The Electronic Scrapbook:
Towards an Intelligent Home-Video Editing System

by

Amy Susan Bruckman

A.B. Physics
Harvard University
Cambridge, MA
1983

SUBMITTED TO THE MEDIA ARTS AND SCIENCES SECTION, SCHOOL OF ARCHITECTURE AND PLANNING, IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN VISUAL STUDIES
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
September 1991

©Massachusetts Institute of Technology 1991
All Rights Reserved

Signature of the Author
.....
Amy Susan Bruckman
Media Arts and Sciences Section
August 9th, 1991

Certified by
.....
Glorianna Davenport
Assistant Professor of Media Technology
Thesis Supervisor

Accepted by
.....
Stephen A. Benton
Chairman
Departmental Committee on Graduate Students

The Electronic Scrapbook:

Towards an Intelligent Home-Video Editing System

by

Amy Susan Bruckman

Submitted to the Media Arts and Sciences Section, School of Architecture and Planning,
on August 9th, 1991, in partial fulfillment of the requirements
for the degree of Master of Science in Visual Studies
at the Massachusetts Institute of Technology

Abstract

How many people's home videos remain unedited and unwatched? Home video is a growing cultural phenomenon; however, few consumers have the time, equipment, and skills needed to edit their work. The Electronic Scrapbook is an environment designed to encourage people to use home video as a creative medium. The system and the user collaborate to create home-video stories.

This work addresses issues of knowledge representation and interface design. Semantic knowledge representation is evaluated as a way to represent information about complex, temporal media. A modified form of case-based reasoning, "knowledge-based templates," is used to explore what a computational model of a home-video story might be.

Thesis Supervisor: Glorianna Davenport
Title: Assistant Professor of Media Technology

This work was supported in part by Apple Computer, Inc., and by The Nichidai Fund.

Acknowledgements

I'd like to thank my advisor, Professor Glorianna Davenport, for her guidance and support.

My thesis readers, Professor Kenneth Haase and David S. Backer, Phd. provided thoughtful comments and technical expertise.

Marc Davis, Pattie Maes, Alan Ruttenberg, and Michael Travers commented on drafts of this thesis. It was Marc who first suggested using a "scrapbook" design metaphor. Marc helped to critique the system design as it evolved. Mike and Alan have been a constant source of technical support, answering many a pesky question when I was first learning lisp.

I'd like to thank all of my colleagues at the Media Lab for providing such a wonderful environment to work and play, often the same thing, especially: Edith Ackermann, Carlos Alston, Joe Chung, Stuart Cody, Eddie Elliot, Henry Lieberman, Ron MacNeil, Dave Rosenthal, and Liza Wirtz.

The narrative-intelligence reading group has provided an intellectual context for research at the Media Lab. In addition to those members already named above, I'd also like to thank Professor Henry Jenkins.

I'd like to thank Mom and Bernie for being there for me.

Finally, thanks go to all those who lent me their home videos -- Linda Haviland Conte, Marie Crowley, Chris Gant, friends of Shawn O'Donnell, Rosalind Picard, Chris Schmandt, and especially Dad, Mari, Alicia, Danielle, Rachael, Enid, Rose, Toby, and Lucy.

Contents

1. Introduction 9
    1.1 An Electronic Scrapbook 9
    1.2 A Sample Interaction 10
    1.3 Overview of this Document 11

2. Background: Representation and Interface Design 13
    2.1 Editing Images 13
    2.2 The Search Problem 13
    2.3 Computer Representations 14
    2.4 How Much Representation is Necessary? 16
    2.5 Interface Design and Coding 17
    2.6 Constructionism 17
    2.7 Visual Interface Design 18

3. A Digital Video Design 20
    3.1 Current Hardware 20
        3.1.1 A Digital Video Scrapbook 20
        3.1.2 Laserdisc as a Stand-in for Digital Video 21
    3.2 Software 21

4. Video Source Material 22
    4.1 Choice of Material 22
    4.2 My Family's Video 23

5. Interface Design and Development 28
    5.1 Titled Icons 28
        5.1.1 Combining Text and Graphics 28
        5.1.2 Local Methods 29
    5.2 Video Controller 29
        5.2.1 Device Control 29
        5.2.2 Paying Full Attention to the Picture 30
    5.3 The Segment-Editor Window 32
        5.3.1 Text Versus Graphics 32
        5.3.2 Panes: Avoiding Screen Clutter 32
    5.4 List Browsers 34
        5.4.1 A Sample List Browser 34
        5.4.2 Lists of Options: Towards Logging Consistency 34

5.4.3 Hierarchical Sorting: Visualizing the Underlying Representation 34
5.5 Video Segments 36
5.6 Scrapbook Windows 37
5.6.1 Spatial Implies Temporal Organization 37

6. A Suburban Ontology 39
6.1 People 39
6.1.1 Extending the Representation: New Buttons and Object Editors 39
6.1.2 Inferencing Capability: Family Relationships 40
6.1.3 The Need for a Graphical Interface 41
6.1.4 But What Can All of This Information Be Used For? 42
6.1.5 Focus Person 42
6.2 People in Audio Only 43
6.2.1 The Need for Separate Representations of Video and Audio 43
6.3 Locations 43
6.3.1 Functional rather than Geographic Information 43
6.3.2 Linking Space and Time 44
6.3.3 Interface Coding for a Multi-level Representation 45
6.3.4 Multiple Inheritance 45
6.4 Events 45
6.4.1 The Power of the "Event" Abstraction 47
6.5 Actions 47
6.5.1 Emotional Actions 48
6.5.2 Complex Information 48
6.6 Significant Objects 49
6.6.1 How Much Representation is Necessary? 50
6.7 Segment Date 50
6.7.1 Qualitative Reasoning 50
6.8 Camera Properties 51
6.8.1 Video Quality 52
6.8.2 Stylistic Constraints 52
6.9 Subjective Properties 53
6.9.1 Lack of Perspective on the Subject Matter 54
6.9.2 Sensitive Subject Matter 54

7. Guessing 55
7.1 Level of Detail 55
7.2 Inheritance 56
7.2.1 Event-based Inheritance 56
7.3 Making Inferences Apparent 56
7.4 Similar Segments 58
7.4.1 The Score Function 58
7.4.2 Selecting Similar Segments 60

8. Story Models 61
8.1 Database Searches 61
8.1.1 Checking the Knowledge Hierarchy 61
8.1.2 Searching for People 62
8.2 Simple Story Models 62
8.2.1 The Basic Search 62
8.2.2 Discards 63
8.2.3 Filtering 64
8.2.4 Removing Overlapping Segments 65
8.2.5 Sorting 66
8.2.6 One Last Pass: the After-sort Function 67
8.3 Compound Story Models 68
8.4 Results of a Simple Story Model: child-at-age 68
8.5 Comparison of Two Approaches: kids-at-same-age and important-firsts 70

9. Related Work 72
9.1 Case-Based Reasoning 72
9.2 Text-Based Story Generation 73
9.2.1 Manipulating a Limited Vocabulary 73
9.2.2 Goals 73
9.3 Scripts 74
9.4 The Cyc Project 74

10. Future Work 75
10.1 More Data 75
10.1.1 Birthday Parties 75
10.1.2 Scaling 75
10.1.3 Storage, Memory, and Speed Issues 75
10.2 Digital Video 76
10.2.1 Still Frames to Represent Segments 76
10.2.2 A Linear Representation of the Source Material 77
10.3 An Interface to Story-Models 77
10.4 A Graphical Interface to the Underlying Knowledge Representation 78
10.5 Making the Interface Reconfigurable 79
10.6 Annotations 79
10.6.1 Titles 79
10.6.2 Histories 80
10.6.3 Why 80
10.7 Scrapbook Page Types 80
10.8 Object Permanence 81
10.8.1 Instances 81
10.8.2 Temporal Extents 82

11. Conclusion 83
11.1 Interface Design 83
11.2 The Benefits of Knowledge Representation 84
11.2.1 Multiple Models 84
11.2.2 The Need for a Common Language 84

Bibliography 86

List of Figures

Figure 2.1: A semantic family tree....15
Figure 5.1: The descriptive icons window....28
Figure 5.2: The video controller....30
Figure 5.3: The segment-editor window....31
Figure 5.4: The segment-editor window, including an editor for the people pane....33
Figure 5.5: A hierarchically-sorted list-browser....35
Figure 5.6: The segments window....36
Figure 5.7: A scrapbook window....38
Figure 6.1: The “edit person” dialog box....39
Figure 7.1 : A segment-editor window and its corresponding “similar segments” window....57
Figure 8.1: The discards window....64
Figure 10.1: A video segment....76

List of Tables

Table 4A: General Contents of the Video Source Material....24
Table 4B: An Excerpt from the Log of the Rough Cut....27
Table 7A: Copying Dates....58
Table 7B: Default Weights for the Similar-Segments Function....60