M A R C   D A V I S

School of Information Management and Systems
University of California at Berkeley

marc@sims.berkeley.edu
www.sims.berkeley.edu/~marc

P U B L I C A T I O N S

Home Media Tools: Animation, Media Editing, and Production for Individuals

Bibliographic Reference:

Marc Davis, et al. “Home Media Tools: Animation, Media Editing, and Production for Individuals.” Research Overview for Interval Research Home Page. Interval Research Corporation, Palo Alto, California, 1997.

Home Media Tools

Animation, media editing, and production for individuals.

Since the dawn of digital technology, almost every measure of raw computing power has increased exponentially with time, and costs have likewise plummeted. Memory capacity, communication bandwidth, and processing power – all have steadily mushroomed over the decades, a phenomenon described by the so-called Moore's Law. This growth is now enabling amateurs at home to manipulate digital media with the computational power formerly available only to professionals.

But improvements in raw power alone do not make hands-on control of digital media appealing to consumers, since the tools have not become correspondingly easier to use. In fact, merely simplifying professional tools for the home may not be enough. We may need to change the basic models of media manipulation. And to be effective, we must develop these models with an eye to how people will use the media.

The key problem today is mandatory microscopic focus. Available media tools often force people to operate at the level of bits, pixels, or frames. Several projects at Interval are giving users control over media at a higher level of abstraction. Certain themes recur through these efforts.

Leveraging Professional Skill in Home Media Tools

Until home computers became powerful enough to manipulate digital media, a small, elite population of professionals created such popular media as movies, TV, cartoons, and music. The media consumer was rarely a media producer.

We are interested in new kinds of home media tools that let amateurs benefit from the skills of a professional in their own media creation, even if that professional is far away in time and space. We'd like to capture the expertise and style of artists and embody them in a tool that gives home users a power assist: a Pegasus users can ride to superior creative efforts of their own. We imagine that these kinds of tools may even enable a revolution in home media production bringing about new forms of media, a type of "garage cinema."

Freedom through Constraints, Structure, and Representation

One way to boost creativity is, paradoxically, through constraints and structure. Consider the popular Lego™ plastic construction blocks. Hand a child a mass of plastic and ask him to build something. He has complete freedom, but he needs to be a Michelangelo with a hammer and chisel to get anywhere. Mold that same plastic into modular shapes with just a few ways to connect, and endlessly impressive results ensue. The constraints imposed by the pegs and holes free children from the task of coping with the basic properties of the material, leaving them to operate at a higher level of structure.

Today, manipulating media is like trying to sculpt that lump of plastic, or like word processing with a paint program. The right high-level representations are not yet available for ordinary users. Also missing are the prerequisite mechanisms to derive structure and apply constraints. Adding these to media would enable users to manipulate them computationally and conceptually.

Content Analysis and Markup

Naturally, much of the work with higher-level representation of media depends for its success on certain types of signal processing and semantic markup of media. There is considerable research at Interval in the area of signal-based media that complements the efforts to provide meaningful higher-level structure for media content. One such effort in the semantic markup of media is Media Streams.

Devices with a Heart

Sometimes new representations of media require new devices. Recording devices available to consumers today are designed with relatively utilitarian needs in mind. They usually target the user who is willing to devote full effort to creating some finished product: such as an audiocassette for grandma or a videotape of the class play.

Interval researchers are interested in new ways of capturing, using, and storing images and recordings. We want Interval's technologies to be useful in recording experiences that matter to people. Research in this area stretches broadly across disciplines ranging from nuts-and-bolts engineering to in-depth study of people's lives.

Example Research Areas

At Interval, we have investigated home media tools in the following areas:

- Iconic visual tools for annotating and retrieving media content so that in making media at home, users aren't limited by what they can shoot, but instead have easy access to existing media content.
- Visual tools that make creating media software more like playing a game than doing work.
- Spacetime-constraints techniques that free the user from some of the time-consuming tasks involved in computer animation by combining physically correct dynamics with learning techniques.

Toward the Future

Just as we may find it hard to imagine our own culture before the advent of widespread literacy in the eighteenth century, in the next century our descendants will find it hard to understand that while everyone consumed media, few had the tools to make it. It may be hard to conceptualize a world in which you engage in a regular practice of making media at home to communicate and play with others, but your grandchildren will not understand how you ever lived without it.

Signal Computation

The compression, analysis, synthesis, and modification of audio, video, and image data for uses in entertainment and media production.

Signal computation is critical to many of Interval's projects. We are especially interested in:

- analyzing audio and video for information about human behaviors,
- modifying audio and video in interesting ways,
- represent and compress large data sets, and
- analyzing video for depth and motion information.

Our approaches tend to be data-intensive and based on statistics.

Some of our most recent work is described below.

Magic Morphin' Mirror: A SIGGRAPH '97 art piece, the Magic Morphin' Mirror is a technology demonstration combining work in face identification, color segmentation, stereo perception, and real-time texture manipulation. The mirror presents an image of space, much like a real mirror, but detects and modifies the region of a face in real time.

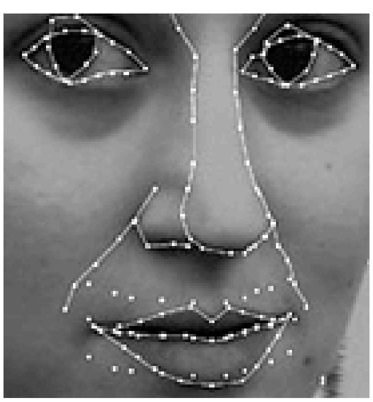
The Magic Morphin' Mirror distorts faces in an image.

Video Rewrite: Video Rewrite is a technology that synthesizes video of a person speaking words they never

spoke. A large database of a person speaking is analyzed automatically. Then given new audio, perhaps spoken by somebody imitating the desired person, we select li p sequences from the database and morph them into a background sequence. The result is a realistic video of a person speaking new words.

A smooth spectrogram, as shown here, is one step in the audio morphing process.

Audio Morphing: Morphing is often used to interpolate between two images. Thus an image of one person smoothly changes into another person's face, all the while preserving the illusion of one object. We've done the same thing with audio. Using the principles of auditor y scene analysis, we analyze a sound into the information which is principally due to the format information from the pitch and voicing information. Using these representations, we can smoothly move from one sound into another.

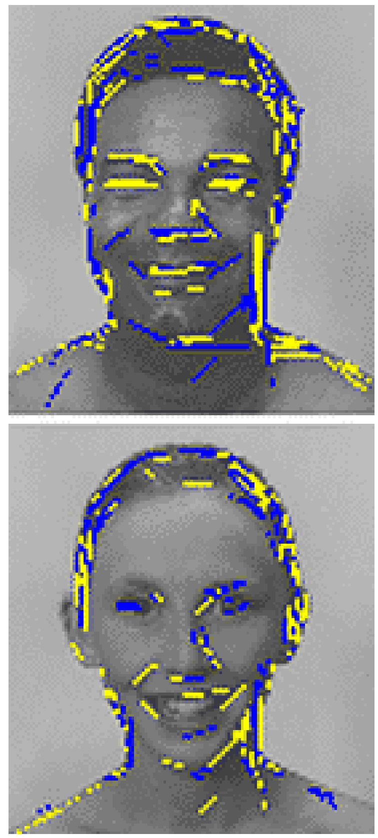


EigenPoints: Many people have used eigenvector techniques to summarize data: The technique has been popular as a way of detecting faces. In EigenPoints we augment the gray-scale image space with a vector of control points. Then, given

These control points were automatically found using EigenPoints.

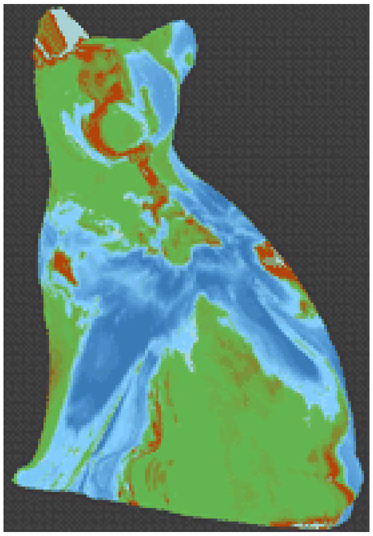
a new image, we can automatically find the associated control points, making facial analysis like that needed for Video Rewrite much easier and more efficient. See also Auto-correspondence.

Auto-correspondence: Auto-correspondence allows images to be automatically morphed. Morphing allows one image to be smoothly changed into another, but usually requires hand-positioned features to be identified in the two images. Using key features of the images, auto-correspondence allows match points to be defined automatically. These match points are used by a conventional morphing program to change the first image into the second. EigenPoints is one example of auto-correspondence. Other approaches that we've used have avoided building models of the images being morphed, instead using cross-correlation and feature-matching approaches.



These corresponding line segments were automatically matched between the two images.

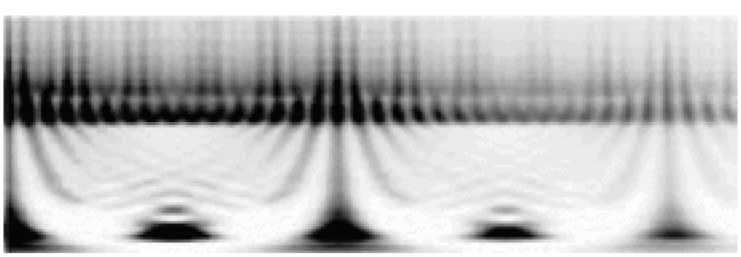
Compression on Manifolds:



The earth's elevation data is mapped and compressed on the surface model of a cat.

We are interested in compression of data defined on complex geometries, such as spheres, general 3D objects, as well as higher dimensional manifolds. Traditionally, data compression methods have been applied to functions defined on simple manifolds like the real line (e.g., audio), a rectangle (e.g., images), or a three-dimensional open-ended box (e.g., video). In order to compress functions defined on high dimensional manifolds, we introduce an approach that combines discrete wavelet transforms with zerotree compression. The result is a highly compressed (more than 100:1 for visually lossless), progressively embedded data representation.

Mach1: Mach1 is a technique for nonuniform time compression of speech, designed to mimic the natural timing of fast speech. There are many techniques that speed up speech, without changing its pitch, but the results are often unnatural. Mach1 looks for the port ions of the speech signal that human speakers compress when they are speaking rapidly. Those portions of the speech signal that are compressible are sped up much faster than other portions.



Auditory Perception: We've been investigating how people perceive

This correlogram is a model of auditory perception and shows the response to a sung vowel.

sounds in a noisy environment. How is it that we can understand one voice in a cocktail party? What is the basic perceptual machinery that allows us to perform this task?

Demographic and Lifestyle Studies

Consumer demographic studies, lifestyle studies, ethnographic field and home studies. Studies of choice in technology related to style, life-stage, spirituality, region, and gender.

Some members of Interval's Research Staff spend much of their time "in the field" talking to and observing people at home, work, and play. Our efforts focus on people's identities, values, sociability, communication habits, purchasing decisions, reminiscence behaviors, sense of community, personal styles, and home decorating styles. We are concerned with their joys, frustrations, worries, and wishes concerning technology.

While consumers are certainly the most qualified experts on their current lives, they are not good predictors of what their lives will be like in the future. Because of this, Interval conducts very basic research to get a better understanding of the fundamental desires, motivations, and concerns that shape people's lives. We combine open, exploratory data collecting methods with rigorous analysis of the results. This approach, which is called Applied Exploration, was developed by Interval researcher Bonnie Johnson and Christopher Ireland of Cheskin Research, a market

Researchers Interview a Subject Outside Her Home research company. An Applied Exploration project about girls and play that we began with Cheskin Research in 1993 evolved into the research for Purple Moon, a 1996 Interval spin-off company.

Some of our information is best gathered in focus group facilities. However, when we are embarking on an Applied Exploration study, we find we need to get out of the confines of a focus group setting to capture the context of a person's own "place." Interval has conducted Applied Exploration studies in locations as varied as private homes, nursing homes, offices, and music festivals. We catch glimpses of current and future prospects for technology by observing how people arrange their furniture, technology, and home entertainment devices. Understanding basic behaviors, such as where people go and what they use throughout the day, helps us develop insights into consumer behavior that research subjects themselves often cannot know today. People we interview often comment on how much they have learned about themselves by participating in our study.


An Informance at Interval

One of the ways in which data from the field is used at Interval is through Informance, a design tool developed by Interval researchers Eric Dishman and Colin Burns.

Informance, or informative performance, helps designers get into other people's shoes. Researchers immerse themselves in the lives of people that have been studied and videotaped in the field. By performing as those people, simulating the same work place or home space, the researchers brainstorm, "bodystorm," and build scenarios as a way of inspiring design.

In 1994, Interval sponsored the Electric Carnival, a tent of sixty-five interactive media exhibits that traveled with the Lollapalooza concert tour. Fifteen researchers from Interval and Cheskin Research conducted interviews in the Electric Carnival tent, around the concert venue, and, subsequently, in focus facilities. The data questionnaires, field video and photographs, focus interviews, and home photos provided us with rare insights into how these teens and young adults view themselves, the world around them, and prospects for technology.

Research Team at Lollapalooza

Field researchers at Interval have collaborated with the HomeNet project. HomeNet is a field trial at Carnegie Mellon University whose purpose is to understand people's use of the Internet at home. Starting in 1995, HomeNet provided 100 families with hardware and Internet connections. Interval is one of the high-technology companies sponsoring HomeNet, and has also collaborated with the Carnegie team on in-home interviews to learn more directly about participants' lives as a context for their use of the Internet.

To learn more about Informance, read the paper written by Eric Dishman and Colin Burns for CHI '96: "Actors, Hairdos and Videotape," Informance Design Conference companion volume to ACM Conference on Human Factors in Computing Systems (CHI'94), p.119.

To learn more about Applied Exploration, read the paper written by Christopher Ireland and Bonnie Johnson: Christopher Ireland, Bonnie Johnson, "Exploring the Future in the Present," Design Management Journal, 6(2), pp. 57-64, Spring 1995.

Here are some of the people we have studied:

1992 Elders in a nursing home – everyday life and technology in an elder home environment.

1993-6 Children – the meaning of "play" for girls and boys. Observing girls at play, and asking girls about play, informed the Interval project that became Purple Moon.

1994 Teens and Young Adults at Lollapalooza – new forms of technology interactivity, new directions in music, products in the future from teens today.

1995 Young Adults just out of college – what happens when easy access to free technology ends?

1996-7 Families in their homes – consumer values, communication habits, purchasing habits, identity issues, social practices, and reminiscence behaviors.

1996-7 HomeNet Families – how does easy Internet access change a family's daily life?

Media Streams: An Iconic Visual Language for Media Annotation and Retrieval

Over the next decades, we will take part in a revolution in media creation by people who could not earlier afford to produce video in their homes, schools, and offices. Just as desktop publishing gave consumers the power of the printing press on their desks (but it took the Internet to make everyone a publisher since without it the distribution channel was lacking), and digital audio samplers gave birth to a whole new genre and population of music makers, home media tools will enable consumers to make video creation a part of their daily communication and entertainment. In the spirit of garage bands, we can think of this new population of motion picture producers as practitioners of "Garage Cinema." These are the people who in the next century will be running a TV station/movie studio out of their homes.

For home media producers, a major obstacle is getting access to media content in order to be able to tell a wider range of stories than they can record or synthesize. The other major challenge lies in having tools which enable them to manipulate media according to their content rather than requiring the specialized skills needed in current motion picture and video production. Once these challenges have been addressed, we can imagine a world in which digital media are produced anywhere by anyone and are accessible to anyone anywhere.

In the future, annotation— the description of the structure of media content— will be fully integrated into the production, archiving, retrieval, and reuse of media data. However, there will remain many annotations which computers won’t be able to automatically encode. A central challenge for computational media technology is to develop a language of description which both humans and computers can read and write and which will enable the integrated annotation, creation, and reuse of media data. In order to overcome the inherent limitations of current keyword-based video annotation and retrieval systems, we need representations that capture the temporal, semantic, and relational content of video data. These representations also need to be convergent and scaleable to a global media archive. We have developed a language for the representation of video content that addresses these issues.

Our prototype system, Media Streams, is an iconic visual language for annotating, retrieving, and repurposing digital video and audio. Within Media Streams, the organization and categories of the Icon Space allow users to browse and compound over 4000 iconic primitives by means of a cascading hierarchical structure that supports compounding icons across branches of the hierarchy. A Media Time Line enables users to visualize, browse, annotate, retrieve and repurpose streams of video and audio content. Media Streams was first developed at the MIT Media Laboratory by Marc Davis, Brian Williams, and Golan Levin, and is an ongoing part of our research in home media tools at Interval.

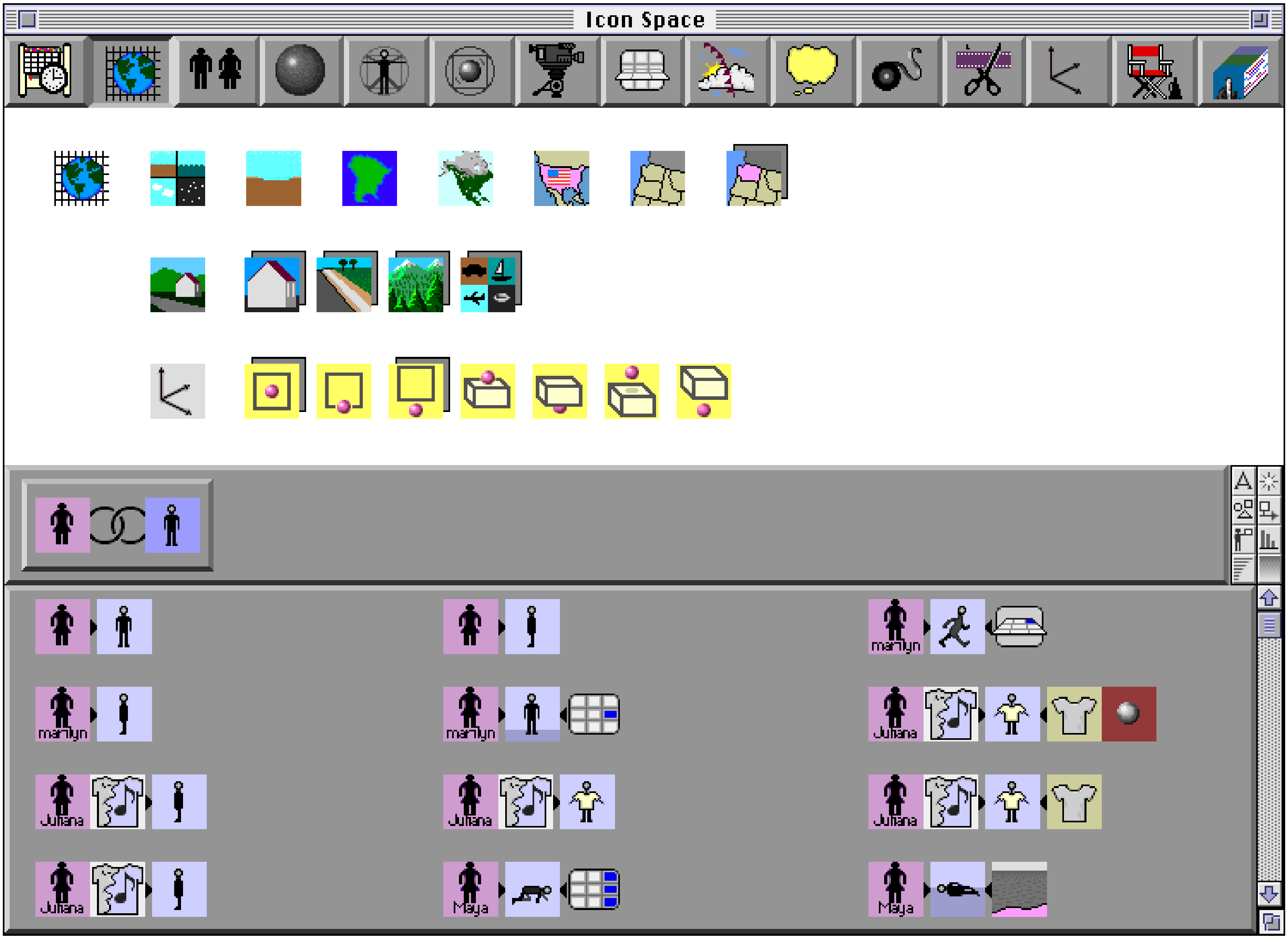


Figure 1: Media Streams Icon Space

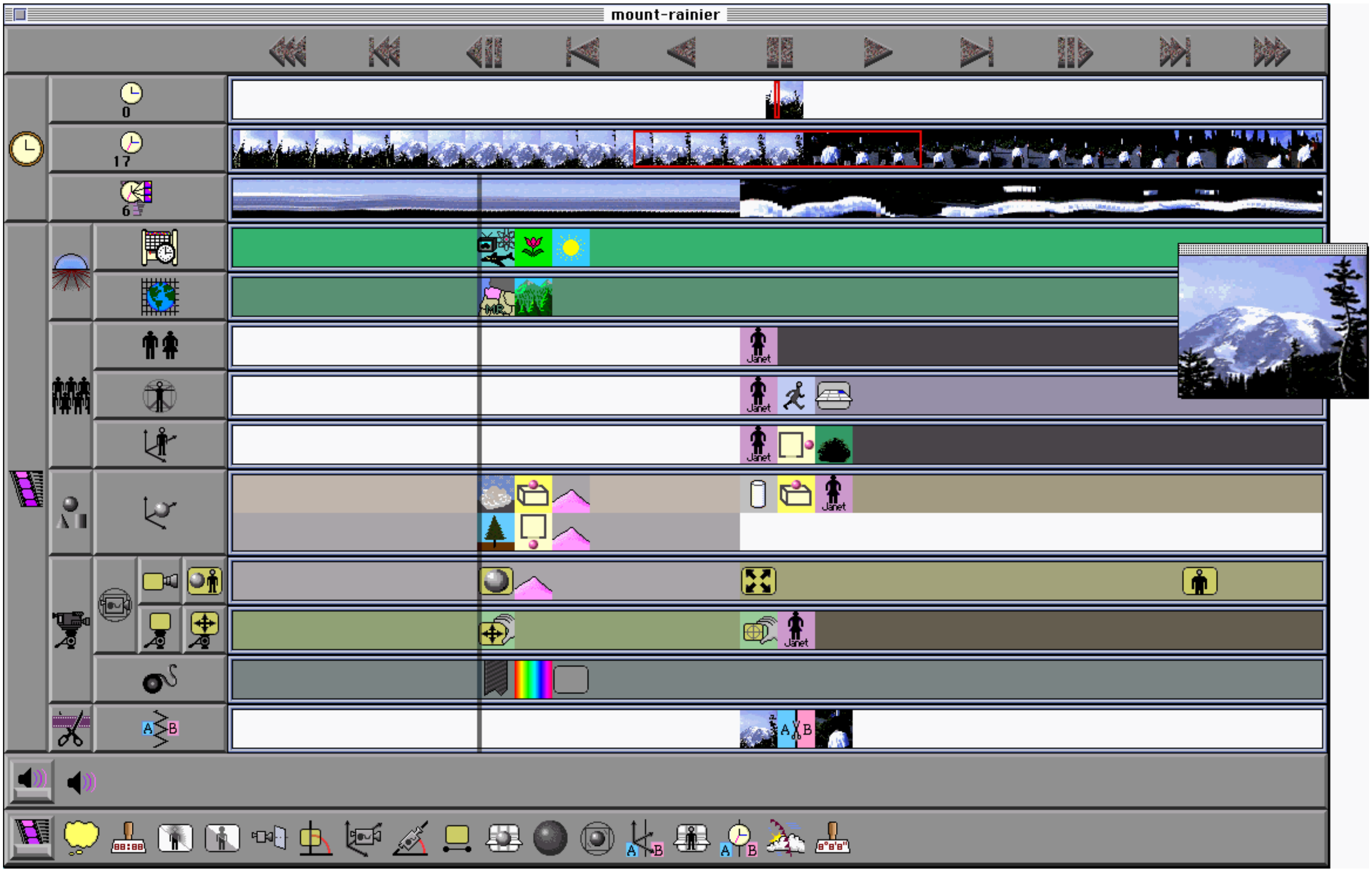


Figure 2: Media Streams Media Time Line

For more information about Media Streams, please refer to the following articles:

Davis, Marc. "Media Streams: An Iconic Visual Language for Video Representation." In: Readings in Human-Computer Interaction: Toward the Year 2000, ed. Ronald M. Baecker, Jonathan Grudin, William A. S. Buxton, and Saul Greenberg. 854-866. 2nd ed., San Francisco: Morgan Kaufmann Publishers, Inc., 1995.

Abstract:

In order to enable the search and retrieval of video from large archives, we need a representation language for video content. Although some aspects of video can be automatically parsed, a sufficient representation requires that video be annotated. We discuss the design of a video representation language with special attention to the issue of creating a global, reusable video archive. Our prototype system, Media Streams, enables users to create multi-layered, iconic annotations of streams of video data. Within Media Streams, the organization and categories of the Icon Space allow users to browse and compound over 3500 iconic primitives by means of a cascading hierarchical structure that supports compounding icons across branches of the hierarchy. A Media Time Line enables users to visualize, browse, annotate, and retrieve video content. The challenges of creating a representation of human action in video are discussed in detail, with focus on the effect of the syntax of video sequences on the semantics of video shots.

Keywords: video archiving, visual language, video indexing, video retrieval, knowledge representation, multimedia, visualization, iconic language, film theory, graphical user interface design, repurposing, computational cinema

Davis, Marc. "Garage Cinema and the Future of Media Technology." Communications of the ACM (50th Anniversary Edition) 40 (2 1997): 42-48.

Abstract:

The twentieth century saw the invention and development of two fundamental, new technologies for creating and manipulating representations of the world: motion pictures and computation. Motion pictures gave us the ability to capture and construct sequences of moving images that enabled the creation of a new language of storytelling and visual experience. Computation provided us a method of constructing universal machines which, by manipulating representations of processes and objects, can create new processes and objects, and even new machines. The deep integration of computation and motion pictures has not yet occurred. The implications of their deeper integration over the next fifty years will have profound technological, linguistic, and social effects. This article traces part of the history and future of computational motion pictures as well as the cultural factors this technology will draw on and foster.

Keywords: visual language, iconic language, semasiography, film theory, multimedia, computational cinema, mass media, popular culture, linguistics, repurposing

Bounce: A Visual Language for Media Software Creation

Within Interval, programs like Bounce take the notion of Lego™-like *kits* whose parts know how to connect, and applies them to software and media creation itself.

Bounce turns software creation into a play activity by providing enough constraints that functional components can 'snap together' and operate instantly as you connect them. Icons represent functional behavior, and simple tools for connecting them let even non-programmers create interactive animation, simulations, characters, and musical instruments -- without waiting, compiling, worrying about typos, or typing at all. Bounce's precursors include *HookUp!* (from David Levitt's Hip Software) and Chuck Blanchard's *Body Electric*.

Bounce and other Interval tools offer ways for researchers to rapidly create new experiments or home media prototypes out of simple parts. Using constrained kits made of understandable parts, we can picture a time where artists, content producers, and even children can understand software processes and create and modify their own with pleasure.

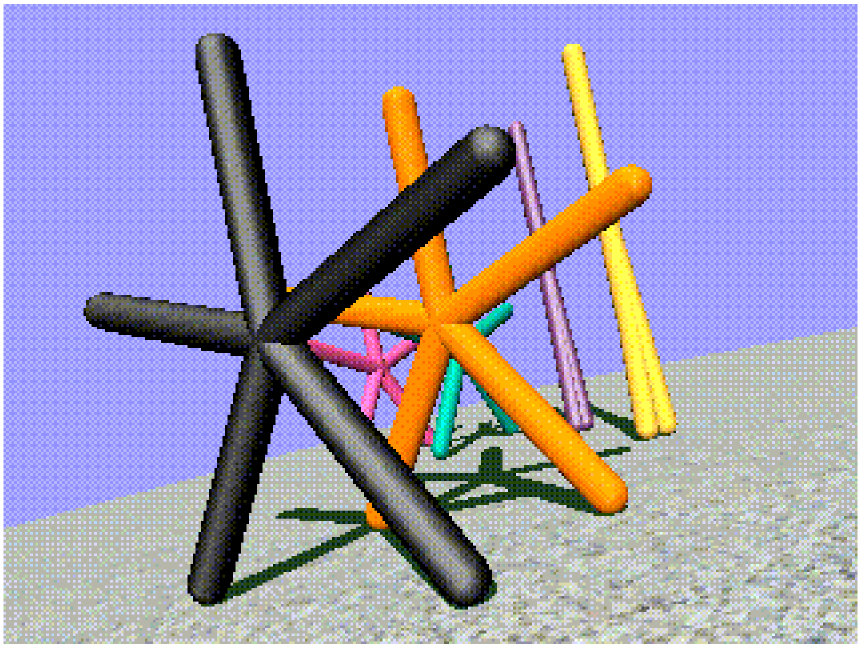
Spacetime Constraints Revisited: Helping the user animate through physics and learning

Animation, whether digital or traditional, is always a time-consuming task. It takes years of experience and infinite patience on the job before one is able to create motion that looks appealing. A master of this art is likely to care every nuance, down to every twitch of the eyes. A casual user, on the other hand, may have only some high-level behavior in mind, such as walking or making a transition from one pose to another. She may be willing to sacrifice control over details in return for not having to worry about them.

One way to relieve the user of these burdens is to incorporate constraints that come from the natural world. We have investigated the use of physics to serve in this role, though other rules and procedures can be and have been used. We have used learning techniques that allow an animated creature to "learn" to perform. A controlled amount of randomness lets the creature exploit lessons from its experience, yet explore new motion strategies.

This combination of techniques has advantages and disadvantages. Delightfully, it surprises the user with unexpected motions. These are often more intricate than one could realistically expect from a human animator, yet seemingly the product of intention. On the other hand, such surprises can make this approach impractical for an animator who has detailed preconceptions about how the resulting motion should appear. We are studying ways to give control back to human users without forcing them to worry about details beyond their skill or patience.

This work was done as part of a collaboration with a number of researchers outside Interval. It refines and extends dissertation work done by Tom Ngo at Harvard University in collaboration with Joe Marks.



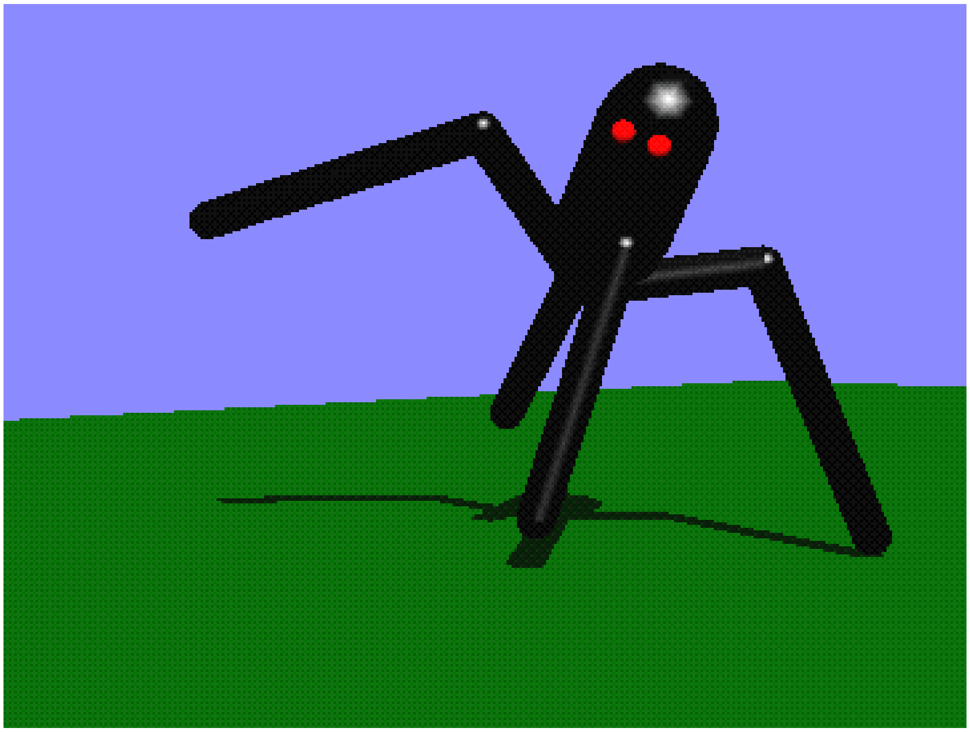
J. Auslander, A. Fukunaga, H. Partovi, J. Christensen, L. Hsu, P. Reiss, A. Shuman, J. Marks, J.T. Ngo, "Further Experience with Controller-Based Automatic Motion Synthesis for Articulated Figures," ACM Trans. on Graphics, 14(4):311-336, Oct. 1995.

Abstract:

We extend an earlier automatic motion-synthesis algorithm for physically realistic articulated figures in several ways. First, we summarize several incremental improvements to the original algorithm that improve its efficiency significantly and provide the user with some ability to influence what motions are generated. These techniques can be used by an animator to achieve a desired movement style, or they can be used to guarantee variety in the motions synthesized over several runs of the algorithm. Second, we report on new mechanisms that support the concatenation of existing, automatically generated motion controllers to produce complex, composite movement. Finally, we describe initial work on generalizing the techniques from 2D to 3D articulated figures. Taken together, these results illustrate the promise and challenges afforded by the automated motion-synthesis approach to computer animation.

Keywords:

Animation, spacetime constraints, heuristic methods, machine learning, stochastic optimization, evolutionary computation, automated choreography



J. Christensen, J. Marks, J. T. Ngo, "Automatic Motion Synthesis for 3D Mass-Spring Models," The Visual Computer, 13(1):20-28, Jan. 1997.

Abstract:

We describe how to synthesize, automatically, motion controllers for locomotive tasks involving animated characters modeled as 3D mass-spring lattices. The motion controllers determine an actuation sequence based on elapsed time, not physical state. Actuation is represented economically with a small, predefined set of global lattice deformations, and stochastic search is used to determine effective values for the controller parameters. Our algorithm generates controllers that produce attractive, visually plausible motion for simple locomotive tasks in under an hour on a standard workstation. This is more than an order of magnitude faster than comparable approaches to motion synthesis for 3D articulated-linkage models.

Keywords:

Animation, motion synthesis, controller synthesis, heuristic methods, stochastic optimization, mass-spring models