptor, and to use that information to enhance the presentation of a particular video. For presenting the results of a query, candidate items may be displayed as icons in a particular region on a map, so that the user can make a choice.

Obviously, having such geographical information:

The map can also serve as a query mechanism, allowing users to search the terabyte library for stories taking place in a selected area of interest.

The approach to extracting descriptors for video fragments is interesting in itself. The two primary sources of information are, respectively, the spoken text and graphic text overlays (which are common in news items to emphasize particular aspects of the news, such as the area where an accident occurs). Both speech recognition and image processing are needed to extract information terms, and in addition natural language processing, to do the actual 'geocoding', that is translating this information to geographical locations related to the story in the video.

Leaving technical details aside, it will be evident that this approach works since news items may relevantly be grouped and accessed from a geographical perspective. For this type of information we may search, in other words, with three kinds of questions:

- *what* – content-related
- *when* – position on time-continuum
- *where* – geographic location

and we may, evidently, use the geographic location both as a search criterium and to enhance the presentation of query results.

mapping information spaces Now, can we generalize this approach to other type of items as well. More specifically, can we use maps or some spatial layout to display the results of a query in a meaningful way and so give better access to large databases of multimedia objects. According to Dodge and Kitchin (2002), we are very likely able to do so:

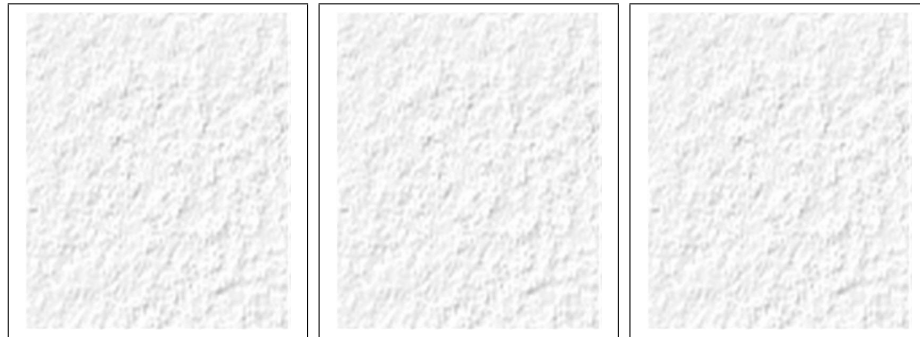
More recently, it has been recognized that the process of spatialization – where a spatial map-like structure is applied to data where no inherent or obvious one does exist – can provide an interpretable structure to other types of data.

Actually, we are taking up the theme of *visualization*, again. In Dodge and Kitchin (2002) visualizations are presented that (together) may be regarded as an *atlas of cyberspace*.

atlas of cyberspace

We present a wide range of spatializations that have employed a variety of graphical techniques and visual metaphors so as to provide striking and powerful images that extend from two dimension 'maps' to three-dimensional immersive landscapes.

As you may gather from chapter 7 and the *afterthoughts*, I take a personal interest in the (research) theme of *virtual reality interfaces for multimedia information systems*. But I am well aware of the difficulties involved. It is an area that is just beginning to be explored!



8

6.3 feature extraction

Manual content annotation is laborious, and hence costly. As a consequence, content annotation will often not be done and search access to multimedia object will not be optimal, if it is provided for at all. An alternative to manual content annotation is (semi) automatic feature extraction, which allows for obtaining a description of a particular media object using media specific analysis techniques.

The Multimedia Database Research group at CWI has developed a framework for feature extraction to support the *Amsterdam Catalogue of Images* (ACOI). The resulting framework for feature extraction is known as the ACOI framework, Kersten et al. (1998).

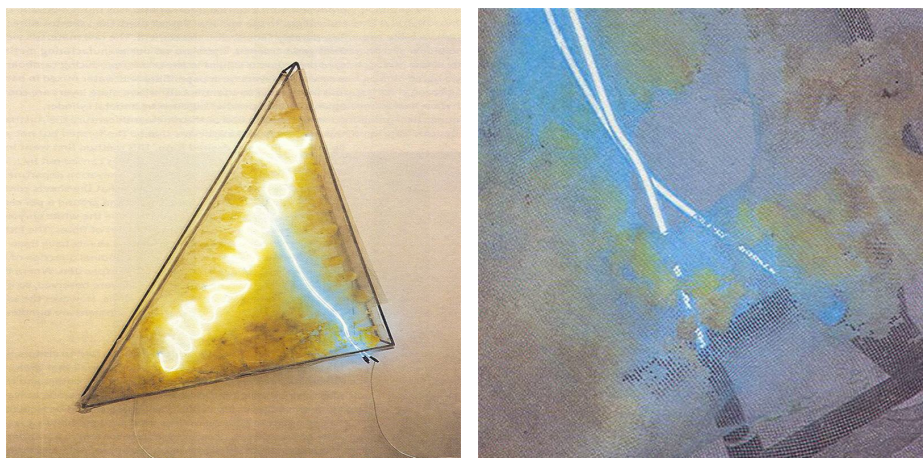
The ACOI framework is intended to accommodate a broad spectrum of classification schemes, manual as well as (semi) automatic, for the indexing and retrieval of arbitrary multimedia objects. What is stored are not the actual multimedia objects themselves, but structural descriptions of these objects (including their location) that may be used for retrieval.

The ACOI model is based on the assumption that indexing an arbitrary multimedia object is equivalent to deriving a grammatical structure that provides a namespace to reason about the object and to access its components. However there is an important difference with ordinary parsing in that the lexical and grammatical items corresponding to the components of the multimedia object must be created dynamically by inspecting the actual object. Moreover, in general, there is not a fixed sequence of lexicals as in the case of natural or formal languages. To allow for the dynamic creation of lexical and grammatical items the ACOI framework supports both *black-box* and *white-box* (feature) detectors. Black-box detectors are algorithms, usually developed by a specialist in the media domain,

that extract properties from the media object by some form of analysis. White-box detectors, on the other hand, are created by defining logical or mathematical expressions over the grammar itself. Here we will focus on black-box detectors only.

The information obtained from parsing a multimedia object is stored in a database. The feature grammar and its associated detector further result in updating the data schemas stored in the database.

formal specification Formally, a feature grammar G may be defined as $G = (V, T, P, S)$, where V is a collection of variables or non-terminals, T a collection of terminals, P a collection of productions of the form $V \rightarrow (V \cup T)$ and S a start symbol. A token sequence ts belongs to the language $L(G)$ if $S \xrightarrow{*} ts$. Sentential token sequences, those belonging to $L(G)$ or its sublanguages $L(G_v) = (V_v, T_v, P_v, v)$ for $v \in (T \cup V)$, correspond to a complex object C_v , which is the object corresponding to the parse tree for v . The parse tree defines a hierarchical structure that may be used to access and manipulate the components of the multimedia object subjected to the detector. See Schmidt et al. (1999) for further details.



anatomy of a feature detector

As an example of a feature detector, we will look at a simple feature detector for (MIDI encoded) musical data. A special feature of this particular detector, that I developed while being a guest at CWI, is that it uses an intermediate representation in a logic programming language (Prolog) to facilitate reasoning about features.

The hierarchical information structure that we consider is defined in the grammar below. It contains only a limited number of basic properties and must be

extended with information along the lines of some musical ontology, see Zimmerman (1998).

feature grammar

```

detector song; # # to get the filename
detector lyrics; # # extracts lyrics
detector melody; # # extracts melody
detector check; # # to walk the tree

atom str name;
atom str text;
atom str note;

midi: song;

song: file lyrics melody check;

file: name;

lyrics: text*;
melody: note*;

```

The start symbol is a *song*. The detector that is associated with *song* reads in a MIDI file. The musical information contained in the MIDI file is then stored as a collection of Prolog facts. This translation is very direct. In effect the MIDI file header information is stored, and events are recorded as facts, as illustrated below for a *note_on* and *note_off* event.

```

event('twinkle',2,time=384, note_on:[chan=2,pitch=72,vol=111]).
event('twinkle',2,time=768, note_off:[chan=2,pitch=72,vol=100]).

```

After translating the MIDI file into a Prolog format, the other detectors will be invoked, that is the *composer*, *lyrics* and *melody* detector, to extract the information related to these properties.

To extract relevant fragments of the melody we use the melody detector, of which a partial listing is given below.

melody detector

```

int melodyDetector(tree *pt, list *tks ){
char buf[1024]; char* _result;
void* q = _query;
int idq = 0;

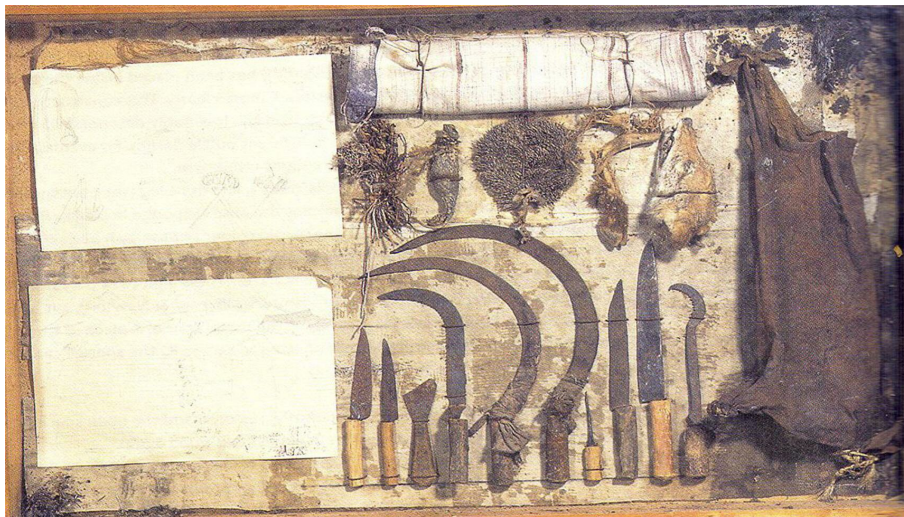
idq = query_eval(q,"X:melody(X)");
while (( _result = query_result(q,idq) ) ) {
    putAtom(tks,"note",_result);
}
}

```

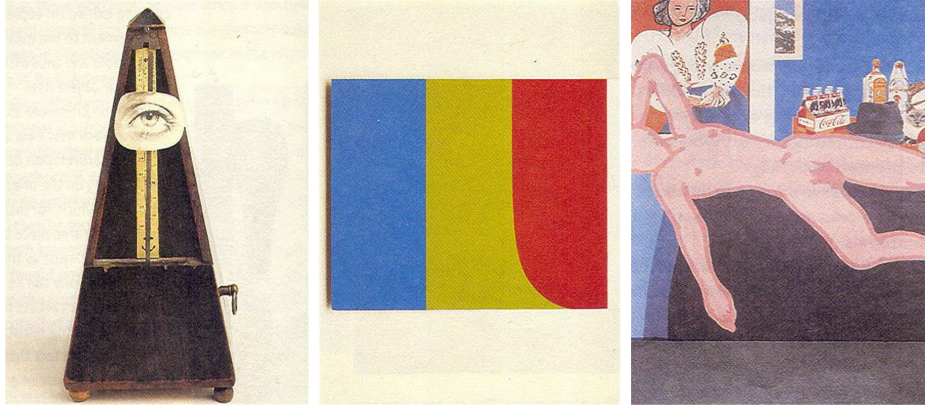
```
    return SUCCESS;  
}
```

The embedded logic component is given the query `X:melody(X)`, which results in the notes that constitute the (relevant fragment of the) melody. These notes are then added to the tokenstream. A similar detector is available for the lyrics.

Parsing a given MIDI file, for example *twinkle.mid*, results in updating the database.



implementation The embedded logic component is part of the *hush* framework, Eliens (2000). It uses an object extension of Prolog that allows for the definition of native objects to interface with the MIDI processing software written in C++. The logic component allows for the definition of arbitrary predicates to extract the musical information, such as the melody and the lyrics. It also allows for further analysis of these features to check for, for example, particular patterns in the melody.



11

example(s) – *modern art: who cares?*

The artworks shown above are taken from Hummelen and Sillé (1999), which bundles the experiences and insights resulting from studying the preservation of contemporary art, under the title: *modern art, who cares?* This project was a precursor to the INCCA⁸⁰ that provided the input to our *multimedia casus*, which is introduced in chapter 10.

Both the INCCA project and the related *Open Archives Initiative*⁸¹, focus on making meta-information available on existing resources for the preservation of contemporary art and cultural heritage in general, including reports, case studies and recordings of artworks, that is images, videos and artists interviews.

12

research directions– *media search*

There is a wealth of powerful search engines on the Web. Technically, search engines rely either on classification schemes (as for example Yahoo) or content-based (keyword) indexing (as for example Excite or AltaVista). Searching on the Web, nowadays, is moderately effective when text-based documents are considered. For multimedia objects (such as images or music) existing search facilities are far less effective, simply because indexing on category or keywords can not so easily be done automatically. In the following we will explore what search facilities there are for music (on the web). We will first give some examples of search based on keywords and categories, then some examples of content-based search and finally we will discuss a more exhaustive list of musical databases and search facilities on the Web. All search facilities mentioned are listed online under *musical resources*.

⁸⁰ www.incca.org

⁸¹ www.openarchives.org

keywords and categories For musical material, in particular MIDI, there are a number of sites that offer search over a body of collected works. One example is the Aria Database, that allows to search for an aria part of an opera based on title, category and even voice part. Another example is the MIDI Farm, which provides many MIDI-related resources, and also allows for searching for MIDI material by filename, author, artist and ratings. A category can be selected to limit the search. The MIDI Farm employs voting to achieve collaborative filtering on the results for a query. Search indexes for sites based on categories and keywords are usually created by hand, sometimes erroneously. For example, when searching for a Twinkle fragment, Bach's variations for Twinkle were found, whereas to the best of our knowledge there exist only Twinkle variations by Mozart Mozart (1787). The Digital Tradition Folksong Database provides in addition a powerful lyrics (free text) search facility based on the AskSam search engine. An alternative way of searching is to employ a meta-search engine. Meta-search engines assist the user in formulating an appropriate query, while leaving the actual search to (possibly multiple) search engines. Searching for musical content is generally restricted to the lyrics, but see below (and section).

content-based search Although content-based search for images and sound have been a topic of interest for over a decade, few results have been made available to the public. As an example, the MuscleFish Datablade for Informix, allows for obtaining information from audio based on a content analysis of the audio object. As far as content-based musical search facilities for the Web are concerned, we have for example, the Meldex system of the New Zealand Digital Library initiative, an experimental system that allows for searching tunes in a folksong database with approximately 1000 records, McNab et al. (1997). Querying facilities for Meldex include queries based on transcriptions from audio input, that is humming a tune! We will discuss the approach taken for the Meldex system in more detail in research directions section, to assess its viability for retrieving musical fragments in a large database.

music databases In addition to the sites previously mentioned, there exist several databases with musical information on the Web. We observe that these databases do not rely on DBMS technology at all. This obviously leads to a plethora of file formats and re-invention of typical DBMS facilities. Without aiming for completeness, we have for example the *MIDI Universe*, which offers over a million MIDI file references, indexed primarily by composer and file length. It moreover keeps relevant statistics on popular tunes, as well as a hot set of MIDI tunes. It further offers access to a list of related smaller MIDI databases. Another example is the aforementioned Meldex system that offers a large collection of tunes (more than 100.000), of which a part is accessible by humming-based retrieval. In addition text-based search is possible against file names, song titles, track names and (where available) lyrics. The Classical MIDI Archive is an example of a database allowing text-based search on titles only. Results are annotated with an indication of "goodness" and recency. The Classical Themefinder Database allows extensive support for retrieval based on (optional) indications of meter,

pitch, pitch-class, interval, semi-tone interval and melodic contour, within a fixed collection of works arranged according to composer and category. The index is clearly created and maintained manually. The resulting work is delivered in the MuseData format, which is a rich (research-based) file format from which MIDI files can be generated, Selfridge (1997). A site which collects librarian information concerning music resources is the International Inventory of Music Resources (RISM), which offers search facilities over bibliographic records for music manuscripts, librettos and secondary sources for music written after c.a. 1600. It also allows to search for libraries related to the RISM site. Tune recognition is apparently offered by the Tune Server. The user may search by offering a WAV file with a fragment of the melody. However, the actual matching occurs against a melodic outline, that is indications of rising or falling in pitch. The database contains approx. 15.000 records with such pitch contours, of which one third are popular tunes and the rest classical themes. The output is a ranked list of titles about which the user is asked to give feedback.

discussion There is great divergence in the scope and aims of music databases on the Web. Some, such as the RISM database, are the result of musicological investigations, whereas others, such as the MIDI Farm, are meant to serve an audience looking for popular tunes. With regard to the actual search facilities offered, we observe that, with the exception of Meldex and the Tune Server, the query facilities are usually text-based, although for example the Classical Themefinder allows for encoding melodic contour in a text-based fashion.

6.4 development(s) – expert recommendations

SERIAL RECOMMENDER MODEL

Admittedly not the best way to do research, although common practice, we found a good starting point for modelling recommender systems, by googling on *serial recommender*, in a paper from Microsoft Research on privacy in distributed recommender systems, Privacy. The model introduced in Privacy, distinguishes between:

$U = user$
 $I = item$
 $B = behavior$
 $R = recommendation$
 $F = feature$

and allows for characterizing observations (from which implicit ratings can be derived) and recommendations, as follows:

- observations – $U \times I \times B$
- recommendations – $U \times I$

In a centralized approach the mapping $U \times I \times B \rightarrow U \times I$ provides recommendations from observations, either directly by applying the $U \times I \rightarrow I \times I$ mapping, or indirectly by the mapping $U \times I \rightarrow U \times U \rightarrow I \times I$, which uses an intermediate matrix (or product space) $U \times U$ indicating the (preference) relation between users or user-groups. Taken as a matrix, we may fill the entries with distance or weight values. Otherwise, when we use product spaces, we need to provide an additional mapping to the range of $[0, 1]$, where distance can be taken as the dual of weight, that is $d = 1 - w$.

In a decentralized approach, Privacy argue that it is better to use the actual features of the items, and proceed from a mapping $I \times F \rightarrow U \times I \times R$. Updating preferences is then a matter of applying a $I \times B \rightarrow I \times F$ mapping, by analyzing which features are considered important.

For example, observing that a user spends a particular amount of time and gives a rating r , we may apply this rating to all features of the item, which will indirectly influence the rating of items with similar features.

$$\begin{aligned} B &= [\text{time} = 20\text{sec}, \text{rating} = r] \\ F &= [\text{artist} = \text{rembrandt}, \text{topic} = \text{portrait}] \\ R &= [\text{artist}(\text{rembrandt}) = r, \text{topic}(\text{portrait}) = r] \end{aligned}$$

Privacy observe that B and R need not to be standardized, however F must be a common or shared feature space to allow for the generalization of the rating of particular items to similar items.

With reference to the CHIP project, mentioned in the previous section, we may model a collection of artworks by (partially) enumerating their properties, as indicated below:

$$\begin{aligned} A &= [p_1, p_2, \dots] \\ \text{where } p_k &= [f_1 = v_1, f_2 = v_2, \dots] \end{aligned}$$

with as an example

$$\begin{aligned} A_{\text{nightwatch}} &= [\text{artist}=\text{rembrandt}, \text{topic}=\text{group}] \\ A_{\text{guernica}} &= [\text{artist}=\text{picasso}, \text{topic}=\text{group}] \end{aligned}$$

Then we can see how preferences may be shared among users, by taking into account the (preference) value adhered to artworks or individual properties, as illustrated in fig. 7.

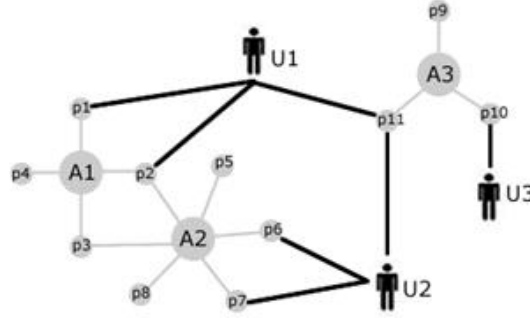


Fig 7. Users, artworks and properties

As a note, to avoid misunderstanding, Picasso's Guernica is not part of the collection of the Rijksmuseum, and does as such not figure in the CHIP studies. The example is taken, however, to clarify some properties of metrics on art collections, to be discussed in the next section.

CONTENT METRICS

To measure similarity, in information retrieval commonly a distance measure is used. In mathematical terms a distance function $d : X \rightarrow [0, 1]$ is distance measure if:

$$\begin{aligned} d(x, y) &= d(y, x) \\ d(x, y) &\leq d(x, z) + d(z, y) \\ d(x, x) &= 0 \end{aligned}$$

From an abstract perspective, measuring the distance between artworks, grouped according to some preference criterium, may give insight in along which dimension the grouping is done, or in other words what attributes have preference over others. When we consider the artworks

$$\begin{aligned} a_1 &= [\text{artist} = \text{rembrandt}, \text{topic} = \text{self-portrait}] \\ a_2 &= [\text{artist} = \text{rembrandt}, \text{name} = \text{nightwatch}] \\ a_3 &= [\text{artist} = \text{picasso}, \text{topic} = \text{self-portrait}] \\ a_4 &= [\text{artist} = \text{picasso}, \text{name} = \text{guernica}] \end{aligned}$$

we may, in an abstract fashion, deduce that if $d(a_1, a_2) < d(a_1, a_3)$ then $r(\text{topic}) < r(\text{artist})$, however if $d(a_1, a_3) < d(a_1, a_2)$ the reverse is true, that is then $r(\text{artist}) < r(\text{topic})$. Somehow, it seems unlikely that a_2 and a_4 will be grouped together, since even though their topic may be considered to be related, the aesthetic impact of these works is quite different, where *selfportraits* as a genre practiced over the centuries indeed seem to form a 'logical' category. Note that we may also express this as $w(\text{artist}) < w(\text{topic})$ if we choose to apply weights to existing ratings, and then use the observation that if $d(a_1, a_3) < d(a_1, a_2)$ then $w(\text{artist}) < w(\text{topic})$ to generate a guided tour in which a_3 precedes a_2 .

For serial recommenders, that provide the user with a sequence of items $\dots, s_{n-1}, s_n, \dots$, and for s_n possibly alternatives a_1, a_2, \dots , we may adapt the (implied) preference of the user, when the user chooses to select alternative a_k instead of accepting s_n as provided by the recommender, to adjust the weight of the items involved, or features thereof, by taking into account an additional constraint on the distance measure. Differently put, when we denote by $s_{n-1} \mapsto s_n/[a_1, a_2, \dots]$ the presentation of item s_n with as possible alternatives a_1, a_2, \dots , we know that $d(s_{n-1}, a_k) < d(s_{n-1}, s_n)$ for some k , if the user chooses for a_k . In other words, from observation B_n we can deduce R_n :

$$\begin{aligned} B_n &= [\text{time} = 20\text{sec}, \text{forward} = a_k] \\ F_n &= [\text{artist} = \text{rembrandt}, \text{topic} = \text{portrait}] \\ R_n &= [d(s_n, a_k) < d(s_n, s_{n+1})] \end{aligned}$$

leaving, at this moment, the feature vector F_n unaffected. Together, the collection of recommendations, or more properly revisions R_i over a sequence S , can be solved as a system of linear equations to adapt or revise the (original) ratings. Hence, we might be tempted to speak of the $R4$ framework, *rate*, *recommend*, *regret*, *revise*. However, we prefer to take into account the cyclic/incremental nature of recommending, which allows us to identify revision with rating.

MEASURES FOR FEEDBACK DISCREPANCY

So far, we have not indicated how to process user feedback, given during the presentation of a guided tour, which in the simple case merely consists of selecting a possible alternative. Before looking in more detail at how to process user feedback, let us consider the dimensions involved in the rating of items, determining the eventual recommendation of these or similar items. In outline, the dimensions involved in rating are:

- positive vs negative
- individual vs community/collaborative
- feature-based vs item-based

Surprisingly, in Shneiderman (1997) we found that negative ratings of artworks had no predictive value for an explicit rating of (preferences for) the categories and properties of artworks. Leaving the dimension *individual vs community/collaborative* aside, since this falls outside of the scope of this paper, we face the question of how to revise feature ratings on the basis of preferences stated for items, which occurs (implicitly) when the user selects an alternative for an item presented in a guided tour, from a finite collection of alternatives.

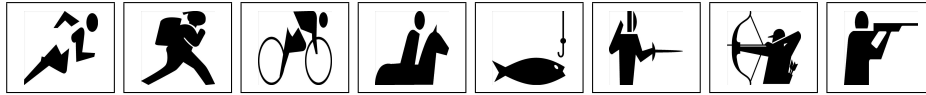
A very straightforward way is to ask explicitly what properties influence the decision. More precisely, we may ask the user why a particular alternative is selected, and let the user indicate what s/he likes about the selected alternative and dislikes about the item presented by the recommender. It is our expectation, which must however yet be verified, that negative preferences do have an impact on the explicit characterization of the (positive and negative) preferences for general artwork categories and properties, since presenting a guided tour, as an

organized collection of items, is in some sense more directly related to user goals (or educational targets) than the presentation of an unorganized collection of individual items. Cf. Hybrid.

So let's look at $s_{n-1} \mapsto s_n/[a_1, a_2, \dots]$ expressing alternative selection options a_1, a_2, \dots at s_n in sequence $S = \dots, s_{n-1}, s_n$. We may distinguish between the following interpretations, or revisions:

- neutral interpretation – use $d(s_n, a_k) < d(s_n, s_{n+1})$
- positive interpretation – increase $w(\text{feature}(a_k))$
- negative interpretation – decrease $w(\text{feature}(s_{n+1}))$

How to actually deal with the revision of weights for individual features is, again, beyond the scope of this paper. We refer however to Eliens (2000), where we used feature vectors to find (dis)similarity between musical fragments, and to Schmidt et al. (1999), on which our previous work was based, where a feature grammar is introduced that characterizes an object or item as a hierarchical structure, that may be used to access and manipulate the component-attributes of an item.



13

questions

content annotation

1. (*) How can video information be made accessible? Discuss the requirements for supporting video queries.

concepts

2. What are the ingredients of an *audio data model*
3. What information must be stored to enable search for video content?
4. What is *feature extraction*? Indicate how feature extraction can be deployed for arbitrary media formats.

technology

5. What are the parameters for *signal-based (audio) content*?
6. Give an example of the representation of *frame-dependent* en *frame-independent* properties of a video fragment.
7. What are the elements of a query language for searching in video libraries?
8. Give an example (with explanation) of the use of *VideoSQL*.

projects & further reading As a project, think of implementing musical similarity matching, or developing an application retrieving video fragments using a simple annotation logic.

You may further explore the construction of media repositories, and finding a balance between automatic indexing, content search and meta information.

For further reading I advice you to *google* recent research on video analysis, and the online material on search engines⁸².

the artwork

1. works from Weishar (1998)
2. faces – from www.alterfin.org, an interesting site with many surprising interactive toys in *flash*, javascript and html.
3. mouth – Annika Karlson Rixon, entitled *A slight Acquaintance*, taken from a theme article about the body in art and science, the Volkskrant, 24/03/05.
4. story – page from the comic book version of *City of Glass*, Auster (2004), drawn in an almost tradional style.
5. story – frame from Auster (2004).
6. story – frame from Auster (2004).
7. story – frame from Auster (2004).
8. *white on white* – typographical joke.
9. modern art – *city of light* (1968-69), Mario Merz, taken from Hummelen and Sillé (1999).
10. modern art – *Marocco* (1972), Krijn Griezen, taken from Hummelen and Sillé (1999).
11. modern art – *Indestructable Object* (1958), Man Ray, *Blue, Green, Red I* (1964-65), Ellsworth Kelly, *Great American Nude* (1960), T. Wesselman, taken from Hummelen and Sillé (1999).
12. signs – sports, van Rooijen (2003), p. 272, 273.

Opening this chapter are examples of design of the 20th century, posters to announce a public event like a theatre play, a world fair, or a festival. In comparison to the art works of the previous chapter, these designs are more strongly *expressive* and more simple and clear in their *message*. Yet, they also show a wide variety of styles and rethorics to attract the attention of the audience. Both the faces and the mouth are examples of using body parts in contemporary art. The page of the comic book version of *City of Glass*, illustrates how the 'logic' of a story can be visualised. As an exercise, try to annoyaye the sequence of frames from the *City of Glass* can be described using the annotation logic you learned in this chapter. The modern art examples should interesting by themselves.

⁸²www.searchtools.com/tools/tools-opensource.html

7. information system architecture

effective retrieval requires visual interfaces

learning objectives

After reading this chapter you should be able to discuss the considerations that play a role in developing a multimedia information system, characterize an abstract multimedia data format, give examples of multimedia content queries, define the notion of virtual resources, and discuss the requirements for networked virtual environments.

From a system development perspective, a multimedia information system may be considered as a multimedia database, providing storage and retrieval facilities for media objects. Yet, rather than a solution this presents us with a problem, since there are many options to provide such storage facilities and equally many to support retrieval. In this chapter, we will study the architectural issues involved in developing multimedia information systems, and we will introduce the notion of media abstraction to provide for a uniform approach to arbitrary media objects. Finally, we will discuss the additional problems that networked multimedia confront us with.



1

7.1 architectural issues

The notion of *multimedia information system* is sufficiently generic to allow for a variety of realizations. Let's have a look at the issues involved.

As concerns the database (that is the storage and retrieval facilities), we may have to deal with homegrown solution, commercial third party databases or (even) legacy sources. To make things worse, we will usually want to deploy a combination of these.

With respect to the information architecture, we may wish for a common format (which unifies the various media types), but in practice we will often have to work with the native formats or be satisfied with a hybrid information architecture that uses both media abstractions and native media types such as images and video.

The notion of media abstraction, introduced in Subrahmanian (1998), allows for uniform indexes over the multimedia information stored, and (as we will discuss in the next section) for query relaxation by employing hierarchical and equivalence relations.

Summarizing, for content organisation (which basically is the information architecture) we have the following options:

content organisation

- *autonomy* – index per media type
- *uniformity* – unified index
- *hybrid* – media indexes + unified index

In Subrahmanian (1998), a clear preference is stated for a uniform approach, as expressed in the *Principle of Uniformity*:

Principle of Uniformity

... from a semantical point of view the content of a multimedia source is independent of the source itself, so we may use statements as meta data to provide a description of media objects.

Naturally, there are some tradeoffs. In summary, Subrahmanian (1998) claims that: metadata can be stored using standard relational and OO structures, and that manipulating metadata is easy, and moreover that feature extraction is straightforward. Now consider, is feature extraction really so straightforward as suggested here? I would believe not. Certainly, media types can be processed and analysis algorithms can be executed. But will this result in meaningful annotations? Given the current state of the art, hardly so!

research directions – *the information retrieval cycle*

When considering an information system, we may proceed from a simple generic software architecture, consisting of:

software architecture

- a database of media object, supporting
- operations on media objects, and offering
- logical views on media objects

However, such a database-centered notion of information system seems not to do justice to the actual support and information system must provide when considering the full information retrieval cycle:

information retrieval cycle

1. specification of the user's information need
2. translation into query operations
3. search and retrieval of media objects
4. ranking according to likelihood or relevance
5. presentation of results and user feedback
6. resulting in a possibly modified query

When we look at older day information retrieval applications in libraries, we see more or less the automation of card catalogs, with search functionality for keywords and headings. Modern day versions of these systems, however, offer graphical userinterfaces, electronic forms and hypertext features.

When we look at the web and how it may support digital libraries, we see some dramatic changes with respect to the card catalogue type of applications. We can now have access to a variety of sources of information, at low cost, including geographically distributed resources, due to improved networking. And, everybody is free to make information available, and what is worse, everybody seems to be doing so. Hence, the web is a continuously growing repository of information of a (very) heterogeneous kind.

Considering the web as an information retrieval system we may observe, following Baeza-Yates and Ribeiro-Neto (1999), that:

- despite high interactivity, access is difficult;
- quick response is and will remain important!

So, we need better (user-centered) retrieval strategies to support the full information retrieval cycle. Let me (again) mention some of the relevant (research) topics: *user interfaces, information visualisation, user-profiling and navigation.*



7.2 media abstractions

Let's have a closer look at media abstractions. How can we capture the characterization of a variety of media types in one common media abstraction. A definition of such a media abstraction is proposed in Subrahmanian (1998). Leaving the formal details aside, a media abstraction has the following components:

media abstraction

- *state* – smallest chunk of media data
- *feature* – any object in a state
- *attributes* – characteristics of objects
- *feature extraction map* – to identify content
- *relations* – to capture state-dependent information
- (inter)relations between 'states' or chunks

Now, that characterization is sufficiently abstract, and you may wonder how on earth to apply this to an actual media database.

However, before giving some examples, we must note that the *feature extraction map* does not need to provide information about the content of a chunk of media data automatically. It may well be a hand-coded annotation.

Our first example is an image database.

example – image database

states: { pic1.gif,...,picn.gif }
features: names of people
extraction: find people in pictures
relations: left-of, ...

In an image database it does not make much sense to speak about relations between 'states' or chunks of media data, that is the images.

For our next example though, video databases, it does make sense to speak about such relations, since it allows us to talk about scenes as sequences of frames.

example – video database

states: set of frames
features: persons and objects
extraction: gives features per frame
relations: frame-dependent and frame-independent information
inter-state relation: specifies sequences of frames

Now, with this definition of media abstractions, we can define a simple multimedia database, simply as

simple multimedia database

- a finite set M of media abstractions

But, following Subrahmanian (1998), we can do better than that. In order to deal with the problems of *synonymy* and *inheritance*, we can define a structured multimedia database that supports:

structured multimedia database

- *equivalence relations* – to deal with synonymy
- *partial ordering* – to deal with inheritance
- *query relaxation* – to please the user

Recall that we have discussed the relation between a 'house of prayer' and 'church' as an example of synonymy in section 4.3. As an example of inheritance we may think of the relation between 'church' and 'cathedral'. Naturally, every cathedral is a church. But the reverse does not necessarily hold. Having this information about possible equivalence and inheritance relationships, we can relax queries in order to obtain better results. For example, when a user asks for cathedral in a particular region, we could even notify the user of the fact that although there are no cathedrals there, there are a number of churches that may be of interest. (For a mathematical characterization of structured multimedia databases, study Subrahmanian (1998).)



3

query languages Having media abstractions, what would a query language for such a database look like? Again, following Subrahmanian (1998), we may extend SQL with special functions as indicated below:

SMDS – functions

Type: object \mapsto type
 ObjectWithFeatures: $f \mapsto \{o \mid \text{object } o \text{ contains } f\}$
 ObjectWithFeaturesAndAttributes: $(f, a, v) \mapsto \{o \mid o \text{ contains } f \text{ with } a = v\}$
 FeaturesInObject: $o \mapsto \{f \mid o \text{ contains } f\}$
 FeaturesAndAttributesInObject: $o \mapsto \{(f, a, v) \mid o \text{ contains } f \text{ with } a = v\}$

Having such functions we can characterize an extension of SQL, which has been dubbed SMDS-SQL in Subrahmanian (1998), as follows.

SMDS-SQL

SELECT – media entities

- m – if m is not a continuous media object

- $m : [i, j] - m$ is continuous, i, j integers (segments)
- $m.a - m$ is media entity, a is attribute

FROM

- $\langle \text{media} \rangle \langle \text{source} \rangle \langle M \rangle$

WHERE

- term IN funcall

As an example, look at the following SMDS-SQL snippet.

example

```
SELECT M
FROM   smds source1 M
WHERE  Type(M) = Image AND
       M IN ObjectWithFeature("Dennis") AND
       M IN ObjectWithFeature("Jane") AND
       left("Jane", "Dennis", M)
```

Note that M is a relation in the image database media abstraction, which contains one or more images that depict Jane to the left of Dennis. Now, did they exchange the briefcase, or did they not?

When we do not have a uniform representation, but a hybrid representation for our multimedia data instead, we need to be able to: express queries in specialized language, and to perform operations (joins) between SMDS and non-SMDS data.

Our variant of SQL, dubbed HM-SQL, differs from SMDS-SQL in two respects: function calls are annotated with media source, and queries to non-SMDS data may be embedded.

As a final example, look at the following snippet:

example HM-SQL

```
SELECT M
FROM   smds video1, videodb video2
WHERE  M IN smds:ObjectWithFeature("Dennis") AND
       M IN videodb:VideoWithObject("Dennis")
```

In this example, we are collecting all video fragments with Dennis in it, irrespective of where that fragment comes from, an (smds) database or another (video) database.

research directions— *digital libraries*

Where *media abstractions*, as discussed above, are meant to be technical abstractions needed for uniform access to media items, we need quite a different set of abstraction to cope with one of the major applications of multimedia information storage and retrieval: digital libraries.

According to Baeza-Yates and Ribeiro-Neto (1999), digital libraries will need a long time to evolve, not only because there are many technical hurdles to be

overcome, but also because effective digital libraries are dependent on an active community of users:

digital libraries

Digital libraries are constructed – collected and organized – by a community of users. Their functional capabilities support the information needs and users of this community. Digital libraries are an extension, enhancement and integration of a variety of information institutions as physical places where resources are selected, collected, organized, preserved and accessed in support of a user community.

The occurrence of digital libraries on the web is partly a response to advances in technology, and partly due to an increased appreciation of the facilities the internet can provide. From a development perspective, digital libraries may be regarded as:

... federated structures that provide humans both intellectual and physical access to the huge and growing worldwide networks of information encoded in multimedia digital formats.

Early research in digital libraries has focussed on the digitization of existing material, for the preservation of our cultural heritage, as well as on architectural issues for the 'electronic preservation', so to speak, of digital libraries themselves, to make them "immune to degradation and technological obsolescence", Baeza-Yates and Ribeiro-Neto (1999).

To bring order in the variety of research issues related to digital libraries, Baeza-Yates and Ribeiro-Neto (1999) introduces a set of abstractions that is known as the 5S model:

digital libraries (5S)

- *streams*: (content) – from text to multimedia content
- *structures*: (data) – from database to hypertext networks
- *spaces*: (information) – from vector space to virtual reality
- *scenarios*: (procedures) – from service to stories
- *societies*: (stakeholders) – from authors to libraries

These abstractions act as "a framework for providing theoretical and practical unification of digital libraries". More concretely, observe that the framework encompasses three technical notions (streams, structures and spaces; which correspond more or less with data, content and information) and two notions related to the social context of digital libraries (scenarios and societies; which range over possible uses and users, respectively).

For further research you may look at the following resources:

D-Lib Forum – <http://www.dlib.org>

Informedia – <http://www.informedia.cs.cmu.edu>

The D-Lib Forum site gives access to a variety of resources, including a magazine with background articles as well as a test-suite that may help you in developing digital library technology. The Informedia site provides an example of a digital library project, with research on, among others, video content analysis, summarization and in-context result presentation.



4

7.3 networked multimedia

For the end user there should not be much difference between a stand-alone media presentation and a networked media presentation. But what goes on *behind the scenes* will be totally different. In this section, we will study, or rather have a glance at, the issues that play a role in realizing effective multimedia presentations. These issues concern the management of resources by the underlying network infrastructure, but may also concern authoring to the extent that the choice of which media objects to present may affect the demands on resources.

To begin, let's try to establish, following Fluckiger (1995), in what sense networked multimedia applications might differ from other network applications:

networked multimedia

- real-time transmission of continuous media information (audio, video)
- substantial volumes of data (despite compression)
- distribution-oriented – e.g. audio/video broadcast

Naturally, the extent to which network resource demands are made depends heavily on the application at hand. But as an example, you might think of the retransmission of television news items on demand, as nowadays provided via both cable and DSL.

For any network to satisfy such demands, a number of criteria must be met, that may be summarized as: throughput, in terms of bitrates and burstiness; transmission delay, including signal propagation time; delay variation, also known as jitter; and error rate, that is data alteration and loss.

For a detailed discussion of criteria, consult Fluckiger (1995), or any other book on networks and distributed systems. With respect to distribution-oriented multimedia, that is audio and video broadcasts, two additional criteria play a role,

in particular: multicasting and broadcasting capabilities and document caching. Especially caching strategies are of utmost importance if large volumes of data need to be (re)transmitted.

Now, how do we guarantee that our (networked) multimedia presentations will come across with the right quality, that is free of annoying jitter, without loss or distortion, without long periods of waiting. For this, the somewhat magical notion of *Quality of Service* has been invented. Quoting Fluckiger (1995):

Quality of Service

Quality of Service is a concept based on the statement that not all applications need the same performance from the network over which they run. Thus, applications may indicate their specific requirements to the network, before they actually start transmitting information data.

Quality of Service (QoS) is one of these notions that gets delegated to the other parties, all the time. For example, in the MPEG-4 standard proposal interfaces are provided to determine *QoS* parameters, but the actual realization of it is left to the network providers. According to Fluckiger (1995) it is not entirely clear how *QoS* requirements should be interpreted. We have the following options: we might consider them as hard requirements, or alternatively as guidance for optimizing internal resources, or even more simply as criteria for the acceptance of a request.

At present, one thing is certain. The current web does not offer *Quality of Service*. And what is worse, presentation formats (such as for example *flash*) do not cope well with the variability of resources. More specifically, you may get quite different results when you switch to another display platform



5

virtual objects

Ideally, it should not make any difference to the author at what display platform a presentation is viewed, nor should the author have to worry about low-quality or ill-functioning networks. In practice, however, it seems not to be realistic to hide all this variability from the author and delegate it entirely to the 'lower layers' as in the MPEG-4 proposal.

Both in the SMIL and RM3D standards, provisions are made for the author to provide a range of options from which one will be chosen, dependent on for example availability, platform characteristics, and network capabilities.

A formal characterization of such an approach is given in Subrahmanian (1998), by defining *virtual objects*.

virtual objects

- $VO = \{(O_i, Q_i, C_i) \mid 1 \leq i \leq k\}$

where

- C_1, \dots, C_k – mutually exclusive conditions
- Q_1, \dots, Q_k – queries
- O_1, \dots, O_k – objects

In general, a virtual object is a media object that consists of multiple objects, that may be obtained by executing a query, having mutually exclusive conditions to determine which object will be selected. Actually, the requirement that the conditions are mutually exclusive is overly strict. A more pragmatic approach would be to regard the objects as an ordered sequence, from which the first eligible one will be chosen, that is provided that its associated conditions are satisfied.

As an example, you may look at the Universal Media proposal from the Web3D Consortium, that allows for providing multiple URNs or URLs, of which the first one that is available is chosen. In this way, for instance, a texture may be loaded from the local hard disk, or if it is not available there from some site that replicates the Universal Media textures.



6

networked virtual environments

It does seem to be an exaggeration to declare *networked virtual environments* to be the ultimate challenge for networked multimedia, considering that such environments may contain all types of (streaming) media, including video and 3D graphics, in addition to rich interaction facilities. (if you have no idea what I am talking about, just think of, for example, Quake or DOOM, and read on.) To be somewhat more precise, we may list a number of essential characteristics of networked virtual environments, taken from Singhal and Zyda (1999):

networked virtual environments

- *shared sense of space* – room, building, terrain
- *shared sense of presence* – avatar (body and motion)
- *shared sense of time* – real-time interaction and behavior

In addition, networked virtual environments offer

- *a way to communicate* – by gesture, voice or text
- *a way to share ...* – interaction through objects

Dependent on the visual realism, resolution and interaction modes such an environment may be more or less 'immersive'. In a truly immersive environment, for example one with a haptic interface and force feedback, interaction through objects may become even threatening. In desktop VEs, sharing may be limited to the shoot-em-up type of interaction, that is in effect the exchange of bullets.

Networked virtual environments have a relatively long history. An early example is SIMNET (dating from 1984), a distributed command and control simulation developed for the US Department of Defense, Singhal and Zyda (1999). Although commercial multi-user virtual communities, such as the *blaxxun* Community server, may also be ranked under networked virtual environments, the volume of data exchange needed for maintaining an up-to-date state is far less for those environments than for game-like simulation environments from the military tradition. Consider, as an example, a command and control strategy game which contains a variety of vehicles, each of which must send out a so-called *Protocol Data Unit* (PDU), to update the other participants as to their actual location and speed. When the delivery of PDUs is delayed (due to for example geographic dispersion, the number of participants, and the size of the PDU), other strategies, such as *dead reckoning*, must be used to perform collision detection and determine possible hits.

To conclude, let's establish what challenges networked virtual environments offers with respect to software design and network performance.

challenges

- *network bandwidth* – limited resource
- *heterogeneity* – multiple platforms
- *distributed interaction* – network delays
- *resource management* – real-time interaction and shared objects
- *failure management* – stop, ..., degradation
- *scalability* – wrt. number of participants

Now it would be too easy to delegate this all back to the network provider. Simply requiring more bandwidth would not solve the scalability problem and even though adding bandwidth might allow for adding another hundred of entities, smart updates and caching is probably needed to cope with large numbers of participants.

The distinguishing feature of networked virtual environments, in this respect, is the need to

manage dynamic shared state

to allow for real-time interaction between the participants. Failing to do so would result in poor performance which would cause immersion, if present at all, to be lost immediately.



7

example(s) – *unreal*

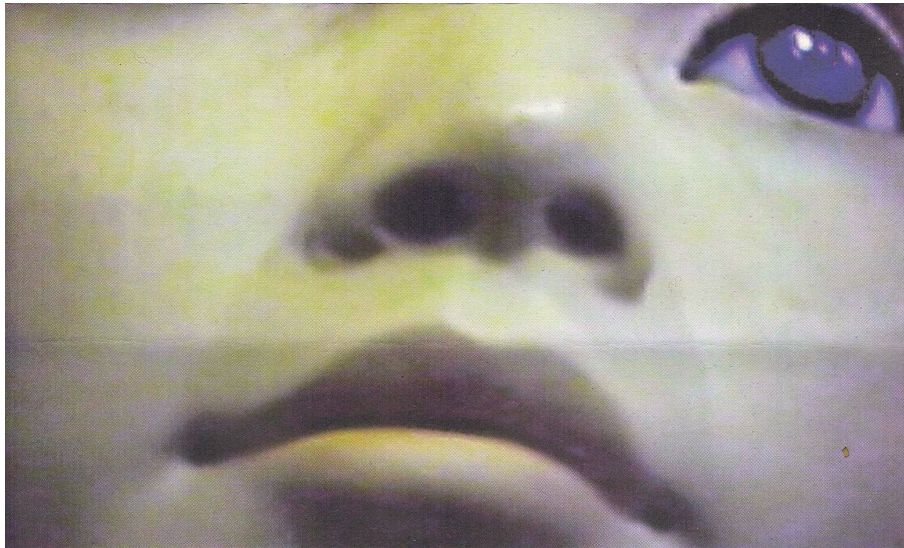
*Unreal Tournament*⁸³ is a highly popular multiplayer game. The storyline is simple, but effective: *It's the year 2362. The most anticipated Tournament ever is about to take place, dwarfing the spectacle and drama of previous events. The finest competitors ever assembled prepare to lay waste to their opponents and claim the Tournament Trophy for themselves.*

There are a number of roles you can associate with:

the corrupt, thunder crash, iron guard, juggernauts, iron skull, sun blade, super nova, black legion, fire storm, hellions, bloof fist, goliath

An interesting feature of the Unreal Tournament games is that they can be adapted and even be re-programmed⁸⁴ by the users themselves, has been done for example for the *Mission Rehearsal Exercise* discussed in section 9.2.

scripting: www.gamedev.net/reference/list.asp?categoryid=76



⁸³www.unrealtournament.com

⁸⁴www.unrealtournament.com/ut2004/screenshots.php

research directions— *architectural patterns*

Facing the task of developing a multimedia information system, there are many options. Currently, the web seems to be the dominant infrastructure upon which to build a multimedia system. Now, assuming that we chose the web as our vehicle, how should we approach building such a system or, in other words, what architectural patterns can we deploy to build an actual multimedia information system? As you undoubtedly know, the web is a document system that makes a clear distinction between *servers* that deliver documents and *clients* that display documents. See Eliens (2000), section 12.1. At the server-side you are free to do almost anything, as long as the document is delivered in the proper format. At the client-side, we have a generic document viewer that is suitable for HTML with images and sound. Dependent on the actual browser, a number of other formats may be allowed. However, in general, extensions with additional formats are realized by so-called *plugins* that are loaded by the browser to enable a particular format, such as *shockwave*, *flash* or *VRML*. Nowadays, there is an overwhelming number of formats including, apart from the formats mentioned, audio and video formats as well as a number of XML-based formats as for example SMIL and SVG. For each of these formats the user (client) has to download a plugin. An alternative to plugins (at the client-side) is provided by Java *applets*. For Java applets the user does not need to download any code, since the Java platform takes care of downloading the necessary classes. However, since applets may be of arbitrary complexity, downloading the classes needed by an application may take prohibitively long.

The actual situation at the client-side may be even more complex. In many cases a media format does not only require a plugin, but also an applet. The plugin and applet can communicate with each other through a mechanism (introduced by Netscape under the name LiveConnect) which allows for exchanging messages using the built-in DOM (Document Object Model) of the browser. In addition, the plugin and applet may be controlled through Javascript (or VBscript). A little dazzling at first perhaps, but usually not too difficult to deal with in practice.

Despite the fact that the web provides a general infrastructure for both (multimedia) servers and clients, it might be worthwhile to explore other options, at the client-side as well as the server-side. In the following, we will look briefly at:

- the Java Media Framework, and
- the DLP+X3D platform

as examples of, respectively, a framework for creating dedicated multimedia applications at the client-side and a framework for developing intelligent multimedia systems, with client-side (rich media 3D) components as well as additional server-side (agent) components.

Java Media Framework The Java platform offers rich means to create (distributed) systems. Also included are powerful GUI libraries (in particular, Swing),

3D libraries (Java3D) and libraries that allow the use and manipulation of images, audio and video (the Java Media Framework). Or, in the words of the SUN web site:

java Media Framework⁸⁵

The Java™ Media APIs meet the increasing demand for multimedia in the enterprise by providing a unified, non-proprietary, platform-neutral solution. This set of APIs supports the integration of audio and video clips, animated presentations, 2D fonts, graphics, and images, as well as speech input/output and 3D models. By providing standard players and integrating these supporting technologies, the Java Media APIs enable developers to produce and distribute compelling, media-rich content.

However, although Java was once introduced as the *dial tone of the Internet* (see Eliens (2000), section 6.3), due to security restrictions on applets it is not always possible to deploy media-rich applets, without taking recourse to the Java plugin to circumvent these restrictions.

DLP+X3D In our DLP+X3D platform, that is introduced in section ?? and described in more detail in appendix ??, we adopted a different approach by assuming the availability of a generic X3D/VRML plugin with a Java-based External Authoring Interface (EAI). In addition, we deploy a high-level distributed logic programming language (DLP) to control the content and behavior of the plugin. Moreover, DLP may also be used for creating dedicated (intelligent) servers to allow for multi-user applications.

The DLP language is Java-based and is loaded using an applet. (The DLP jar file is of medium size, about 800 K, and does not require the download of any additional code.) Due, again, to the security restrictions on applets, additional DLP servers must reside on the site from where the applet was downloaded.

Our plugin, which is currently the *blaxxun* VRML plugin, allows for incorporating a fairly large number of rich media formats, including (real) audio and (real) video., thus allowing for an integrated presentation environment where rich media can be displayed in 3D space in a unified manner. A disadvantage of such a unified presentation format, however, is that additional authoring effort is required to realize the integration of the various formats.

7.4 development(s) – living in a virtual economy

living in a Virtual Economy – Applications

Mashups on the Web are interesting representatives of what one may call a *virtual economy*, with a business-model that is not grounded in traditional *production* and *trade* values, but rather consists of value-added services with an indirect, albeit substantial, financial spin-off, due to recommendations and referrals. The basic mechanisms in a recommender economy are, according to Economy:

⁸⁵java.sun.com/products/java-media/jmf/reference/api

- cross sale – users who bought A also bought B
- up sale – if you buy A and B together ...

Where the principles underlying this virtual economy have definitely proven their value in first (ordinary) life economy, what are the chances that these principles are also valid in Second Life?

According to the media companies selling their services to assist the creation of presence in Second Life, there are plenty *New Media Opportunities In The Online World Second Life*⁸⁶, to a possibly even greater extent, as they boldly claim, as in what they call *the predecessor of Second Life, the World Wide Web*.

To assess the role web services, including semantic web services, may play in Second Life, it seems worthwhile to investigate to what extent web services can be deployed to deliver more traditional media, such as *digital TV*. To support the business model of digital TV, which in outline may be summarized as *providing additional information, game playing and video on demand*, with an appropriate payment scheme, DTV argue in favor of the use of a SOA (Service Oriented Architecture), to allow for a unified, well-maintainable approach in managing collections of audio-visual objects. Such services would include meta-data annotation, water-marking for intellectual property protection, and search facilities for the end-user. Framework even propose to wrap each individual audio-visual object in a (semantic) web service and provide compound services based on semantic web technologies such as OWL-S⁸⁷ (OWL-based Web Service Ontology) and WSMO⁸⁸ (Web Service Modelling Ontology) using semi-automatic methods together with appropriate semantic web tools⁸⁹, for the description and composition of such services. Obviously, there is a great technical challenge in creating such self adjusting service environments.

With respect to the application of web services in Second Life, however, a far more modest aim, it seems that nevertheless the business model associated with the delivery of media items through digital TV channels may profitably be used in Second Life, and also the idea of wrapping media items in web services has in some way an immediate appeal.

Leaving the economic issues aside we will briefly characterize two applications that we envisage to realize within the virtual campus of VU @ Second Life. The first application is a game meant to give information about climate change. More specifically, as a submission for the Dutch Contest for the Communication of Science, we have developed, in close collaboration with the VU Climate Center, the *Clima Futura*⁹⁰ game, that invites players to manage resources related to climate change, over a period of 100 years, and also allows the player to play minigames, to gain access to audio-visual information about scientific developments as well as political issues related to the *inconvenient truth* of climate change. Although we do not intend to realize the whole game in Second Life, we actually use *flash* to reach an audience as wide as possible, as a pilot parts

⁸⁶www.youtube.com/watch?v=8NOHRJB9uyI

⁸⁷www.daml.org/services/owl-s/

⁸⁸www.wsmo.org/

⁸⁹composing-the-semantic-web.blogspot.com/

⁹⁰www.climafutura.nl/

of the game could fruitfully be realized in the VU virtual campus, in particular the search for knowledge, that is looking for an expert in a particular area of (climate-related) research. A similar quest was implemented in our Half Life 2 based game VULife, Eliens and Bhikharie (2006), where the player had to visit nine information spots, which resulted in displaying in a HUD nine square matrix the location of a hidden treasure, which was then actually the power to use arms. Technical issues in realizing Clima Futura in Second Life are support for ranking, as well as meta-information with respect to locations where relevant information can be found, which may be realized with the techniques indicated in section 5. Another issue is giving flexible access to video material related to specific topics in climate change.

Another application we wish to discuss is PANORAMA, already mentioned in section 5. In developing PANORAMA we proceeded from the assumption that people somehow like to have a feel of what is going on in the workspace, although not in any detail, and also like to see items of personal interest, including (their own) birth-announcements and sport troffes. The original version of PANORAMA was developed in DirectX, and was meant to be displayed on a large screen, with a static viewpoint on a dynamically changing scene, reflecting the activity in the workspace and the self-reflective contributions in an artful way. Moving from a central large screen in first life to a visualisation embedded in Second Life, which is moreover subject to first-person viewpoint dynamics, is quite a challenge. The most simple solution would be to project the large PANORAMA screen onto a display object in Second Life using live video streams but this would leave that challenge unanswered. Embedding PANORAMA in Second Life would allow us to observe, in more detail than in a previous user study, the behavior of users, that is, to be more precise, the proximity to particular objects of interest, the duration of their presence, and, using the mechanisms of recommendation, their interest in related items. Technically, such monitoring can be achieved using the sensors and listeners described in section 5. To make sense of such data, however, we need some model that allows for an interpretation that is more meaningful than the mere registration of presence.



9

questions

information system architecture

1. (*) What are the issues in designing a (multimedia) information system architecture. Discuss the tradeoffs involved.

concepts

2. What considerations would you have when designing an architecture for a multimedia information system.
3. Characterize the notion of *media abstraction*.
4. What are the issues in *networked multimedia*.

technology

5. Describe (the structure of) a video database, using *media abstractions*.
6. Give a definition of the notion of a *structured multimedia database*.
7. Give an example (with explanation) of querying a *hybrid multimedia database*.
8. Define (and explain) the notion of *virtual objects* in *networked multimedia*.

projects & further reading As a project, you may implement a multi-player game in which you may exchange pictures and videos, for example pictures and videos of celebrities.

Further you may explore the development of a data format for text, images and video with appropriate presentation parameters, including positioning on the screen and intermediate transitions.

For further reading you may study information system architecture patterns⁹¹, and explore the technical issues of constructing server based advanced multimedia applications in Li and Drew (2004).

the artwork

1. examples of dutch design, from Betsky (2004).
2. idem.
3. screenshots – from *splinter cell: chaos theory*, taken from Veronica/Gammo⁹², a television program about games.
4. screenshots – respectively *Sekken 5*, *Sims 2*, and *Super Monkey Ball*, taken from insidegamer.nl⁹³.
5. screenshots – from Unreal Tournament⁹⁴, see section 7.3.
6. idem.
7. idem.
8. *resonance* – exhibition and performances, Montevideo⁹⁵, april 2005.
9. signs – sports, van Rooijen (2003), p. 274, 275.

Opening this chapter are examples of *dutch design*, taken from the book *False Flat*, with the somewhat arrogant subtitle *why is dutch design so good?*. It is often noted that dutch design is original, functional and free from false traditionalism. Well, judge for yourself.

⁹¹www.opengroup.org/architecture/togaf8-doc/arch/p4/patterns/patterns.htm

⁹²www.gammo.nl

⁹³<http://www.insidegamer.nl>

⁹⁴www.unrealtournament.com/ut2004/screenshots.php

⁹⁵www.montevideo.nl

The screenshots from the various games are included as a preparation for chapter 9, where we discuss realism and immersion in games, and also because multiplayer games like *Unreal Tournament* have all the functionality a serious application would ever need.

part iv. applications & tools

a journey of a thousand miles begins with the first step
chinese/japanese proverb

chapters:

- 8. virtual environments
- 9. digital content creation
- 10. application development



2

reading directives In this part we will look in more detail at virtual environments as an interface to complex multimedia information spaces. In chapters 9 and 10, we will consider the issues that come into focus when creating digital content, and more in general, when developing a multimedia application. In chapter 10, the final chapter, we will look at some examples of multimedia application development.

Essential sections are section 8.1, which argues how virtual reality interfaces may be relevant, and sections 9.2 and 10.2, which provide examples of multimedia application development. Dependent on your experience you may skip sections 9.1 and 10.1, which provide rules of thumb for respectively content creation and application development.

perspectives Even in a more practical sense there are many perspectives that may characterize your academic stance. Design and application development, obviously, does not only consist of aesthetic or technical issues. So, non-exhaustively, you may look at such issues from the following perspectives:

perspectives – multimedia applications

- technical – algorithmic effects
- sociological – stakeholders and teamwork
- tool selection – Maya vs 3DSMAX
- political – negotiating support
- scientific – experience design
- computer science – tools and technologies
- artistic – portfolio as a design product

For example, the issues you may come across in an actual project may have to do more with people than any thing else, in other words may be more of a political nature, than of an aesthetic nature.

essay topics Even when you are primarily interested in the practice of developing digital content, it might well pay off to reflect on more theoretical issues. For example, consider writing an essay about:

- 2D vs 3D aesthetics animation – stills, sequences and stories
- elements of style – diversity and confluence
- models of creativity – a critical evaluation

In writing about such issues you should always beware of the risk of abstract speculation. So, look for examples in the domain of art, design or street culture to demonstrate your point.



3

the artwork

1. walking figure – sculpture by Alberto Giacometti, Hohl (1971).
2. signs – meteorological symbols, van Rooijen (2003), p. 214, 215.
3. photographs – Jaap Stahlie⁹⁶, commissioned work, using traditional non-digital techniques.

⁹⁶ www.jaapstahlie.com

8. virtual environments

augmented virtuality acts as an intelligent looking glass

learning objectives

After reading this chapter you should be able to characterize the notion of virtual context, discuss the issue of information retrieval in virtual environments, explain what is meant about intelligent multimedia and discuss the potential role of intelligent agents in multimedia applications.

From a user perspective, virtual environments offer the most advanced interface to multimedia information systems. Virtual environments involve the use of (high resolution) 3D graphics, intuitive interaction facilities and possibly support for multiple users. In this chapter, we will explore the use of (desktop) virtual environments as an interface to (multimedia) information systems. We will discuss a number of prototype implementations illustrating, respectively, how paintings can be related to their context, how navigation may be seen as a suitable answer to a query, and how we can define intelligent agents that can interact with the information space. Take good notice, the use of virtual environments as an interface to information systems represents a major challenge for future research!



1

8.1 virtual context

Imagine that you walk in a museum. You see a painting that you like. It depicts the Dam square in 17th century Amsterdam. Now, take a step forwards and suddenly you are in the middle of the scene you previously watched from some distance. These things happen in movies.

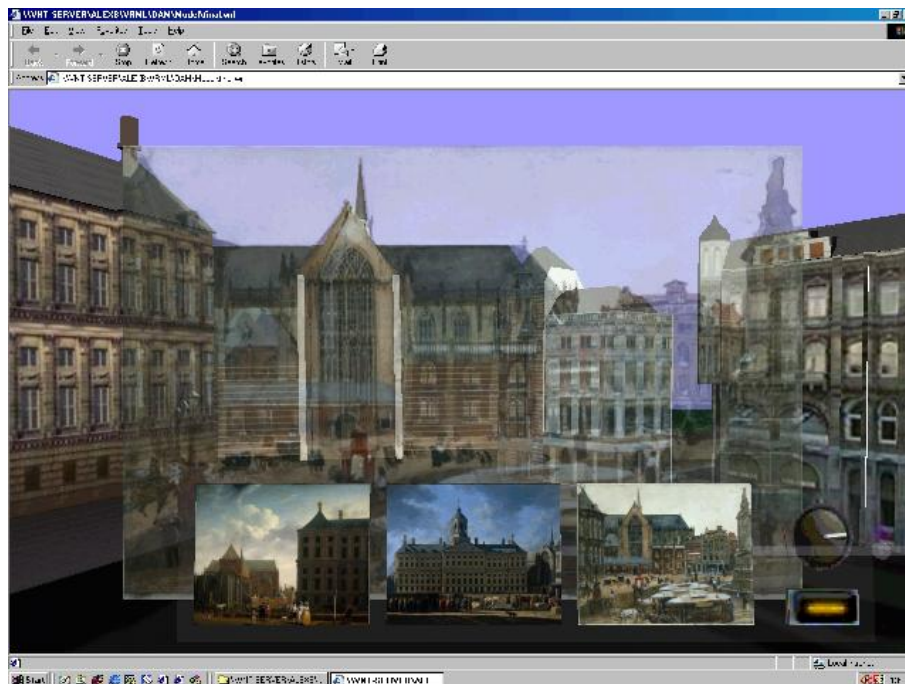
Now imagine that you are walking on the Dam square, some sunday afternoon in May 2001, looking at the Royal Palace, asking yourself is this where Willem-Alexander and Maxima will get married. And you wonder, what did this building and the Dam square look like three centuries ago. To satisfy your curiosity you go to the Royal Museum, which is only a half hour walk from there, and you go to the room where the 17th century city-scape paintings are. The rest is history.

We can improve on the latter scenario I think. So let's explore the options. First of all, we may establish that the Dam square represents a rich information space. Well, the Dam Square is a 'real world' environment, with it has 700 years of (recorded) history. It has a fair amount of historical buildings, and both buildings and street life have changed significantly over time.

So, we can rephrase our problem as

how can we give access to the 'Dam square' information space

But now we forget one thing. The idea underlying the last scenario is that we somehow realize a seamless transition from the real life experience to the information space. Well, of course, we cannot do that. So what did we do?



2

Look at the screenshot from our *virtual context* prototype. You can also start the VRML demo version that is online, by clicking on the screenshot. What you see is (a model of) the Dam square, more or less as it was in 2001. In the lower part, you see a panel with paintings. When you click on one of these painting,

your viewpoint is changed so that you observe the real building from the point of view from which the painting was made. Then using the controls to the right of the panel, you can overlay the real building with a more or less transparent rendering of the painting. You can modify the degree of transparency by turning the dial control. You may also make the panel of paintings invisible, so that it does not disrupt your view of the Dam and the chosen overlay.

In other words, we have a VR model of Dam square and a selection of related paintings from the Royal Museum, that are presented in a panel from which the user can choose a painting. We deploy viewpoint adjustment, to match the selected painting, and we use overlay of paintings over buildings, in varying degrees of transparency, to give the user an impression of how the differences between the scene depicted in the painting and the actual scene in (the virtual) reality.

We have chosen for the phrase *virtual context* to characterize this prototype, since it does express how virtual reality technology enables us to relate an information space to its original context.

From the perspective of virtual reality, however, we could also have characterized our prototype as an application of *augmented virtual reality*, since what we have is a virtual reality model of a real-life location that is augmented with information that is related to it, (almost) without disrupting the virtual reality experience. In summary, we may characterize our approach as follows.

augmented virtual reality

- give user sense of geographic placement of buildings
- show how multiple objects in a museum relate to each other
- show what paintings convey about their subject, and how

Considering the fact that many city-scape paintings of Amsterdam have been made, many of which are in the Royal Museum, and that paintings may say many things about their subject, we believe that our approach is viable for this particular instance. The augmented virtual reality approach would also qualify as a possible approach to cultural heritage projects, provided that sufficient pictorial material is available or can be reconstructed.

Although we were quite satisfied with what we accomplished, there are still many things that can be done and also a number of open problems. Guided tours are a well-known phenomenon. But how to place them in our virtual context is not entirely clear. As another problem, our approach does not seem suited to account for buildings that do no longer exist. Another thing we have to study is how to change the temporal context, that is for example change from a model of the dam in 2001 to a model of the Dam in 1850. We would then also like to have 'viewpoint transitions' over space and time!

Finally, to give better access to the underlying information space we must also provide for textual user queries, and find an adequate response to those queries.

<name a=ex-t-7-1>

VRML To realize our prototype we used VRML, which limits us to medium quality desktop VR. At this stage, VRML is a good option, since it is a relatively

stable format with a reasonable programmatic model. In short, what VRML offers is

VRML

- declarative means for defining geometry and appearance
- prototype abstraction mechanism
- powerful event model
- relatively strong programmatic capabilities

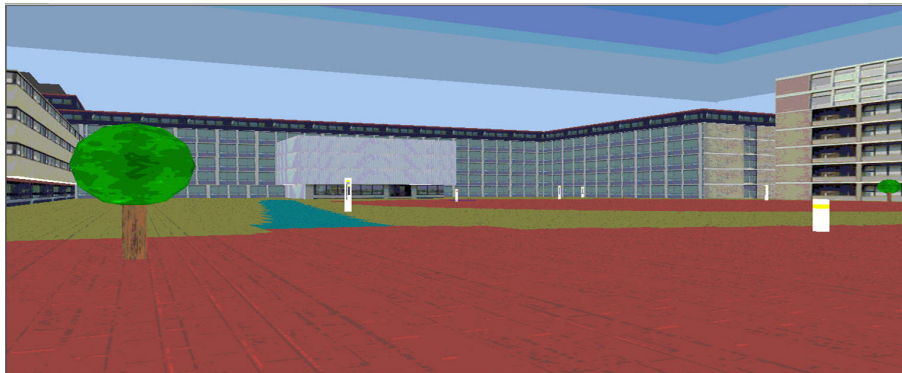
Although VRML allows for writing models (including geometry and appearance) using a plain text editor, many tools support export to VRML. As a consequence, often tools are used to create more complex models.

In addition, VRML allows for defining prototype abstractions, so reuse of models and behavior can be easily realized.

Defining dynamic behavior involves the routing of events that may come from a variety of built-in sensors (for example a TimeSensor for animations) to scripts or so-called interpolators, that allow for the manipulation of geometry and appearance parameters of the model.

In particular, the use of scripts or the *External Authoring Interface* (EAI), that allows for defining behavior in Java, is essential for realizing complex behavior.

Summarizing, VRML is a sufficiently rich declarative language for defining 3D scenes, with a relatively powerful programming model for realizing complex behavior. Some may think that VRML is dead. It isn't. The underlying model is endorsed in both the X3D and RM3D standards, simply since it has proven its worth.



3

research directions— *augmented virtuality*

Given an information space, there is a duality between information and presentation. For an audience or user to be able to digest a presentation, the amount of information must be limited. Effective presentation, moreover, requires the use of proper rethorics (which may be transcoded as *ways of presenting*) that belong

to the medium. Using VR, which is (even in its desktop format) a powerful presentation vehicle, one should always beware of the question *what is it good for?* Generally one may ask, what is the added value of using VR? In an abstract fashion the answer should be, to bridge the gap between information content and presentation. Or, in other words, to resolve the duality between information and presentation!

Let's look at an example, a site about archeology, announced as a site offering *Virtual Archeology*. Perhaps it is good to bring to your attention that the *virtual*, in Computer Science, means nothing but another level of indirection to allow for a (more) flexible usage of entities or objects. See Eliens (2000), section 1.2.

virtual archeology

- variety of archeological sites
- various paths through individual site
- reconstruction of 'lost' elements
- 'discovery' of new material
- glossary – general background knowledge

For a site about archeology, *virtual* means the ability to present the information in a number of ways, for example as paths through a particular site, with the possibility to explore the reconstruction of lost or perished material, and (for students) to discover new perspectives on the material. In addition, for didactic reasons there may also be a glossary to explain concepts from archeology.

Now, how would you construct such a site about virtual archeology? As a collection of HTML pages and links? It seems that we can do better, using VR and rich interaction mechanisms!

So, what is meant by *augmented virtuality*? Nothing that hasn't been expressed by the notion of *augmented virtual reality*, of which an example has been given in this section. The phrase *augmented virtuality* itself is just one of those potentially meaningless fancy phrases. It was introduced simply to draw your attention to the duality between information and presentation, and to invite you to think about possible ways to resolve this duality.

8.2 navigation by query

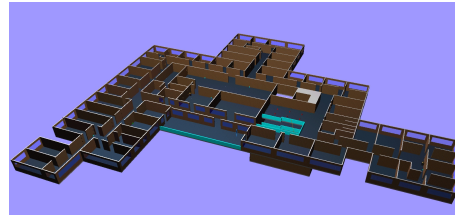
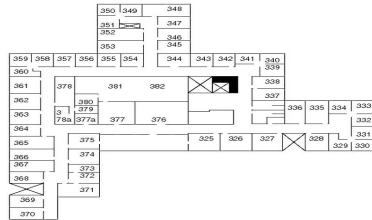
Virtual worlds form (in itself) a rich repository of multimedia information. So, when working on the musical feature detector, sketched in section 6.3, the thought occurred to ask funding for a research project on information retrieval in virtual worlds. This project is called RIF, which stands for

RIF

Retrieval of Information in Virtual Worlds using Feature Detectors

For the RIF project, we decided to develop a small multi-user community of our own, using the *blaxxun* Community Server. Then, during the development of our own virtual environment, the question came up of how to present the results of a

query to the user. The concept we came up with was *navigation by query*, and in this section we will look at the prototype we developed to explore this concept.



On the left is the 2D map of the third floor of CWI, on the right the model generated from it.

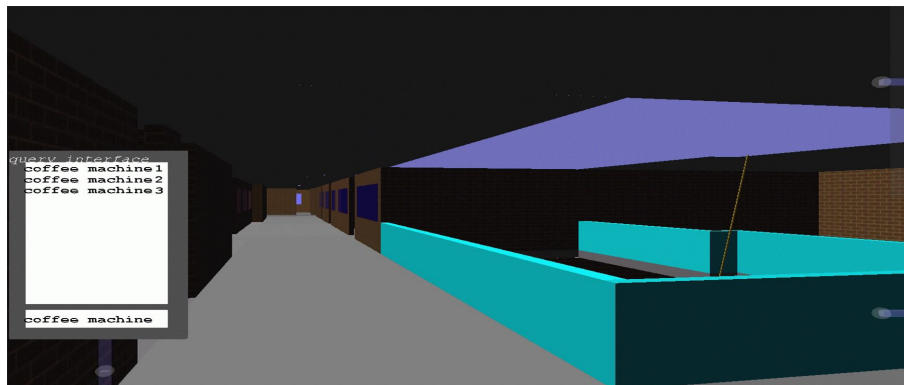
4

case study – CWI

For our prototype, we took one of the worlds of our virtual environment, the third floor of the CWI. The reason for this is that we were (at the time) doing our research there, and so there were plenty locations of interest, such as the rooms of our colleagues, the printer room, and not to forget, the coffee corner.

We started out by taking a map of the third floor, and developed a model of it, using a tool developed by a student, who needed such a tool for realizing his game *Out of the Dark*.

When dwelling around in (this part of) our virtual environment, the user may pose (arbitrary) queries, for example *where is the coffee machine*.



5

Remind, that after a few hours of research, coffee might be needed to get fresh ideas!

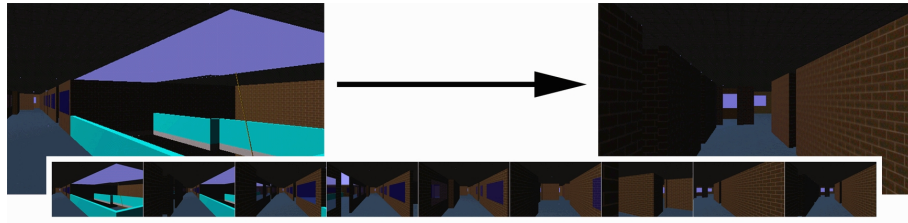
As a result, the user is then so to speak taken by the hand and led to one of the coffee machines that can be found on the third floor. In effect, with knowledge

of the layout of the building a viewpoint transformation is executed, in a tempo that allows the user to

explore and discover

the (model of the) third floor of the CWI.

The idea is rather straightforward. Some have asked us why *navigation by query* might be useful. Well, simply, it seems to offer an interesting alternative to navigation by explicit interaction and navigation in the form of a guided tour. Our primary goal in developing the prototype, however, was to see whether navigation by query is feasible, and under what conditions.



6

information in virtual worlds

Developing the prototype has forced us to think more explicitly about what information is available in virtual worlds, and (perhaps more importantly) how to gain access to it. So the question we asked ourselves was

what are we searching for?

Now, in a virtual world, such as the ones built with VRML, we can distinguish between the following types of information: viewpoints, that is positions in the world from where interesting things can be looked at or accessed in any other way; areas of interest, where those interesting things are located; objects, that may provide information or offer particular kinds of functionality; persons, that is other users that are visiting the world; and even text, which might be on billboards or slides.

Some of this information is, so to speak, hard-wired in the model and may be accessed anytime, in some cases even by scanning the VRML file. Other information, however, is of a more dynamic nature, since it might be due to the presence of multiple users, the execution of scripts, or events that happen in response to user interaction. Some information may even be explicitly hidden, such as for example the actions one should take in solving a puzzle or playing a game.

When the virtual world is loaded, all the information (or at least most of it) is present in the so-called scenegraph, the structure that is built to render the

world. Using the software interface to access the scenegraph (which is usually browser-specific), we can look for annotations, node types and textual content to extract information from the world. This information may then be stored in a database, and be reused later for other users and queries. In principle, more advanced techniques could be used to extract information from the materials used, and even from textures and geometry.

presentation issues

In our prototype, we aimed at solving the question how to present the results of a query, using navigation. First of all, we had to

choose a metaphor

for navigation. Dependent on the object of interest a viewpoint can be selected. For a viewpoint, it is just that viewpoint. For an area of interest, the viewpoint selected must enable the user to view the area, and when objects or persons are chosen, care must be taken not to block the users' view by some obstacle.

Now answering a query then comes down to planning a suitable route and apply a series of viewpoint transformations along that route.

Not surprisingly, the navigation metaphor we chose was

walking

as the preferred mode of viewpoint transformations.

the prototype

The structure of the prototype is depicted in the figure below.

In realizing the prototype, we made the following (simplifying) assumptions.

We avoided a number of difficulties by choosing for explicit annotations (which indicate locations and areas of interest), and by avoiding the intricacies of route planning and advanced text processing.

The requirements laid down before hand just stated that we would have a database and that we would avoid superfluous user interface elements. Instead, we used control and input panels written in VRML, in order to provide a 3D(pseudo-immersive) interface.

Now, our assumptions may in principle be relaxed. For example, annotation might be done incrementally by users that visit the world or to some extent even automatically, by using feature extractors. Instead of explicit maps, we may dynamically create maps based on users' navigation patterns. And, instead of simple keyword matching, we may apply more advanced text retrieval techniques. But this is left as future work. Anyway, we were satisfied that we could state the following conclusions:

conclusions

- navigation by query is feasible and may help users to find locations and objects
- determining suitable navigation routes without an explicitly defined map is hard

As is often the result with good research, you solve one problem and a number of other problems come up. So, one of the questions that remains was: how can we improve on navigation? What additional navigation support can we provide?

research directions— *extended user interfaces*

Is desktop VR a suitable candidate as an interface technology for multimedia information systems? And if so, what needs to be done to apply this technology effectively?

At first sight, our vision of applying VR as an interface to multimedia systems seems to be doomed to fail. As Ben Schneiderman, in a keynote for the Web3D Symposium 2002, observes:

3D GUI

Wishful thinking about the widespread adoption of three-dimensional interfaces has not helped spawn winning applications. Success stories with three-dimensional games do not translate into broad acceptance of head-tracking immersive virtual reality. To accelerate adoption of advanced interfaces, designers must understand their appeal and performance benefits as well as honestly identify their deficits. We need to separate out the features that make 3D useful and understand how they help overcome the challenges of dis-orientation during navigation and distraction from occlusion.

Ben Shneiderman

So, even if advanced (3D) user interfaces might be useful, there are a number of questions to raise. Again, following Ben Schneiderman:

Does spatial memory improve with 3D layouts? Is it true that 3D is more natural and easier to learn? Careful empirical studies clarify why modest aspects of 3D, such as shading for buttons and overlapping of windows are helpful, but 3D bar charts and directory structures are not. 3D sometimes pays off for medical imagery, chemical molecules, and architecture, but has yet to prove beneficial for performance measures in shopping or operating systems.

Ben Shneiderman

In particular, according to Schneiderman, we must beware of *tacky 3D*, gadgets in 3D space that are superfluous and only hindering the user to perform a task. Well-spoken and based on adequate observations! Nevertheless, at this stage, we should (in my opinion) adopt a slightly more liberal attitude and explore in what ways the presentation of (multimedia) information could be augmented by using (desktop) VR. But enough about *augmentation*. Let's discuss technology, and investigate what is required for the effective deployment of VR from the point of view of intelligent agents!



7

8.3 intelligent agents

Visitors in virtual environments are often represented by so-called avatars. Wouldn't it be nice to have intelligent avatars that can show you around, and tell you more about the (virtual) world you're in.

Now, this is how the idea came up to merge the RIF project, which was about information retrieval, with the WASP project, another acronym, which stands for:

WASP

Web Agent Support Program

The WASP project aims at realizing intelligent services using both client-side and server-side agents, and possibly multiple agents. The technical vehicle for realizing agents is the language DLP, which stands for

DLP

Distributed Logic Programming

Merging the two projects required providing the full VRML EAI API in DLP, so that DLP could be used for programming the dynamic aspects of VRML worlds.

background Historically, the WASP project precedes the RIF project, but we started working on it after the RIF project had already started. Merging these two projects had more consequences than we could predict at the time. The major consequence is that we shifted focus with respect to programming the dynamics of virtual environments. Instead of scripts (in Javascript), Java (through the EAI), and even C++ (to program *blaxxun* Community Server extensions), we introduced the distributed logic programming language DLP as a uniform computational platform. In particular, for programming intelligent agents a logic programming

language is much more suited than any other language. All we had to do was merge DLP with VRML, which we did by lifting the Java EAI to DLP, so that function calls are available as built-ins in the logic programming language.

When experimenting with agents, and in particular communication between agents, we found that communication between agents may be used to maintain a shared state between multiple users. The idea is simple, for each user there is an agent that observes the world using its 'sensors' and that may change the world using its 'effectors'. When it is notified by some other agent (that is co-located with some other user) it can update its world, according to the notification. Enough background and ideas. Let's look at the prototypes that we developed.



8

multi-user soccer game

To demonstrate the viability of our approach we developed a multi-user soccer game, using the DLP+VRML platform.

We chose for this particular application because it offers us a range of challenges.

multi-user soccer game

- *multiple (human) users* – may join during the game
- *multiple agents* – to participate in the game (e.g. as goalkeeper)
- *reactivity* – players (users and agents) have to react quickly
- *cooperation/competition* – requires 'intelligent' communication
- *dynamic behavior* – sufficiently complex 3D scenes, including the dynamic behavior of the ball

Without going into detail, just imagine that you and some others wish to participate in a game, but there are no other players that want to join. No problem, we just add some intelligent agent football players. And they might as well be taken out when other (human) players announce themselves.

For each agent player, dependent on its role (which might be *goal-keeper*, *defender*, *mid-fielder* and *forward*), a simple cognitive loop is defined: sensing, thinking, acting. Based on the information the agent gets, which includes the agent's position, the location of the ball, and the location of the goal, the agent decides which action to take. This can be expressed rather succinctly as rules in the logic programming formalism, and also the actions can be effected using the built-in VRML functionality of DLP.

Basically, the VRML-related built-ins allow for obtaining and modifying the values of *control points* in the VRML world.

control points

- get/set – position, rotation, viewpoint

These control points are in fact the identifiable nodes in the scenegraph (that is, technically, the nodes that have been given a name using the DEF construct).

This approach allows us to take an arbitrarily complex VRML world and manipulate it using the control points. On the other hand, there are also built-ins that allow for the creation of objects in VRML. In that case, we have much finer control from the logic programming language.

All in all we estimate that, in comparison with other approaches, programming such a game in DLP takes far less time than it would have taken using the basic programming capabilities of VRML.

agents in virtual environments

Let us analyse in somewhat more detail why agents in virtual environments may be useful. First of all, observe that the phrase *agents in virtual environments* has two shades of meaning:

agents in virtual environments

- virtual environments with embedded autonomous agents
- virtual environments supported by ACL communication

where ACL stands for *Agent Communication Language*. Our idea, basically is to use an ACL for realizing shared objects, such as for example the ball in the soccer game.

The general concept of multi-user virtual environments (in VRML) has been studied by the *Living Worlds Working Group*. Let's look at some definitions provided by this working group first. A *scene* is defined as a geometrically bounded, continuously navigable part of the world. Then, more specifically a *world* is defined as a collection of (linked) scenes.

Now, multi-user virtual environments distinguish themselves from single-user virtual environments by allowing for so-called *Shared Objects* in scenes, that is objects that can be seen and interacted with by multiple independent users,

simultaneously. This requires synchronization among multiple clients, which may either be realized through a server or through client-to-client communication.

Commonly, a distinction is made between a *pilot* object and a *drone* object.

Shared Object

- *pilot* – instance that will be replicated
- *drone* – instance that replicates pilot

So, generally speaking, pilot objects control drone objects. There are many ways to realize a pilot-drone replication scheme. We have chosen to use agent technology, and correspondingly we make a distinction between *pilot agents*, that control the state of a shared object, and *drone agents*, that merely replicate the state of a shared object.

Since we have (for example in the soccer game) different types of shared objects, we make a further distinction between agents (for each of which there is a pilot and a drone version). So, we have *object agents*, which control a single shared object (like the soccerball). For these agents the pilot is at the server, and the drone is at the client. We further have agents that control the users' avatars, for which the pilot is at user/client side, and the drone either at the server or the client. Finally, we have autonomous agents, like football players, with their own avatar. For those agents, the pilot is at the server, and the drones at the clients.

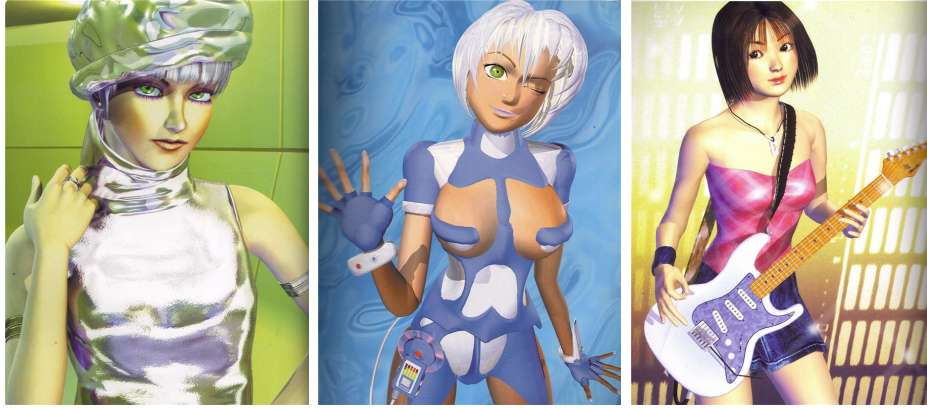
Now, this classification of agents gives us a setup that allows for the realization of shared objects in virtual environments in an efficient manner. See Huang et al. (2002) for details.

The programming platform needed to implement our proposal must satisfy the following requirements.

programming platform

- VRML EAI support
- distributed communication capabilities (TCP/IP)
- multiple threads of control – for multiple shared objects
- declarative language – for agent support

So, we adapted the distributed logic programming language DLP (which in its own right may be called an agent-oriented language *avant la lettre*), to include VRML capabilities. See the online reference to the AVID project for a further elaboration of these concepts.



9

PAMELA The WASP project's chief focus is to develop architectural support for web-aware (multi) agent systems. So, when we (finally) got started with the project we developed a taxonomy along the following dimensions:

taxonomy of agents

- 2D/3D – to distinguish between text-based and avatar embodied agents
- client/server – to indicate where agents reside
- single/multi – as a measure of complexity

A classification along these dimensions results in a lattice, with as the most complex category a *3D-server-multi-agent system*, of which the distributed soccer game is an example. See Huang et al. (2000).

When we restrict ourselves to *3D-client-single-agent systems*, we may think of, for example, navigation or presentation agents, that may help the user to roam around in the world, or that provide support for presenting the results of a query as objects in a 3D scene.

Our original demonstrator for the WASP project was an agent of the latter kind, with the nickname *PAMELA*, which is an acronym for:

PAMELA

Personal Assistant for Multimedia Electronic Archives

The PAMELA functional requirements included: autonomous and on-demand search capabilities, (user and system) modifiable preferences, and multimedia presentation facilities. It was, however, only later that we added the requirement that PAMELA should be able to live in 3D space.

In a similar way as the soccer players, PAMELA has control over objects in 3D space. PAMELA now also provides animation facilities for its avatar embodiment.

To realize the PAMELA representative, we studied how to effect facial animations and body movements following the *Humanoid Animation Working Group* proposal.

H-Anim

- control points – joints, limbs and facial features

The H-Anim proposal lists a number of control points for (the representation of the) human body and face, that may be manipulated upto six degrees of freedom. Six degrees of freedom allows for movement and rotation along any of the X,Y,Z axes. In practice, movement and rotation for body and face control points will be constrained though.

presentation agent Now, just imagine how such an assistant could be of help in multimedia information retrieval.

presentation agent

Given any collection of results, PAMELA could design some spatial layout and select suitable object types, including for example color-based relevance cues, to present the results in a scene. PAMELA could then navigate you through the scene, indicating the possible relevance of particular results.

persuasion games But we could go one step further than this and, taking inspiration from the research field of *persuasive technology*, think about possible persuasion games we could play, using the (facial and body) animation facilities of PAMELA:

persuasion games

- single avatar persuasive argumentation
- multiple avatar dialog games

Just think of a news reader presenting a hot news item. or a news reader trying to provoke a comment on some hot issue. Playing another trick on the PAMELA acronym, we could think of

PAMELA

Persuasive Agent with Multimedia Enlightened Arguments

I agree, this sounds too flashy for my taste as well. But, what this finale is meant to express is, simply, that I see it as a challenge to create such synthetic actors using the DLP+VRML platform.



research directions— *embodied conversational agents*

A variety of applications may benefit from deploying embodied conversational agents, either in the form of animated humanoid avatars or, more simply, as a 'talking head'. An interesting example is provided by *Signing Avatar*, a system that allows for translating arbitrary text in both spoken language and sign language for the deaf, presented by animated humanoid avatars. Here the use of animated avatars is essential to communicate with a particular group of users, using the sign language for the deaf.

Other applications of embodied conversational agents include e-commerce and social marketing, although in these cases it may not always be evident that animated avatars or faces actually do provide added value.

Another usage of embodied conversational agents may be observed in virtual environments such as Active Worlds, *blaxxun* Community and Adobe Atmosphere. Despite the rich literary background of such environments, including Neil Stephenson's *Snow Crash*, the functionality of such agents is usually rather shallow, due to the poor repertoire of gestures and movements on the one hand and the restricted computational model underlying these agents on the other hand. In effect, the definition of agent avatars in virtual environments generally relies on a proprietary scripting language which, as in the case of *blaxxun* Agents, offers only limited pattern matching and a fixed repertoire of built-in actions.

In contrast, the scripting language for *Signing Avatar* is based on the H-Anim standard and allows for a precise definition of a complex repertoire of gestures, as exemplified by the sign language for the deaf. Nevertheless, also this scripting language is of a proprietary nature and does not allow for higher-order abstractions of semantically meaningful behavior.

scripting behavior In this section we introduced a software platform for agents. This platform not only offers powerful computational capabilities but also an expressive scripting language (STEP) for defining gestures and driving the behavior of our humanoid agent avatars.

The design of the scripting language was motivated by the requirements listed below.

STEP

- *convenience* – for non-professional authors
- *compositional semantics* – combining operations
- *re-definability* – for high-level specification of actions
- *parametrization* – for the adaptation of actions
- *interaction* – with a (virtual) environment

Our scripting language STEP meets these requirements. STEP is based on dynamic logic Harel (1984) and allows for arbitrary abstractions using the primitives and composition operators provided by our logic. STEP is implemented on top of DLP,

As a last bit of propaganda:

DLP+X3D

The DLP+X3D platform provides together with the STEP scripting language the computational facilities for defining semantically meaningful behaviors and allows for a rich presentational environment, in particular 3D virtual environments that may include streaming video, text and speech.

See appendix D for more details.

evaluation criteria The primary criterium against which to evaluate applications that involve embodied conversational agents is whether the application becomes more effective by using such agents. Effective, in terms of communication with the user. Evidently, for the *Signing Avatar* application this seems to be quite obvious. For other applications, for example negotiation in e-commerce, this question might be more difficult to answer.

As concerns the embedding of conversational agents in VR, we might make a distinction between *presentational VR*, *instructional VR* and *educational VR*. An example of educational VR is described in Johnson et al. (2002). No mention of agents was made in the latter reference though. In instructional VR, explaining for example the use of a machine, the appearance of a conversational agent seems to be quite natural. In presentational VR, however, the appearance of such agents might be considered as no more than a gimmick.

Considering the use of agents in applications in general, we must make a distinction between *information agents*, *presentation agents* and *conversational agents*. Although the boundaries between these categories are not clearcut, there seems to be an increasing degree of interactivity with the user.

From a system perspective, we might be interested in what range of agent categories the system covers. Does it provide support for managing information and possibly information retrieval? Another issue in this regard could be whether the system is built around open standards, such as XML and X3D, to allow for the incorporation of a variety of content.

Last but not least, from a user perspective, what seems to matter most is the naturalness of the (conversational) agents. This is determined by the graphical quality, as well as contextual parameters, that is how well the agent is embedded in its environment. More important even are emotive parameters, that is the mood and style (in gestures and possibly speech) with which the agents manifest themselves. In other words, the properties that determine whether an agent is (really) convincing.

8.4 development(s) – the metaverse revisited

In less than a decade after the publication of William Gibson's novel *Neuromancer*, the *metaverse* was realized, albeit in a primitive way, through the introduction of VRML⁹⁷, introduced at the Int. Web Conference of 1992. Cf. Anders (1999). The German company *blaxxun*⁹⁸, named after the virtual environment in Neil

⁹⁷www.web3d.org

⁹⁸www.blaxxun.com

Stephenson's *Snowcrash*, was one of the first to offer a 3D community platform, soon to be followed by *AlphaWorld*, already mentioned in the introduction, which offered a more rich repertoire of avatar gestures as well as limited in-game building facilities. However, somehow 3D virtual communities never seemed to realize their initial promises. Furthermore the adoption of VRML as a 3D interface to the Web never really took off.

The history of Second Life is extensively described in the official Second Life guide, SecondLife. Beginning 2004, almost out of the blue, *Second Life*⁹⁹ appeared with a high adoption and low churn rate, now counting, March 2007, over 4 million inhabitants. Considering the cost of ownership of land, which easily amounts to 200 euro per month rent after an initial investment of 1500 euro for a single piece of land measuring 65,536 square meters, the adoption of Second Life by individuals as well as companies such as ABN-AMRO, Philips and institutions such as Harvard is surprising.

What is the secret of the success of Second Life? We don't know! But in comparison to other platforms for immersive worlds, including MMORPGs such as *World of Warcraft*¹⁰⁰ and *Everquest*¹⁰¹, Second Life seems to offer an optimal combination of avatar modification options, gesture animations, in-game construction tools, and facilities for communication and social networking, such as chatting and instant messaging. Cf. Utz. Incorporating elements of community formation, commonly denoted as Web 2.0, and exemplified in MySpace, YouTube and Flickr, the immersive appearance, perhaps also the built-in physics and the inclusion of elementary economic principles, seem to be the prime distinguishing factors responsible for the success of Second Life. In addition, the possibility of recording collaborative enacted stories, Stories, using built-in *machinima*¹⁰² certainly also contributes to its appeal. Later on, after discussing Second Life from a more technical perspective, we will speculate further on the possible reasons for the success and adoption of Second Life as a platform for communication and immersive presence.

What has been characterized as a shift of culture, from a media consumer culture to a participatory culture, Participatory, where users also actively contribute content, is for our institution one of the decisive reasons to create a presence in Second Life, to build a virtual platform that may embody our so-called *community of learners*, where both staff and students cooperate in contributing content, content related to our sciences, that is.

BUILDING A VIRTUAL CAMPUS

In December 2006, we discussed the idea of creating presence in Second Life. Our initial targets were to build a first prototype, to explore content creation in Second Life, to create tutorials for further content creation, and to analyze technical requirements and opportunities for deployment in education and research.

⁹⁹secondlife.com

¹⁰⁰www.worldofwarcraft.com

¹⁰¹everquest.station.sony.com

¹⁰²www.machinima.org

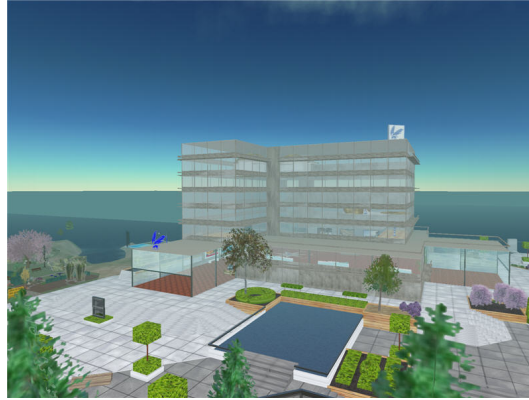


Fig 1. VU Campus – outside view

Two and a half months later, we are online, with a virtual campus, that contains a lecture room, a telehub from which teleports are possible to other places in the building, billboards containing snapshots of our university's website from which the visitors can access the actual website, as well as a botanical garden mimicking the VU Hortus, and even a white-walled experimentation room suggesting a 'real' scientific laboratory. All building and scripting were done by a group of four students, from all faculties involved, with a weekly walkthrough in our 'builders-meeting' to re-assess our goals and solve technical and design issues.

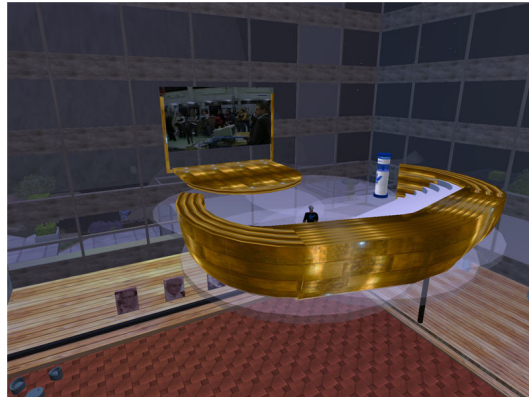


Fig 2. VU Campus – inside view

The overall style is realistic, although not in all detail. Most important was to create a visual impression of resemblance and to offer the opportunity to present relevant information in easily accessible, yet immersive, ways. Cf. Bolter and Grusin (2000), *Reality*. Our virtual campus, see figs. 1 and 2, is meant to serve as an *information portal* and as a *meeting ground*, where students, staff and visitors can meet and communicate, as well as a place where teachers and researchers can conduct experiments aimed at discovering new ways of teaching and doing research.

SECOND LIFE SUCCESS FACTORS

Will Second Life become the new (immersive 3D) mass medium of our participatory culture of the 21th century, as once the immersive *panorama* was the propaganda/art medium for the masses in the 19th century? Cf. Grau (2003). In thinking about possible reasons why Second Life is so successful, we observed that Second Life does provide:

- convergence of social networking and content creation
- immersive networked 3D environment
- inclusion of elementary economic principles

However, we also see that other factors may contribute to the success of Second Life, such as:

- don't miss the boat effect
- free and easy accessible 3D design tool set
- adoption by big companies like IBM, Reebok, ...
- marketing of Second Life by Linden Lab (?)
- the promise to make (real) money (?)

According to Philip Rosedale, CEO of Linden Lab, (interview in .NET magazine, issue 158, January 2007) the success of SL is due to the fact that (1) it offers a set of capabilities, which are in many different ways superior to the real world, (2) the decision to allow residents to own the intellectual property rights to their creations and (3) because Second Life is full of creative possibilities, and opportunities for innovation.

In order to establish what constitutes the success of Second Life in a more rigorous manner, we must subject Second Life to a (game) *reference model* as introduced in Juul (2005), which we have also applied to (serious) service management games in Serious. A first tentative characterization of Second Life according to our reference model would be:

reference model

- *rules* – construct and communicate!
- *outcome* – a second world
- *value* – virtual and real (monetary) rewards
- *effort* – requires elementary skills
- *attachment* – a virtual identity
- *consequences* – transfer to first life

Second Life clearly has a wider scope and more freedom than just gaming. Apart from elementary rules, that more or less require of the (serious) visitor to *construct and communicate*, there are almost no fixed rules, no in-game strategies, but many opportunities for inter-personal contact and the establishment of relations world-wide, affecting (possibly) the Second Lifer's first life (*consequences*).

Whether Second Life will turn out to be a veritable media-supported augmentation of our first life, cf. Zielinski (2006), remains to be seen. Chances are also that Second Life will end up as another item on the *dead media projects*¹⁰³ list, to be replaced by an alternative participatory framework or environment.



Fig 3. VU @ SL – visitors outside

HOLD YOUR BREATH – GOING LIVE

The 1st of March 2007, we went live. In the evening there was a news item on national television, RTL4 news, featuring the students showing the virtual campus and our project leader explaining the reasoning behind our presence in Second Life and how to give a course in the virtual classroom. A similar item appeared at AT5, local Amsterdam television, and various newspapers, among which Parool, Telegraaf and Volkskrant, spent a multiple-column article to report on our efforts. As a note, not surprisingly, all items focused on what we have characterized as the naive interpretation of our efforts, exemplifying the old credo *the medium is the message*. To be clear, our intention is not to provide a virtual replica, nor to provide an analogon of the Open University, in Second Life.

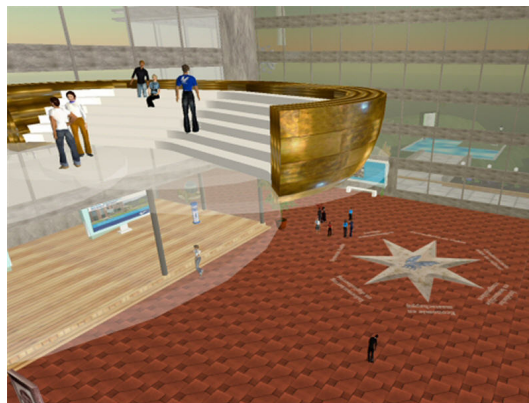


Fig 4. VU @ SL – visitors inside

¹⁰³www.cs.vu.nl/~eliens/media/project-deadmedia.html

After the news broadcasts, the number of visitors increased dramatically, having stayed at a modest below 100 during the day, see figs. 3 and 4. In the evening, however, just after the news items on the national television, the number of visitors increased rapidly. Since, presently, we do have only one island it appeared to be very difficult to separate internal experimental activities from visitors just asking for additional information, and to exclude potentially malicious visitors. In that evening, we were even surprised by the invasion of an army of Mario Brothers. Hilarious and non-harmful. But enough reason to sit back and limit access to our campus for students and staff only the day after our open day. A few days later, after the first turbulent days following the TV broadcasts, we re-opened our virtual campus to allow visitors to walk/fly around, and enjoy our news items and informative videos. So far, the results exceeded our expectations, the students were praised for the results of their building efforts, and as a team we may continue to think about how to deploy Second Life as a platform for education and research projects.



11

questions

virtual environments

1. (*) Discuss how *virtual environments* may be used for giving access to (*multimedia*) *information*. Give a brief characterization of *virtual environments*, and indicate how *information (hyper) spaces* may be projected in a virtual environment.

concepts

2. What is meant by *virtual context*?
3. Give an example of *navigation by query*, and indicate its possible advantages.
4. Discuss the deployment of (*intelligente*) *navigation agents*.

technology

5. Give a brief characterization of: VRML.
6. What is a *viewpoint transformation*?
7. What kinds of navigation can you think of?
8. How may intelligent avatars be realized? Give an example.

projects & further reading As a project, I suggest the implementation of storytelling in virtual environments, with (possibly) an embodied agent as the narrator. You may further explore or evaluate the role of agents in multimedia applications and virtual environments.

For further reading in (real) VR I advice Sherman and Craig (2003), and for gaining an understanding in story telling and applications you may try to get hold of the proceedins, of TIDSE 2003¹⁰⁴, and TIDSE 2004¹⁰⁵.

the artwork

1. another series of *dutch light*¹⁰⁶.
2. *virtual context* – Dam Square, Amsterdam, see 8.1.
3. VU Campus in VRML – student project.
4. CWI 3th floor, floormap and model, see 8.2..
5. query – on 3th floor of CWI.
6. navigation – on 3th floor of CWI.
7. soccer game – image from WASP project, see section 8.3.
8. *digital beauties* – taken from Wiedermann (2002).
9. *digital beauties* – taken from Wiedermann (2002).
10. signs – sports, van Rooijen (2003), p. 276, 277.

Another sequence of *dutch light*, opening this chapter, is meant to make you wonder about *realism*. Is virtual reality less 'real'? With a reference to section 2.3, where I quoted Bolter and Grusin (2000) on *re-mediation*, I may remark that the graphic style chosen for presenting the virtual environment strongly determines whether the environment is experienced as 'realistic'. In our culture this is generally a *photorealistic style*, as for example in the *Mission Rehearsal Exercise* discussed in the next chapter, section 9.2. The *digital beauties* are not only a pleasure to look at, but do also display a wide range of postures and styles.

¹⁰⁴www.zgdv.de/TIDSE03

¹⁰⁵www.zgdv.de/TIDSE04

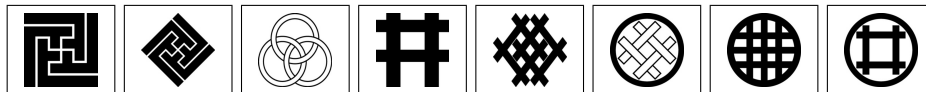
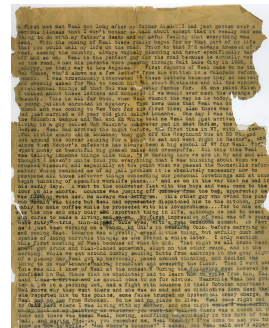
¹⁰⁶www.dutchlight.nl

part v. game development

man is a playful animal
johan huizinga

chapters:

- 11. game technology for serious applications
- 12. towards an aesthetics for interaction



2

reading directives In the final two chapters of this book, we will look at game development, with an emphasis on both implementation and design. In particular, we will discuss the motivation for using game technology to develop serious application, which in summary may be characterized as the requirement to allow for *immersion*, understood as experiential involvement of the user.

Essential sections are 11.2, in which we describe the development of VU-Life using the Source Half-Life 2 SDK, and 12.1, which presents a semi-formal game model, that may serve as a reference for designing more complex games.

perspectives Game development is essentially teamwork, and generally involves both artists as well as more technical developers. In addition, with to the commercial impact of games, many factors influence the design and implementation

of games. As a consequence, we may look at game development from (among others) the following perspectives:

perspectives – game development

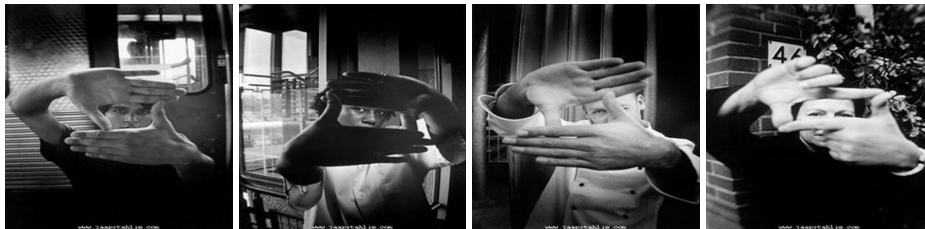
- artistic – plot, narrative, style
- technical – choice of game engine (SDK)
- sociological – sharing within game communities
- tool selection – supporting the workflow
- commercial – success factors
- management – teamwork

It is only very recently that game development became a topic with academic credibility. From an academic perspective, game technology may be regarded as enabling technology, that is related to many areas of computer science, including graphics, computer organization, distributed systems and software engineering.

essay topics In accordance with the variety of perspectives, an essay may focus on the commercial aspects of games, or, for example, software engineering aspects, or the design of community games. Consider writing an essay about:

- the success of games – criteria for comparison
- game engine architecture – options for extensibility
- community games – models for sharing information

When writing, make sure that you provide enough information about the actual history of games. And, remind that games require real-time performance, whereas the movies that may have triggered your phantasy may get away with effects that require many hours of processing. Afterall, games are about interaction!



3

the artwork

1. manuscript – used as a desktop by my favorite student.
2. signs – abstract symbols, van Rooijen (2003), p. 214, 215.
3. photographs – Jaap Stahlie¹⁰⁷, commissioned work.

¹⁰⁷ www.jaapstahlie.com

afterthought(s)

The world of multimedia may be looked at in many ways. In fact, the phrase *multimedia* is too generic to be meaningful in any way. Nevertheless, multimedia has become a subject of interest for academia. This book has been written from an academic perspective. Let me clarify this perspective, to provide you with some context that might help you in understanding this book and use it more effectively in either education, research, or even your artistic endeavors.

As a starting point, let's look (again) at the *media equation*, quoted in the *research directions* of section 9.1:

media equation(s) 1/4

We regularly exploit the media equation for enjoyment by the willing suspension of our critical faculties. Theatre is the projection of a story through the window of a stage, and typically the audience gets immersed in the story as if it was real.

This suspension of our critical faculties seems opposed to what we are used to in academic practice. And, indeed, there is an often noted conflict between the arts and the sciences, a conflict that the introduction of multimedia in the academic curriculum cannot resolve.

If we try to delineate the 'meaning' of multimedia more precisely, we might come up with pseudo-equation such as the following.

multimedia equation(s) 2/4

$$\text{multimedia} = \text{presentation} + \text{context}$$

where *presentation* includes the sensory and aesthetic part and *context* everything else. Now, at the risk of getting too much involved in 'funny mathematics' we might define *context* by another series of pseudo-equations

multimedia equation(s) 3/4

- context = convergence + information + architecture

where

multimedia equation(s) 4/4

- convergence = data + platform + distribution
- information = storage and retrieval
- architecture = compression + components + connectivity

Clearly, and this is exactly what this exercise in funny mathematics intended to illustrate, this book is about the contextual aspects of multimedia. Contextual aspects that may be the subject of academic research.

Is there any hope to include the presentational or aesthetic aspects in the academic curriculum? Based on a thought experiment, that explored the possibility of algorithmic art and aesthetics, Eliens (1988), I would say no. And as a matter of fact, I strongly disagree with a recipe-based approach to developing multimedia presentations, as seems to be popular in a number of the academic multimedia courses.

There is another shade of meaning that may be attributed to the notion of *context*, namely context of application. Evidently, multimedia has become a natural ingredient of almost any application you can think of. In 1998, I organized a course on multimedia for Ph.D. students, entitled *Multimedia in Context*. This course dealt with some of the issues in distributed multimedia and multimedia information retrieval, as well as applications in the publishing industry, travel advertisement and medical diagnosis. To announce the course, I used an image from medieval alchemy, see part I, and a phrase characterizing 'perfect solutions'.

perfect solutions

Much more than the art of turning base metals into gold, alchemy is a system of cosmic symbolism. The alchemist learns how to create within a sealed vessel a Model of the Universe in which the opposing complementary forces of Male and Female, Earth and Air, Fire and Water attain the perfect synthesis of which gold is the emblem.

Risking obscurity at this point, I wish to equate multimedia with alchemy, to emphasize that the engineering of multimedia is an art that takes a lifetime to master. Repeating the quote from section 9.1:

multimedia engineering

"engineering is the art of moulding materials we do not wholly understand ... in such a way that the community at large has no reason to suspect the extent of our ignorance."

multimedia in context

Originally the book, that is chapters 1-7, were written for the *Multimedia and Culture* curriculum at the Vrije Universiteit, Amsterdam, that started in 2001. In particular, the book contains the course notes for the first year course *introduction multimedia*.

Then I constructed four follow-up courses:

multimedia courses¹⁰⁸

- multimedia authoring – Web3D/VRML
- intelligent multimedia technology – Virtual Environments

¹⁰⁸www.cs.vu.nl/~eliens/multimedia

- visual design – digital content creation
- multimedia casus – digital dossier(s)

The first of these courses deals with the technology for creating 3D scenes and worlds see appendix B, whereas the second is about providing intelligent services in virtual environments, as discussed in chapter 8 and appendix E. In addition, *Multimedia and Culture* students are required to take a course in *visual design* and to work on a *multimedia casus* to bring what they learned into practice. The structure and content of these courses are reflected in chapters 9 and 10.

Due to faculty politics, the *Multimedia and Culture* curriculum was reduced to a minor in Information Science, which made it less appealing, both for students and staff, including me. Over time, all the course mentioned above were repurposed for the newly created specialisation *multimedia* in Computer Science, which attracts more technically oriented students, with better skills for actual multimedia and game application development. Although not technical in nature, chapters 11 and 12 were written with these students in mind. And very likely, or hopefully, the specialisation *multimedia* will soon become *multimedia and game development*.

Faculty politics is not a nice topic, but unfortunately has an effect on daily life, even to the extent that I sometimes regret that I gave up the, within an academic environment, relatively safe and simple discipline of software engineering and object-oriented software development. The truth of the matter is that, whatever the reasons, multimedia and game development does not fit in well in the standard academic context of computer science. Although liked by students, it is certainly not well accepted, and for that matter understood, by the senior staff. But although anecdotes about the many conflicts about research directions and scientific merit would be interesting for those who like gossip, the sad fact is that the multidisciplinary background of multimedia and game development would require an intellectually and artistically more rich environment than a department of computer science can offer.

explorative development

From the perspective of research, the situation is not much better. It is my strong belief, right or wrong, that relevant research in the area of multimedia and game development requires *explorative development*, that is the design and implementation of prototype applications that embody the realization of an idea, as with our research on the *digital dossier(s)*, an idea that includes technical as well as cultural and presentational aspects. But how hard it is to perform such multi-disciplinary research in an environment that is by tradition pre-dominantly mono-disciplinary.

Back to the book, apart from providing an introduction to a number of issues and research areas in the world of multimedia, this book also defines, in an implicit way, a research program that concerns the development and use of

*virtual reality interfaces for multimedia information systems*¹⁰⁹

¹⁰⁹www.cs.vu.nl/~eliens/research

All aspect covered in this book contribute, one way or another, to that (implicit) research program that may be classified under the heading of *intelligent multimedia*, of which a tentative definition is given in appendix E. And, admittedly, there are many aspects that are not covered, in particular those that are related to more advanced multimedia, virtual reality technology, and artificial intelligence.

the skill(s) of relevance

At this stage you may still wonder why I chose to name the book *topical media*. Let me explain. The phrase *topical*, as an adjective, has the following meanings:

topical

1. designed for involving local application (as an anesthetic),
2. relating to, or arranged by, topics,
3. referring to the topics of the day or place.

Although certainly not meant as an anesthetic, and even though it is arranged by, or at least refers to, topics, the intended meaning is due to the motivation to write a book that is relevant for the topics of (interest of) today's world. And, although it may not teach you the actual skills necessary to survive in today's world of multimedia and game development, it is certainly meant to help you in acquiring the *skill of relevance*, see Bruner (1972), in this area, a skill that you will need to find your proper place and direction, anytime, anywhere.

Amsterdam, 9/8/2006

A. Eliëns

appendix

D. resources, tools and technology

What do you need to have to start working on your multimedia project? that depends, naturally, on what you want to do. In the following, I will give a brief overview of resources, tools and technologies that you might find useful or that you might want to explore. This overview consists mainly of urls and a brief characterization and in some cases an indication of a price range.

This overview is definitely not meant to be complete, and is only included for your convenience, so that you don't have to *google*¹¹⁰ it yourself. In the online version of the book more (online) resources are given, as well as a (clickable) list of all urls that appear (as a footnote) in the book.

resource(s)

This section contains a variety of itmes, including a selection of online tutorials and thesauri. Some examples are given of online museum tours and listings are included of the media art and cultural heritage institutes mentioned in the book. But we will start with introducing briefly with what you need for 3D authoring and rendering, since this is what we have primarily focused on in this book.

3D authoring & conversion

- vrmtpad – www.parallelgraphics.com/products/vrmtpad
- polytrans – www.okino.com/products.htm
- maya – www.alias.com
- 3dsmax – www.discreet.com
- sketchup – sketchup.google.com/download.html
- flux studio – www.mediamachines.com/products.html

The *polytrans* tool from Okino has been included, since it allows you to convert almost any format into you format of choice, which is a great asset for (re) using models.

¹¹⁰www.google.com

3D rendering

- blaxxun – www.blaxxun.com/en/products/contact
- virtools – www.virtools.com
- flux web3d – sourceforge.net/projects/flux
- mediamachines flux – www.mediachines.com/products.html

As concerns price, VRML-based solutions for authoring and rendering are clearly low-cost, whereas tools such as *Maya* and *Studio Max* require more investment, not only in money but also in learning time. Also *Virtools* is in the higher price range.

tutorials

- html – www.mcli.dist.maricopa.edu/tut
- javascript – www.javascriptkit.com
- php – www.php.net/docs.php
- rdf – www.w3.org/TR/rdf-primer
- vrml – web3d.vapourtech.com/tutorials/vrml97
- java – java.sun.com/docs/books/tutorial
- 3D modeling – www.raph.com/3dartists/tutorials/t-3dsmax.html
- games in VRML – www.3dezine.com/3DEZine/gamestory.html
- ria – www.macromedia.com/resources/business/rich_internet_apps/whitepapers.html

In many cases it is (more) convenient to have working examples at hand. Personally, I advice my students to learn using HTML, VRML, Javascript and the like from one of the online tutorials, which do provide such examples. The *php* documentation is not really a tutorial but does provide useful help and examples.

visual design

- collage – www.artlex.com/ArtLex/c/collage.html
- storyboard – www.thestoryboardartist.com/links.html
- drawing – www.thestoryboardartist.com/tutorial.html

For *visual design* it might be worthwhile to look at some examples, or even take a complete course in drawing.

museum

- van gogh – www.vangoghmuseum.nl
- rijksmuseum – www.rijksmuseum.nl
- canada – www.virtualmuseum.ca/English/index_flashFT.html
- zkm – www.zkm.de
- tate – www.tate.org.uk
- louvre – www.louvre.fr

More inspiration can perhaps be obtained from looking at what musea have to offer. It also gives you an opportunity to update your knowledge of the history of art.

media art

- montevideo – www.montevideo.nl
- V2 – www.v2.nl
- electronic arts intermix – www.eai.org/eai
- cinemanet – www.cinemaneteurope.com
- variable media – www.variablemedia.net
- net art – www.jodi.org/100cc/index.html
- mediamatic – www.mediamatic.net

Listed above are institutions that play a role in the preservation and dissemination of contemporary media art. Not an institution, but an early pioneer of art on the internet, is *jodi* from *net art*.

virtual tours

- amsterdam – www.channels.nl
- panoramic amsterdam – www.panoramsterdam.nl
- rijksmuseum – www.rijksmuseum.nl/collectie/meesterwerken/?lang=en
- groningen – www.kalamiteit.nl/world/no_cache/museum/vrml/connect.html
- mondriaan – www.artmuseums.harvard.edu/mondrian

Many cities nowadays have virtual tours. And also many musea allow the (online) visitor to have a look at their collection.

cultural heritage

- incca – www.incca.org
- delos – www.delos.info
- echo – echo.mpiwg-berlin.mpg.de/home
- eu – www.iue.it/ECArchives
- cidoc – www.cidoc.icom.org
- collate – www.collate.de
- cimwos – www.xanthi.ilsp.gr/cimwos
- library of congress – www.loc.gov/
- scriptorium – sunsite.berkeley.edu/scriptorium
- tei – www.tei-c.org
- open archives – www.tei-c.org
- topia – topia.telin.nl

Above is a mixed collection of references to organizations and institutions that are in some way involved in cultural heritage projects, either related to traditional art or contemporary art.

thesaurus

- webopedia – www.webopedia.com
- visual – www.visualthesaurus.com
- 3D glossary – www.nvidia.com/page/pg_20010527107687.html
- art & architecture – www.getty.edu/research/conducting_research/vocabularies/aat/
- modern art – en.wikipedia.org/wiki/Modern_art
- (new) media art – en.wikipedia.org/wiki/New_Media_art
- art online – www.art-online.com
- multimedia – www.insead.fr/CALT/Encyclopedia/Media/multimedia.html
- virtual reality – www.insead.fr/CALT/Encyclopedia/ComputerSciences/VR
- gaming – www.insead.fr/CALT/Encyclopedia/ComputerSciences/Gaming
- mathematics – www.cs.brown.edu/people/scd/facts.html
- mpeg – www.m4if.org/mpeg4
- wikipedia – en.wikipedia.org/wiki/Multimedia

There is a wealth of online information sources, including glossaries and thesauri. Beware, not all of them are properly authorized. Nevertheless, it might be interesting to note that the online version of this book is referred to in the *wikipedia*, for the entry *multimedia*.

games

- gamasutra – www.gamasutra.com
- gamedev – www.gamedev.net
- developer – www.gdmag.com/resources.html
- and more – www.lostlogic.com/postnuke
- games at school – www.freewebs.com/schoolgamemaker
- gamemaker – www.gamemaker.nl/
- game learning – www.gamelearning.net
- scripting – <http://www.lua.org>
- open source – www.delta3d.org
- free source – www.thefreecountry.com/sourcecode/games.shtml

For games, there are several popular sites providing information about new upcoming games, as well as developer's resources, including software available for download.

A recommended open source game engine is *Delta3D*. This package contains a variety of open source software, well-integrated due to the efforts of a dedicated team at the Naval Postgraduate School in Monterey, CA/USA.

serious games

- play2learn – www.play2learn.nl
- nitrogenius – www.serc.nl/play2learn/products/nitrogenius
- at school – rla.oakland.edu/~ist_699
- primary games – www.primarygames.com

- games at school – www.freewebs.com/schoolgamemaker
- arcade – www.educationarcade.org
- never winter – nwn.bioware.com

Serious games are a new brand of games. Not really new in terms of technology, but new with respect to focus and intent.

tool(s)

There is a great variety of tools, with huge differences in prize. Often, however, you can download a fully functional trial version that will last for a month, and thus may determine the length of your project. A number of tools, however, come with a free (such as Maya) or limited price (such as 3DSMax) student version.

imaging and graphics

- photoshop – www.adobe.com/products/photoshop
- illustrator – www.adobe.com/products/illustrator
- snagit – www.techsmith.com/products/snagit
- camtasia – www.techsmith.com/products/studio

Perhaps the most popular tools among designers are *photoshop* and *illustrator*. Both for capture and image catalogue maintenance I have benefited from *snagit* and *camtasia*, both from *techsmit*.

3D modeling

- vrmlpad – www.parallelgraphics.com/products/vrmlpad
- polytrans – www.okino.com/products.htm
- maya – www.alias.com
- 3dsmax – www.discreet.com
- houdini – www.sidefx.com
- bodystudio – www.reiss-studio.com
- poser – www.curious-labs.com

In addition to the modeling tools already mentioned before, there are many additional tools and add-ons, such as *houdini* for procedural modeling, *bodystudio* for importing poser models in maya, 3dsmax and other tools, and *poser*, a somewhat outdated tool voor modeling humanoids, with a large collection of ready-made feature material.

Alias Wavefront Maya

- information – www.alias.com
- tutorials – www.alias.com/eng/community/tutorials
- community – www.alias.com/eng/community

A high end 3D modeling tool, with a respectable history and a large community of users. It is in the high end price range and requires significant effort to master.

Discreet 3D Studio Max

- information – www.discreet.com
- tutorials – www.pixel2life.com/tutorials/3dsmax.php?tut=16
- vrml – www.dform.com/inquiry/tutorials/3dsmax

Popular within the game community, *studio max* which includes *character studio* appears to be somewhat more straightforward than maya.

technology

Again, the technology overview is certainly not exhaustive. There are many commercial game engines that are well worth looking at when you engage in a real project. I have included a limited number of open source libraries and toolkits to provide you with a starting point for further exploration.

DirectX SDK 9

- information – www.microsoft.com/directx
- show + 3d – msdn.microsoft.com/library/default.asp?url=/library/en-us/dnwmnt/html/vmr_d3d.asp
- SDK – msdn.microsoft.com/library/default.asp?url=/library/en-us/directx9_c/directx/directx9cpp.asp
- frames – www.jkarlsson.com/Articles/loadframes.asp
- animation controller – www.jkarlsson.com/Articles/animation.asp

Direct X is an advanced, yet complicated multimedia platform. The managed code version is significantly less powerful than the C++ version. As indicated in section 4.2 there is a great many of books about DirectX. Some helpful online tutorials are listed above.

Wild Tangent

- information – www.wildtangent.com
- developers – www.wildtangent.com/developer

Wild Tangent is very appropriate for developing games. It provides convenience layer around DirectX 7, and enables applications to be run via a Web browser, by deploying the COM interfaces for DirectX. It allows for authoring content and dynamics in Javascript and Java. However, also the original X meshes, the file format for DirectX, can be used.

Virtools Software Suite

- information – www.virttools.com

Virtools is announced to be a *comprehensive development platform, for games, virtual reality/simulations and marketing/multimedia applications*.

OpenML

- information – www.khronos.org/openml

OpenML might be the candidate platform for those that wish to develop platform-independent (read non Microsoft windows-tied) multimedia applications. It is a *royalty-free, cross-platform programming environment for capturing, transporting, processing, displaying, and synchronizing digital media - including 2D/3D graphics and audio/video streams*. *OpenML 1.0 defines professional-grade sample-level stream synchronization,*

OpenGL extensions for accelerated video processing, the MLdc professional display control API and the ML framework for asynchronous media streaming between applications and processing hardware.

open source technology

- plib – plib.sourceforge.net
- OpenSceneGraph – www.openscenegraph.org
- OpenSound – www.cnmat.berkeley.edu/OpenSoundControl
- ARToolkit – artoolkit.sourceforge.net
- Mixed Reality Toolkit – www.cs.ucl.ac.uk/staff/rfreeman
- OpenNap – opennap.sourceforge.net
- ImageMagick – www.imagemagick.org
- *cygwin* – www.cygwin.com

There are many open source software toolkits and libraries. My experience with these is mixed. Anyway, when you start working with these make sure that you have sufficient programming skills, and patience. But then the results might be better than you could obtain with more expensive commercial technology. If you run Linux, then open source is probably the only way to go. For windows users, with a unix background, there is *cygwin*, which is a linux-like environment for windows.

XML

- XML Entities – tech.irt.org/articles/js212
- W3C – www.w3.org/Style/XSL
- resources – www.xml.org/xml/resources_cover.shtml
- saxon – saxon.sourceforge.net
- online tutorial – www.zvon.org/HTMLOnly/XSLTutorial/Books/Book1/index.html
- Xena XML editor – www.alphaworks.ibm.com/tech/xena
- X3D Edit setup – sdk.web3d.org/spring2002disk2/tools/X3D-Edit/index.html

For XML there is a number of generic editors, such as Xena, which has been adapted for X3D, see appendix B. There are also XSLT processing tools, such as *saxon*, which is the only one I have experience with.

Java

- information – <http://www.javasoft.com>
- art with Java – <http://java.khm.de>
- java media framework – <http://java.sun.com/products/java-media/jmf/2.1.1/guide/JMFTOC.html>
- slide show – <http://developer.java.sun.com/developer/technicalArticles/Threads/applet/index.html>
- basics – <http://developer.java.sun.com/developer/onlineTraining/Programming/BasicJava1/compile.html>
- tutorial – <http://java.sun.com/docs/books/tutorial/index.html>
- advanced – <http://developer.java.sun.com/developer/onlineTraining/Programming/JDCBook/>
- sound API – <http://java.sun.com/products/java-media/sound/samples/JavaSoundDemo>
- imaging – <http://developer.java.sun.com/developer/technicalArticles/Media/AdvancedImage>

Java is the programming language of choice for many computer science curricula. It is a well-documented, relatively easy to use language and the java framework provides a rich collection of libraries. There is also a version for mobile platforms.

student multimedia facilities

To conclude this overview of resources, tools and technologies, I have included a brief description of the student facilities we have for multimedia work at the Vrije Universiteit, spring 2005.

computers 14 fujitsu siemens scenico P320, AMD64 3400+ MHz, 1G memory, 80 GB serial ATA disk, 6 x USB, XFX Geforce 6600 GT 128 Mb AGP, dual display, 2 LCD monitors.

software

- Parallel Graphics VrmIPad – site license
- Alias Maya Complete (5.0 & 6.0) – 10 floating licenses
- 3D Studio Max 7 – 15 floating licenses
- DirectX9c SDK – <http://www.microsoft.com/directx>
- WildTangent WebDriver & SDK – <http://www.wildtangent.com/developer>
- CG Toolkit – <http://developer.nvidia.com/page/tools.html>
- RenderMonkey & SDK – <http://www.ati.com/developer/rendermonkey>
- Illustrator & Photoshop CS – <http://www.adobe.com>

There is no need to say that this is not the end of the story. More software will be installed, among which *virttools*, hopefully soon. And whenever the opportunity is there, we will no doubt upgrade to more powerful hardware as well!

E. write an essay!

Even when you prefer to do practical work, it might well pay off to take a step back, reflect on your approach and study one aspect of multimedia in more detail. When you plan to work in an academic situation, it is very likely that at some point you must report about your work and provide some theoretical context to it. These few closing paragraphs are meant to give you some hints about how to approach writing a paper or report.

Independent of how you tackle the process of collecting material, organizing notes and writing it all down, keep in mind that the end result must consist of:

outline

title – *indicating the topic*
name – *to tell who you are*
abstract – *giving the 'message' of your efforts*
introduction – *clarifying the approach and structure*
background – *explaining the context of the subject*
sections – *to elaborate on the subject*
related work – *characterizing related approaches*
conclusion(s) – *summarizing the main point(s)*
references – *listing the literature you consulted*
appendices (optional) – *providing extra information*

It is surprising how often students forget, for example, an abstract or a proper introduction. Often the familiarity with the material, built up when working with it, seems to make them forget that for the reader these items are important and cannot be missed to grasp the point(s) of their efforts. Also, I wish to note that, although the discipline of giving references is in computer science much less strict than in, for example, philosophy, sufficiently clear references are necessary for the reader to check and verify your claims.

As I already indicated I do not wish to elaborate on how to gather material, how to organize your collection of potentially useful notes, or how to convert these notes into readable text. Rather, I wish to discuss the distinction, or tension, between form and content. Form, I would say, is determined by the perspective from which you approach the material and the goal you set yourself when writing the paper or report. Possible perspectives, or if you prefer forms, are:

perspective(s)

- review/background – *sketch perspectives, history, viewpoints*
- case study – *analyse assumptions, gather empirical data, and explain!*
- technical analysis – *technology-oriented, work out the details*

- formal study – *clarify in a formal manner, conceptualize and formalize*
- tutorial – *explain for the laymen, but do it very good*

To be clear, the phrase perspectives as used here is only vaguely related to the use of perspectives when used to introduce the parts, where it meant to indicate the scientific discipline or point of view from which to look at a particular topic.

Content, as opposed to form, may be characterized as the collection of possible subjects, which in the area of multimedia include authoring, digital convergence, standards and information retrieval. Obviously, some subjects are better matched with particular forms or perspectives than others. For example, a formal study is suitable for discussing standards, but, to my mind, less so for explaining multimedia authoring. To get an idea of how I look at the problem of reconciling form and content when writing a paper about multimedia, consult the matrix:

	authoring	convergence	standards	retrieval
review/background	-	++	++	+
case study	+	+	+	+
technical analysis	-	++	++	++
formal study	-	-	++	-
tutorial	-	-	?	-

You may wonder why I don't think of tutorials as a suitable form for writing about multimedia. Well, in fact I do think that the form of a tutorial is an excellent way to write about multimedia technology, but it is not a very rewarding form for getting academic credits. When you want to be an academic, you'd better learn to write a technical analysis or case study. However, by that time perhaps the scientific paper generators¹¹¹ might have matured to the extent that writing has become a superfluous activity.

¹¹¹ www.pdos.lcs.mit.edu/scigen

references

- Adams J. (2002), *Programming Role Playing Games with DirectX*, Premier Press
- Adams J. (2003), *Advanced Animation with DirectX*, Premier Press
- Alexander T., ed. (2005), *Massively Multiplayer Game Development 2*, Charles River Media
- Anders P. (1999), *Envisioning Cyberspace – Designing 3D Electronic Spaces*, McGraw-Hill
- Angel E. (1997), *Interactive Computer Graphics – A top-down approach with OpenGL*, Addison-Wesley
- Arnheim R. (1957), *Film as Art*, The University of California Press
- Astlener H. (2000), Designing emotionally sound instruction: the FEASP approach, *Instructional Science* 28, pp. 169-198
- Auster P. (2004), *City of Glass*, Faber and Faber, adaptation by Karasik P. and Mazzuchelli D.
- Badler N., Bindigavale R., Bourne J., Palmer M., Shi J., Schuler W. (1998), A Parameterized Action Representation for Virtual Human Agents, Workshop on Embodied Conversational Characters, WECC98, Lake Tahoe, CA, Oct 12-15, 1998.
- Baeza-Yates R. and Ribeiro-Neto B. (1999), *Modern Information Retrieval*, Addison-Wesley, 1999
- Ballegooij A. van and Eliens A. (2001), Navigation by Query in Virtual Worlds, Web3D 2001 Conference, Paderborn, Germany, 19-22 Feb 2001
- Barcelo S.A., Forte M., Sanders D.H., eds. (2000), *Virtual Reality in Archeology*, Bar International Series 843, 2000
- Barron T. (2003), *Strategy Game Programming with DirectX 9.0*, Wordware Publishing
- Baxter W., Wendt J., and Lin M. (2004), IMPaSTo: A Realistic, Interactive Model for Paint, Proc. of NPAR 2004, The 3rd Int. Symp. on Non-Photorealistic Animation and Rendering. June 7-9 2004, Annecy, France. pp. 45–56.
- Berners-Lee T., Hendler J., Lassila O. (2001), The semantic web, *Scientific American*, may 2001, pp. 28-37
- Betsky A. with Eeuwens A. (2004), *False Flat – Why Dutch Design is so Good*, Phaidon
- Bolter J.D and Grusin R. (2000), *Remediation – Understanding New Media*, MIT Press
- Boncz P.A. and Kersten M.L. (1995), Monet – an impressionist sketch of an advanced database system, In Proc. BIWIT95

- Bongers B. (2006), *Interactivation – Towards an e-ecology of people, our environment, and the arts*, Ph.D. Thesis, Vrije Universiteit
- Bordwell D. and Thomson K. (2003), *Film art, an introduction*, McGraw-Hill, 7th edn.
- Briggs A. and Burke P. (2001), *A social history of the media – from Gutenberg to the Internet*, Polity Press
- Broll W. (1996), VRML and the Web: A basis for Multi-user Virtual Environments on the Internet, In Proceedings of WebNet96, H. Maurer (ed.), AACE, Charlottesville, VA (1996), 51-56.
- Broll W., Schäfer L., Höllerer T., Bowman D. (2001), Interface with Angels: the future of VR and AR interfaces, IEEE Computer Graphics, November/December 2001, pp. 14-17
- Bruner J.S. (1972), *Relevance of Education*, Penguin Education
- Bush V. (1945), As we may think, Atlantic Monthly, July 1945
- CACM 44:3 (2001), The next 1000 years, Communications of the ACM, March 2001, 44:3
- Capin T.K., Pandzic I.S., Magnenat-Thalmann N., Thalmann D. (1999), *Avatars in Networked Virtual Environments*, Wiley
- Carey R. and Bell G. (1997), *The Annotated VRML 2.0 Reference Manual*, Addison-Wesley
- Cesa-Bianchi N. and Lucosi G. (2006), *Prediction, Learning, and Games*, Cambridge University Press
- Chang S.C. and Costabile M.F. (1997), Visual Interfaces to Multimedia Databases, In [Handbook]
- Chapman N. and Chapman J. (2004a), *Digital Multimedia*, Wiley, 2nd edn.
- Chapman N. and Chapman J. (2004b), *Digital Media Tools*, Wiley, 2nd edn.
- Cheng F. (2006), *Five meditations on beauty (in french)*, Editions Albin Michel
- Christel, M., Olligschlaeger, A., Huang, C. (2000), Interactive maps for digital video, IEEE Multimedia 7(1), pp. 60-67
- Churchill E.F., Snowdon D.N. and Munro A.J., eds. (2001), *Collaborative Virtual Environments – Digital Places and Spaces for Interaction*, Springer
- Conger D. (2004), *Physics Modeling*, Thomson
- Conklin J. (1987), Hypertext: An Introduction and Survey, *IEEE Computer* 20(9), pp. 17-41
- Danaher S. (2004), *Digital 3D Design*, Thomson
- Davenport G.(2000), Your own virtual story world, Scientific American, november 2000, pp. 61-64
- Davison A. (1993), A survey of Logic Programming-based Object-Oriented Languages, In Research Directions on Concurrent Object-Oriented programming, G. Agha, P. Wegner, A. Yonezawa (eds.), MIT Press (1993)
- Davison A. (2001), Enhancing VRML97 Scripting, Euromedia2001, Valencia, Spain, April 18-20. available from: <http://fivedots.coe.psu.ac.th/~ad>

- Davison A. (2001b), Logic Programming Languages for the Internet, Chapter in Computational Logic: From Logic Programming into the Future, Antonis Kakas, Fariba Sadri (eds.), Springer Verlag (2001). available from: <http://fivedots.coe.psu.ac.th/~ad>
- Diehl G. (1973), *Vasarely*, Crown Publishers Inc.
- Diehl S. (2001), *Distributed Virtual Worlds – Foundations and Implementation Techniques using VRML, Java and Corba*, Springer
- Dijkstra K., Zwaan R. Graesser A. and J. Magliano (1994), Character and reader emotions in literary texts, *Poetics* 23, pp. 139-157
- Dix A., Finlay J., Abowd G., Beale R. (1998), *Human-Computer Interaction (1998)*, Prentice Hall, 2nd edn.
- Dodge M. and Kitchin R. (2000), *Mapping Cyberspace*, Routledge
- Dodge M. and Kitchin R. (2002), *Atlas of Cyberspace*, Addison-Wesley
- Earnshaw R. and Vince J., eds. (2002), *Intelligent Agents for Mobile and Virtual Media*, Springer
- Eberly D.H. (2001), *3D Game Engine Design*, Morgan Kaufmann Publishers
- Eberly D.H. (2004), *Game Physics*, Morgan Kaufmann Publishers
- Eco U. (1994), *Six walks in the fictional woods*, Harvard University Press
- Eliens A. (1988), *Computational Art*, Leonardo, MIT Press
- Eliens A. (1992), *DLP – A language for Distributed Logic Programming*, Wiley
- Eliens A. (1996), workshop: Logic Programming and the Web, <http://www.cs.vu.nl/~eliens/online/workshops/www6>
- Eliens A., Welie M. van, Ossenbruggen J. van and Schonhage S.P.C (1997) , Jamming (on) the Web, In: Proceedings of the 6th International World Wide Web Conference — Everone, Everything Connected, O'Reilly and Associates, Inc. April 1997, pp. 419-426
- Eliens A., Niessink F., Schonhage S.P.C., van Ossenbruggen J.R., Nash P. (1996), Support for BPR – simulation, hypermedia and the Web, Proceedings Euromedia'96, Euromedia, London 1996
- Eliens A. (2000), *Principles of Object-Oriented Software Development*, Addison-Wesley Longman, 2nd edn.
- Eliens A. (2002), intelligent multimedia @ VU, <http://www.cs.vu.nl/~eliens/research>
- Eliens A., Huang Z., and Visser C. (2002), A platform for Embodied Conversational Agents based on Distributed Logic Programming, AAMAS Workshop – Embodied conversational agents - let's specify and evaluate them!, Bologna 17/7/2002
- Eliens A., Dormann C., Huang Z. and Visser C. (2003), A framework for mixed media – emotive dialogs, rich media and virtual environments, Proc. TIDSE03, 1st Int. Conf. on Technologies for Interactive Digital Storytelling and Entertainment, Göbel S. Braun N.,n Spierling U., Dechau J. and Diener H. (eds>), Fraunhofer IRB Verlag, Darmstadt Germany, March 24-26, 2003
- Eliens A., Huang Z., Visser C. (2005), Presentational VR – *What is the secret of the slides?*, in preparation

- Eliens A., van Riel C., Wang Y. (2006), Navigating media-rich information spaces using concept graphs – the abramovic dossier, accepted for: International Conference on Multidisciplinary Information Sciences and Technologies (InSciT2006) October, 25-28th 2006, Merida, Spain www.instac.es/inscit2006
- Eliens A. (2006b), Odyssee – explorations in mixed reality theatre, accepted for: GAME'ON-NA'2006, September 19-21, 2006 - Naval Postgraduate School, Monterey, USA
- Eliens A., S.V. Bhikharie (2006), game @ VU – developing a masterclass for high-school students using the Half-life 2 SDK, GAME'ON-NA'2006, September 19-21, 2006 - Naval Postgraduate School, Monterey, USA
- Engel W.F. (2003), *Beginning Direct3D Programming*, Premier Press
- Engel W.F., ed. (2004a), *Shader X2, Introduction & Tutorials with DirectX 9*, Wordware Publishing
- Engel W.F., ed. (2004b), *Shader X2 – Shader Programming, Tips & Tricks with DirectX 9*, Wordware Publishing
- Engel W.F., ed. (2005), *Shader X3, Advanced Rendering with DirectX and OpenGL*, Charles River Media
- Engelbart D. (1963), A Conceptual Framework for the Augmentation of Man's Intellect, *Vistas in Information Handling* 1(9)
- England E. and Finney A. (1999), *Managing Multimedia – project Management for Interactive Media*, Addison-Wesley
- Entabaciaz A. (2003), *Les metamorphoses d'Ulysse – reecritures de l'Odyssée*, Editions Flammarion, Paris
- Ericson C. (2005), *Real-time Collision Detection*, Morgan Kaufmann Publishers
- Fan J., Ries E. and Tenitchi C. (1996), *Black Art of JAVA Game Programming*, Waite Group Press
- Faulkner X. (2000), *Usability Engineering*, MacMillan Press
- Fay T.M. with Selfon S. and Fay T.J. (2004), *DirectX Audio Exposed – Interactive Audio Development*, Wordware Publishing Inc.
- Fernando R. and Kilgard M.J. (2003), *The Cg Tutorial – The Definitive Guide to Programming Real-Time Graphics*, Addison-Wesley
- Fischetti M. (2000), The future of digital entertainment, *Scientific American*, november 2000, pp. 31-33
- Fluckiger R. (1995), *Understanding networked multimedia – applications and technology*, Prentice Hall, 1995
- Fogg B.J. (2003), *Persuasive Technology – Using Computers to Change What We Think and Do*, Morgan Kaufmann Publishers
- Foley J.D. and van Dam A. (1982), *Fundamentals of Interactive Computer Graphics*, Addison-Wesley
- Forman P. and Saint John R.W. (2000), Digital Convergence, *Scientific American*, november 2000, pp. 34-40
- Fuhr, N., Gövert, N., Rölleke, Th. (1998), DOLORES: A System for Logic-Based Retrieval of Multimedia Objects, In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 257-265. ACM, New York

- Gaver W. (2002), Designing for Homo Ludens, i3 Magazine
- Gee J.P. (2003), *What video games have to teach us about learning and literacy*, Palgrave Macmillan
- Gibson W. (1986), *Neuromancer*,
- Gonzales R.C. and Wintz P. (1987), *Digital Image Processing*, Addison-Wesley, 2nd edn.
- Grau O. (2003), *Virtual Art – From Illusion to Immersion*, The MIT Press
- Grosky W., Jain R., Mehrotra R., eds. (1997), *The Handbook of Multimedia Information Management*, Prentice Hall, 1997
- Hardman L., Bulterman D., van Rossum G. (1994), The Amsterdam Hypermedia Model: Adding Time and Context to the Dexter Model, CACM 37(2), pp. 50-62, febr 1994
- Harel D. (1984), Dynamic Logic, In: Handbook of Philosophical Logic, Vol. II, D. Reidel Publishing Company, 1984, pp. 497-604
- Hartman J. and Wernecke J. (1996), *The VRML 2.0 Handbook – Building Moving Worlds on the Web*, Addison-Wesley
- Hawkins B. (2005), *Real-time cinematography for games*, Charles River Media
- Hewlett W.B. and Selfridge-Field E., eds. (1998) , *Melodic Similarity – Concepts, Procedures and Applications*, MIT Press
- Hildebrand M., Eliens A., Huang Z. and Visser C. (2003), Interactive Agents Learning their Environment, Proc. Intelligent Virtual Agents 2003, Irsee, September 15-17, 2003 J.G. Carbonell and J.Siekman (eds.), LNAI 2792, Springer, pp. 13-17
- Hix D. and Hartson R. (1993), *Developing User Interfaces*, Wiley
- Hoggar S.G. (1992), *Mathematics for Computer Graphics*, Cambridge University Press
- Hohl H. (1971), *Giacometti*, Abrams Publishers
- Hoorn J., Eliens A., Huang Z., van Vugt H.C., Konijn E.A., Visser C.T., Agents with character: Evaluation of empathic agents in digital dossiers, Emphatic Agents, AAMAS 2004 New York 19 July - 23 July, 2004
- Huang Z., Eliens A., van Ballegooij A., De Bra P. (2000), A Taxonomy of Web Agents, IEEE Proceedings of the First International Workshop on Web Agent Systems and Applications (WASA 2000), 2000.
- Huang Z., Eliens A., and De Bra P. (2001), An Architecture for Web Agents, Proceedings of the Conference EUROMEDIA 2001, 2001.
- Huang Z., Eliens A., Visser C. (2001), Programmability of Intelligent Agent Avatars, Proceedings of the Agent'01 Workshop on Embodied Agents, June 2001, Montreal, Canada
- Huang Z., Eliens A., Visser C. (2002), 3D Agent-based Virtual Communities, In: Proc. Int. Web3D Symposium, Wagner W. and Beitler M.(eds), ACM Press, pp. 137-144
- Huang Z., Eliens A., Visser C. (2005), *Intelligent Multimedia Technology: An Approach to Combine Agent Technologies with Multimedia*, in preparation
- Huang, Z., Eliens, A., and Visser, C., *STEP: a Scripting Language for Embodied Agents*, in: Helmut Prendinger and Mitsuru Ishizuka (eds.), Life-like Characters, Tools, Affective Functions and Applications, Springer-Verlag, 2003.

- Huang, Z., Eliens, A., and Visser, C. (2003d), "Is it within my reach?" – an agents perspective , Proc. Intelligent Virtual Agents 2003, Irsee, September 15-17, 2003, J.G. Carbonell and J.Siekmann (eds.), LNAI 2792, Springer, pp. 150-158
- Huang Z., Eliens A., and Visser C. (2004), Facial Animation in STEP, AAMAS 2004 New York 19 July - 23 July, 2004
- Huang, Z., Eliens, A., and Visser, C. (2003c), *STEP: a Scripting Language for Embodied Agents*, in: Helmut Prendinger and Mitsuru Ishizuka (eds.), Life-like Characters, Tools, Affective Functions and Applications, Springer-Verlag, (to appear).
- Huang Z., Eliens A., Visser C. (2003b), XSTEP: A Markup Language for Embodied Agents, Proc. CASA03, The 16th Int. Conf. on Computer Animation and Social Agents
- Huang Z., Eliens A., Visser C. (2003a), Implementation of a scripting language for VRML/X3D-based embodied agents, Proc. Web3D 2003 Symposium, Saint Malo France, S. Spencer (ed.) ACM Press, pp. 91-100
- Huang Z., Eliens A., Visser C. (2003a), Implementation of a scripting language for VRML/X3D-based embodied agents, Proc. Web3D 2003 Symposium, Saint Malo France, S. Spencer (ed.) ACM Press, pp. 91-100
- Huang Z., Eliens A., Visser C. (2002b), STEP – a scripting language for Embodied Agents, PRICAI-02 Workshop – Lifelike Animated Agents: Tools, Affective Functions, and Applications, Tokyo, 19/8/2002
- Hughes B. (2000), *Dust or Magic – Secrets of Successful Multimedia Design*, Addison-Wesley, 2000
- Hummelen IJ. and Sillé D. (1999), *Modern Art: Who cares?*, Foundation for the Conservation of Modern Art and the Netherlands Institute for Cultural Heritage
- Huron D. (1997) , Humdrum and Kern: Selective Feature Encoding, In [*Beyond*], pp. 375-400
- Jain R. (2000), Digital Experience, [*1000*], pp. 38-40
- JavaSoft (2001), Java3D API, <http://java.sun.com/products/java-media/3D>
- Johnson A., Moher T., Cho Y.-J., Lin Y.-J., Haas D., and Kim J. (2002), Augmenting Elementary School Education with VR, IEEE Computer Graphics and Applications, March/April
- Johnson B., Skibo C., Young M. (2003), *Inside Microsoft Visual Studio .NET*, Microsoft Press
- Juul J. (2005), *Half Real – Video Games between Real Rules and Fictional Worlds*, MIT Press
- Kalawsky R.S. (1993), *The Science of Virtual Reality and Virtual Reality Environments*, Addison-Wesley
- Kay M. (2001), *XSLT Programmer's Reference*, Wrox Press
- Kersten M. L., Nes M., Windhouwer M.A. (1998), A feature database for multimedia objects, CWI Report INS-R9807, July 1998
- Kientzle T. (1998), *A programmer Guide to Sound*, Addison-Wesley
- Kimmel R. (2004), *Numerical Geometry of Images – Theory, Algorithms, and Applications*, Springer,

- Koegel Buford J.F., ed. (1994), *Multimedia Systems*, Addison-Wesley, 1994
- Koenen R. (1999), MPEG-4 - Multimedia for our time, IEEE Spectrum, Vol. 36, No. 2, February 1999, pp. 26-33.
- Koenen R. (ed.) (2000), Coding of moving pictures and audio, ISO/ITEC JTC1/SC29/WG11 N3747
- Kress G. and van Leeuwen T. (1996), *Reading Images: The Grammar of Visual Design*, Routledge
- Krikke J. (2003), Samurai Romanesque, and the battle for mobile cyberspace, IEEE Computer, Jan/Febr 2003
- Kuipers J.B. (1999), *Quaternions and Rotation Sequences – A Primer with Applications to Orbits, Aerospace, and Virtual Reality*, Princeton University Press
- Labrou L., Finin T. and Peng Y. (1999), The current landscape of Agent Communication Languages, IEEE Intelligent Systems, Vol. 14, no. 2, 1999
- LaMothe A. (2002), *Tricks of the Windows Game Programming Gurus*, Sams, 2nd edn.,
- Lengyel E. (2004), *Mathematics for 3D Game Programming & Computer Graphics*, Charles River Media
- Li Z-N. and Drew M.S. (2004), *Fundamentals of Multimedia*, Prentice Hall, Pearson
- Luna F. (2003), *Introduction to 3D Game Programming with DirectX 9.0*, Wordware Publishing Inc.
- Manco T. (2004), *Street logos*, Thames & Hudson
- McCool M. and Du Toit S. (2004), *Metaprogramming GPUs with Sh*, A K Peters, Ltd.
- McCuskey M. (2002), *Special Effects Game Programming with DirectX*, Premier Press
- McKnight C., Dillon A. and Richardson J. (1991), *Hypertext in Context*, Cambridge University Press
- McNab R.J., Smith L.A., Bainbridge D. and Witten I.M. (1997), The New Zealand Digital Library MELody inDEX, D-Lib Magazine, May 1997
- McQuail D. and Siune K. (1989), *Media Policy – Convergence, Concentration & Commerce*, Euromedia Research Group, Sage Publications
- Messick P. (1998), *Maximum MIDI – Music Applications in C++*, Manning
- Moortgat, M. (1996), Categorical Type Logics, Handbook of Logic and Language, J. van Benthem and A. ter Meulen (eds.), Elsevier Amsterdam (1996), pp. 93-177
- Moortgat, M. (2002), Categorical Grammar and Formal Semantics, Encyclopedia of Cognitive Science, Nature Publishing Group, Macmillan Publishers Ltd.
- Mongeau M. and Sankoff D. (1990), Comparison of musical sequences, Computers and the Humanities 24, pp. 161-175, 1990
- Morrison M. (2005), *Beginning Mobile Phone Game Programming*, Sams Publishing
- Mozart W.A. (1781), Twinkle, twinkle little star (Oh! vous dirai-je, mamam), K300, 1781-82
- Munro A., Höök, K. and Benyon D.R. (1999), Footprints in the Snow, In: Social Navigation of Information Space, Springer

- Murray J. (1997), *Hamlet on the Holodeck – The future of narrative in Cyberspace*, Free Press
- Negroponte N. (1995), *Being Digital*, New Riders
- Nelson T. (1980), *Literary Machines*, Mindfull Press
- Nielsen J. and Mack R.L., eds. (1994), *Usability inspection methods*, Wiley
- Norman D.A. (2004), *Emotional Design*, Basic Books
- Pachet F., Roy P. and Cazaly D. (1999), A combinatorial approach to content-based music selection, Proc. IEEE Int. Conf. on Multimedia Systems, Firenze
- Ossenbruggen J. van and Eliens A. (1994), Music in Time-based Hypermedia, Proc. European Conference on Hypermedia Technology 1994, pp. 224-227
- Ossenbruggen J. van (2001), *Processing Structured Hypermedia – A matter of style*, Ph.D. Thesis, Free University, 2001
- Ossenbruggen J. van, Geurts J., Cornelissen F., Rutledge L., Hardman L. (2001), Towards Second and Third Generation Web-Based Multimedia, In Proc. of The Tenth International World Wide Web Conference pp. 479-488, May 1-5, 2001, Hong Kong
- Patin S. (2006), *Regards sur les Nympheas (in french)*, Editions de la reunion des musees nationaux.
- Pesce M. (2003), *Programming Microsoft DirectShow for digital video and television*, Microsoft Press
- Picard R.W. (1998), *Affective Computing*, MIT Press
- Preece J., Rogers Y., Sharp H., Benyon D., Holland S., Carey T. (1994), *Human-Computer Interaction*, Addison-Wesley
- Purvis A.W. and Le Coutre M.F. (2003), *Graphic Design 20th Century*, BIS Publishers
- Ratner P. (2003), *3-D Human Modeling and Animation*, Wiley, 2nd edn.
- Riel C. van, Eliens A., Wang Y (2006), Exploration and guidance in media-rich information spaces: the implementation and realization of guided tours in digital dossiers, accepted for: International Conference on Multidisciplinary Information Sciences and Technologies (InSciT2006) October, 25-28th 2006, Merida, Spain www.instac.es/inscit2006
- Riel C. van, Wang Y. and Eliens A. (2006b), Concept map as visual interface in 3D Digital Dossiers: implementation and realization of the Music Dossier, CMC2006, Costa Rica, Sept 5-8 2006
- Rooijen P. van, ed. (2003), *Signs & Symbols*, Pepin Press
- Rosenblum L. and Macedonia M., eds. (2002), Projects in VR – From Urban Terrain Models to Visible Cities, IEEE Computer Graphics and Applications, 22:4, pp. 10-15
- Rosenblum L. and Macedonia M., eds. (2005), Projects in VR – Tangible Photorealistic Virtual Museum, IEEE Computer Graphics and Applications, 25:1, pp. 15-17
- Rutledge L., van Ballegooij A., Eliens A. (2000), Virtual Context - relating paintings to their subject, Culture Track of WWW9 in Amsterdam, The Netherlands, May 16th, 2000

- Ruttkay M., Huang Z. and Eliens A. (2003), Reusable gestures for interactive web agents, Proc. Intelligent Virtual Agents 2003, Irsee, September 15-17, 2003 J.G. Carbonell and J.Siekmann (eds.), LNAI 2792, Springer, pp. 80-87
- Ruttkay Z., Huang Z. and Eliens A., *The Conductor: Gestures for Embodied Agents with Logic Programming*, in: Recent Advances in Constraints K.R. Apt, F. Fages, F. Rossi, P. Szeredi and J. Vancza (eds.) LNAI 3010, Springer 2004
- Sanchez-Crespo Dalmau D. (2004), *Core Techniques and Algorithms in Game Programming*, New Riders
- Santiago D. (2005), *Creating 3D Effects for Film, TV and Games*, Thomson
- Santos Lobao A. and Hatton E. (2003), *.NET Game Programming with DirectX 9.0*, The Authors Press
- Schmidt, A.R. Windhouwer M.A., Kersten M.L. (1999), Indexing real-world data using semi-structured documents, CWI Report INS-R9902
- Schneider P.J. and Eberly D.H. (2003), *Geometric Tools for Computer Graphics*, Morgan Kaufmann Publishers
- Schönhage S.P.C., van Ballegooij A. and Eliens A. (2000), 3D Gadgets for Business Process Visualization: a case study, Proc. Int. Web3D/VRML Conference - 2000, Monterey CA, Febr 2000, ACM Press pp. 131-138
- Schuytema P. (2007), *Game Design: A Practical Approach*, Charles River Media
- Selfridge-Field E., ed. (1997), *Beyond MIDI – The Handbook of Musical Codes*, MIT Press 1997
- Selfridge-Field E. (1998), Conceptual and Representational Issues in Melodic Comparison, In [Similarity], pp. 3-64
- Sherman W.R. and Craig A.B. (2003), *Understanding Virtual Reality – Interface, Application and Design*, Morgan Kaufmann Publishers
- Sherrod A. (2006), *Ultimate Game Programming with DirectX*, Charles River Media
- Shneidermann B. (1997), *Designing the user interface – strategies for effective human-computer interaction*, Addison-Wesley (3rd edn)
- Shneiderman B (2003), *Leonardo's Laptop*, Mit Press
- Singhal S. and Zyda M. (1999), *Networked Virtual Environments*, Addison-Wesley, 1999
- Snook G. (2003), *Real-time 3D terrain engines using C++ and DirectX 9*, Charles Riven Media
- Spalter A.M. (1999), *The Computer in the Visual Arts*, Addison-Wesley
- Stephenson N. (1992), *Snowcrash*,
- St-Laurent S. (2004), *Shaders for Game Programmers and Artists*, Thomson
- Subrahmanian V.S. (1998), *Principles of Multimedia Databases*, Morgan Kaufmann
- Sullivan A. (2005), 3-Deep, IEEE Spectrum 42-4, April 2005, pp. 22-27
- Tarau P. (1999), Jinni: Intelligent Mobile Agent Programming at the Intersection of Java and Prolog, Proc. of PAAM99, London, UK, April, see also <http://www.binnetcorp.com/Jinni>
- Temperley D. and Sleator D. (1999), Modeling Meter and Harmony: A Preference-Rule Approach, Computer Music Journal 23(1), 1999, pp. 10-27

- Trogemann G. & Viehoff J. (2004), *CodeArt – Eine elementare Einführung in die Programmierung als künstlerische Praktik*, Springer
- Vasudev B. and Li W. (1997), Memory management: Codecs, In [_Handbook], pp. 237-278
- Vaughan T. (1998), *Multimedia – Making It Work*, Osborne/McGraw-Hill, 4th edn
- Vince J. (1998), *Essential Virtual Reality*, Springer
- Visser C. and Eliens A. (2000), A High-Level Symbolic Language for Distributed Web Programming, Internet Computing 2000, June 26-29, Las Vegas
- Vyas, D., van der Veer, G. C. (2006), Experience as Meaning: Some Underlying Concepts and Implications for Design. , Proceedings of 13th European Conference on Cognitive Ergonomics. ACM Press: NY, (2006), 81-91.
- Vyas, D., Chisalita, C., van der Veer, G. C. (2006), Affordance in Interaction. , Proc. of 13th European Conference on Cognitive Ergonomics. ACM Press: NY, (2006), 92-99.
- W3C SMIL Working Group (2001), Synchronized Multimedia Integration Language (SMIL 2.0), W3C Recommendation 07 August 2001
- Wagner W. and Beitler M. (eds), 2002, Proc. Int. Web3D Symposium, ACM Press
- Walsh P. (2003), *Advanced 3D Programming with DirectX 9.0*, Wordware Publishing Inc.
- Wang Y., Eliens A., van Riel C. (2006), Content-oriented presentation and personalized interface of cultural heritage in digital dossiers, accepted for: International Conference on Multidisciplinary Information Sciences and Technologies (InSciT2006) October, 25-28th 2006, Merida, Spain www.instac.es/inscit2006
- Watt A. and Policarpo F. (2001), *3D Games – Real-time Rendering and Software Technology*, Addison-Wesley
- Web3D Consortium (2001), VRML97 Standard, http://www.web3d.org/fs_specifications.htm
- Weishar P. (1998), *Digital Space – Designing Virtual Environments*, McGraw-Hill
- Welie M. and Eliens A. (1996), Chatting on the Web, Proc. ERCIM W4G Workshop on CSCW and the Web, 7-9 febr 1996 GMD St Augustin, Germany, 1996
- Welie M. avn, G.C. van der Veer G.C., and Eliens A. (1998) , Euterpe - Tool support for analyzing cooperative environments, Proceedings of the Ninth European Conference on Cognitive Ergonomics , August 24-26, 1998, University of Limerick, Ireland, pp. 25-30
- Wiedermann J., ed. (2002), *Digital Beauties*, Taschen Verlag
- Wiedermann J., ed. (2004), *Animation now*, Taschen Verlag
- Wilson S. (2002), *Information Arts – Intersections of Art, Science and Technology*, The MIT Press
- Yellowlees D. and Hargadon A. (2000), The pleasure principle: immersion, engagement, flow, Proc. of the 11th ACM Conf. on Hypertext and Hypermedia, May 2000, San Antonio, pp. 153-160
- Zielinski S. (2006), *Deep Time of the Media – Towards an archaeology of Hearing and Seeing by Technical Means*, The MIT Press

- Zimmermann D., Modeling Musical Structures, Aims, Limitations and the Artists Involvement, Proc. Constraints techniques for artistic applications, Workshop at ECAI98 25th August, 1998, Brighton, UK, <ftp://ftp.csl.sony.fr/pub/pachet/workshopEcai>

index

Adams (2003), ix, 100
Arnheim (1957), ix, 192, 261
Auster (2004), 139

Baeza-Yates and Ribeiro-Neto (1999),
ix, 72, 73, 105, 106, 116, 117,
143, 146, 147
Ballegooij and Eliëns (2001), ix
Baxter et al. (2004), 86
Betsky (2004), 44, 155
Bolter and Grusin (2000), ix, 38–40, 44,
177, 265
Briggs and Burke (2001), 9, 15, 20, 21,
44
Broll (1996), 289
Broll et. al (2001), 73
Bruner (1972), 4, 258, 270
Bush (1995), 31

CACM (2001), 22
Cesa-Bianchi and Lucosi (2006), 263
Chang and Costabile (1997), ix, 26, 31
Chapman and Chapman (2004a), ix
Chapman and Chapman (2004b), ix

Davenport (2000), ix, 5
Davison (2001), 289
Diehl 1973, 44
Dodge and Kitchin (2002), 131

Eco (1994), 44
Eliëns (1988), 199, 268

Eliëns (1992), 283
Eliëns (2000), 35, 46, 73, 89, 135, 153,
154, 163, 216, 284
Eliëns et al. (2002), 283
Eliëns *et al.* (1997), 35
Engel (2004a), 86
Engel (2004b), 86
Engel (2005), 86
Engelbart (1963), 31
Entanaclaz (2003), 195

Faulkner (2000), 190
Fay et al. (2004), 100
Fernando and Kilgard (2003), ix, 80,
86
Fluckiger (1995), ix, 148, 149
Forman and Saint John (2000), ix
Fuhr et al. (1998), 107

Gonzales and Wintz (1987), 117
Grau (2003), ix, 4, 246, 248, 262

Hardman et al. (1994), 35
Harel (1984), 174
Hawkins (2005), ix
Hewlett and Selfridge-field (1998), 121,
184
Hildebrand et al. (2003), 263
Hohl (1971), 158
Huang et al. (2000), 172
Huang et al. (2002), ix, 171
Hughes (2000), ix, 32, 184–186
Hummelen and Sillé (1999), 136, 139
Huron (1997), 124

- Jain (2000), ix, 6
 Johnson et al. (2002), 175
 Juul (2005), ix
- Kay (2001), 30, 282
 Kersten et al. (1998), ix, 132
 Koenen (1999), 54
 Koenen (2000), ix, 53, 56, 76
 Kress and van Leeuwen (1996), ix, 232, 245–247, 260
 Krikke (2003), 235
- Li and Drew (2004), 76, 155
 Luna (2003), ix, 100, 249
- Manco (2004), 201
 McCuskey (2002), ix, 226
 McKnight et al. (1991), 33
 McNab et al. (1997), ix, 123, 124, 137
 Mongeau and Sankoff (1990), 121, 123
 Morrison (2005), 22
 Mozart (1787), 121, 137
 Murray (1997), 41, 42
- Negroponte (1995), 3
 Nelson (1980), 31
 Norman (2004), 187, 188, 191
- Ossenbruggen (2001), ix, 35, 36
 Ossenbruggen & Eliëns (1994), 35
 Ossenbruggen et al. (2001), 69
- Pesce (2003), 89, 130
 Picard (1998), 75
 Preece et al. (1994), 209, 215
- Rosenblum and Macedonia (2002), 14, 23
 Rosenblum and Macedonia (2005), 208, 226
- Sanchez-Crespo Dalmau (2004), 234
 Santos Lobao and Hatton (2003), 22
 Schmidt et al. (1999), 133
 Schuytema (2007), 232
 Selfridge (1997), 138
 Selfridge (1998), 121, 122
 Sherman and Craig (2003), 176, 186
 Sherrod (2006), ix, 233
 Shneiderman (1997), 37
 Shneiderman (2003), 22
 Singhal and Zyda (1999), 150, 151
 St-Laurent (2004), 86
 Subrahmanian (1998), ix, 103, 104, 107, 109–112, 114, 115, 120, 125–128, 142, 144, 145, 149
 Sullivan (2005), 92, 100
- Tarau (1999), 289
 Temperley and Sleator (1999), 124
 Trogemann & Viehoff (2004), 98
- van Rooijen (2003), 2, 23, 44, 46, 76, 100, 102, 118, 139, 156, 158, 176, 201, 227, 230, 250, 265
 Vasudev and Li (1997), ix, 48–50
 Visser and Eliëns (2000), ix
 Vyas and van der Veer (2006), 254
- W3C (2001), 53
 Weishar (1998), 139
 Welie et al. (1998), 216, 227
 Wiedermann (2002), 176, 249
 Wiedermann (2004), 76, 201
 Wilson (2002), 4, 248

Zielinski (2006), ix, 4, 5

Zimmerman (1998), 134

topical media & game development

This book provides a concise and comprehensive introduction to multimedia. It arose out of the need for material with a strong academic component, that is (simply) material related to scientific research. Indeed, studying multimedia is not (only) fun. Compare it with obtaining a driver license. Before you are allowed to drive on the highway, you have to take a theory exam. So why not take such an exam before entering the multimedia circus. Don't complain, and take the exam. After all it makes you aware of the rules governing the (broadband) digital highway. The book and accompanying material is available at <http://www.cs.vu.nl/eliens/media>