M A R C D A V I S
www.marcdavis.me

info@marcdavis.me

Editing Out Video Editing

Bibliographic Reference:

Marc Davis. "Editing Out Video Editing." *IEEE MultiMedia*, 10 (2). April–June 2003. 54–64.

Editing Out Video Editing

Marc Davis
University of California at Berkeley

This article outlines a paradigm shift in media production: the advent of computational media production that will automate the capture, editing, and reuse of video content. By integrating metadata creation and (re)use throughout the media production process, we'll enable the mass customization of video.

For the majority of people to not just watch, but make video on a daily basis, the current media production process must be transformed from a mechanical process to a computational one. To investigate this transformation, let's begin with a thought experiment—consider these four simple questions:

1. How many of you reading this article read text on a daily basis (including email, Web pages, newspapers, and so on)? Answer: All of you.

2. How many of you write text on a daily basis (including email, notes, and so forth)? Answer: Very nearly all of you.

3. How many of you watch video on a daily basis (including all media forms of moving images with sound—movies, television, and videos). Answer: Many more of you than you might care to admit.

4. How many of you make video on a daily basis (including all media forms of moving images with sound—movies, television, and videos)? Answer: ?

Although a majority of *IEEE MultiMedia* readers work on digital media technology, probably only a few of you answered "yes" to the fourth question. In this article I describe the technology and a vision for computational media production that will enable us to answer "yes" to the fourth question. I initially articulated this vision of the widespread daily production and sharing of video content for the 50th anniversary edition of the *Communications of the ACM*.[1] In that article, the key technological challenges I identified that would enable a new "garage cinema" were tools for accessing and manipulating content. While the challenges of making media accessible and manipulable are still with us, my further research has revealed that for video to become a daily means of communication we need to invent new forms of computational media production that don't merely computerize existing media production tools and methods.

What's wrong with media production today?

The current paradigms for media production arose within the context of social, economic, legal, and technological factors that differ greatly from a situation in which video could be a widespread, daily medium of many-to-many communication in the way that text is today.

It takes teams of highly trained people to create professionally produced video content (including movies, television, and videos) for mass audiences. The professional video production process usually requires three distinct phases:

1. *Preproduction*: concept formation, scriptwriting, storyboarding, and production planning.

2. *Production*: video and audio recording.

3. *Postproduction*: video and audio editing, special effects, soundtrack composition, and rerecording video and audio.

This media production methodology is expensive in time, money, and human resources. It's especially time-consuming and can range from many hours (television news) to many years (Hollywood feature films). The current media production process requires a variety of expertise at each step as well as expensive personnel and equipment. The media production process itself is also often wasteful of effort and lossy of information.

Rarely do we orient current media production toward creating reusable media assets. Because of the difficulties of finding and reusing appropriate media assets, new production often occurs when media reuse could have been an option. In addition, we create far more footage in professional media production than we use in the final edit (from 10 to 1 to 100 to 1, depending on the type of production). Furthermore, almost all metadata created during the various media production phases are neither easily available to subsequent

1070-986X/03/$17.00 © 2003 IEEE

Published by the IEEE Computer Society

From Mechanical to Computational

To envision the future of media production, it's helpful to understand its past. The invention of motion pictures occurred in the late 1890s and while the technologies for motion picture cameras and editing systems have changed, the underlying paradigms of capturing and manipulating motion pictures haven't (see Figure A). First let's examine the history of motion picture capture technologies. This history is detailed and complex,[1] but for our purposes, a thumbnail sketch will suffice.

In 1895, the Lumière brothers invented the Cinématographe, a portable film camera that also functioned as a film developing unit and projector. In 1951, the first videotape recording was invented for use in television broadcasting, and it took another 20 years—until the invention of the Port-a-pak—for a video camera connected to a videotape recording unit to become portable.[2] With the advent of the video camcorder in the 1980s, we've nearly come full circle to the capabilities of the Lumière Cinématographe. While modern video camcorders use computation to automate many important functions related to capturing focused, properly exposed video, the paradigm of video capture (like that of video editing) has remained unchanged. Video camcorders encode almost no metadata about what they're recording and don't proactively assist in capturing video assets that we can easily edit and reuse—they simply record moving images and sound.

Motion picture editing technology underwent three major technological phases in the last century:

- physical film cutting,
- electronic videotape editing, and
- digital nonlinear editing.[3]

Film editing began with the direct manipulation of the film reel: a *cut* was literally that, the cutting of the film reel at one point and the taping together of it and another length of film at the cut point. The invention of electronic videotape editing in the early 1960s, while a cost and time savings over film editing, was actually a step backward in the manipulability of motion picture content given the difficulty of insert editing and the invisibility of all but the current frame being used. The advent of nonlinear editing in the late 1980s was in many ways a return to the affordances of early film editing with its ability for the editor to both see and manipulate sequences of frames. While the technological developments in motion picture editing in the last part of the 20th century involved computers, the paradigm of video editing remained and remains mechanical rather than computational. The most advanced video editing software today still manipulates motion picture content in much the same way as the first film editors. They provide direct manipulation interfaces for cutting, pasting, and trimming sequences of frames, while both the representation of the motion picture content and the methods of its restructuring reside in the editor's mind. What we need is technology that can embody knowledge of both motion picture content and structure into a new computational media production process.

I'd like to further situate the technologies and methods of media production within the history of industrial production

continued on p. 4

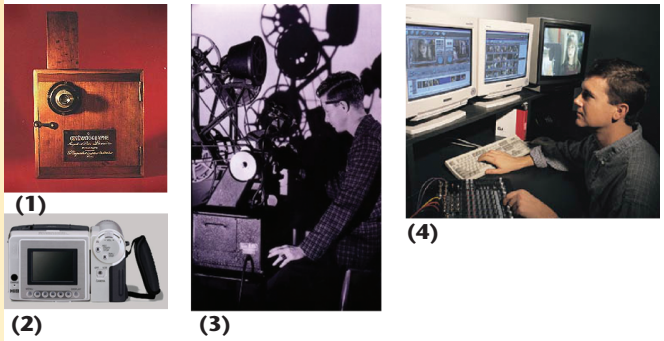

Figure A. While motion picture technology has changed in the last 100 years, the underlying paradigms for motion picture capture and editing haven't. (1) Lumière Cinématographe. (2) Viewcam camcorder. (3) Moviola film editing machine (courtesy of the University of Southern California's Moving Image Archive). (4) Nonlinear editing suite.

phases, nor easily accessible after finishing production. As a result, most research on creating video metadata through signal analysis attempts to recreate metadata that was available (but lost) at various points in the production process. The expense and inefficiencies of current professional media production are symptomatic of it still being in the craft mode of production (for more about the history of media production, see the sidebar "From Mechanical to Computational").

Amateur video production by consumers using video camcorders is fraught with its own difficulties because the tools and methods available to consumers are modeled after those of professional video production, while consumers usually possess neither the time, money, nor expertise that professional production methods require. Quite simply, far more amateur video is shot than watched, and people almost never edit it. Introducing supposedly consumer-friendly video editing tools such as Apple's iMovie doesn't solve the problem of consumer video production except for the devoted hobbyist and enthusiast. For consumers to quickly, easily, and

continued from p. 3
processes. Industrial production underwent three major developmental phases in the last 300 years. Before the Industrial Revolution, skilled craftspeople custom-made goods for consumers. Production was individually tailored, but expensive in time and money. With the advent of standardized interchangeable parts and mass production processes, goods were no longer created for individual consumers, but could be produced cheaply and quickly. In the late 20th century, mass customization processes began to combine the personalization of the customized production of goods with the efficiencies of mass production methods.[4] When we look at modern media production in light of the development of industrial production methods, we see that media production is still largely in the mode of craft-based, customized production. Skilled craftspeople work to create one media production; they can then use mass production methods to reproduce and distribute the media production. However, the production process itself doesn't take advantage of the efficiencies of industrial production, especially of standardized interchangeable parts and production methods that don't require skilled craftspeople. For media production to become a process of mass customization, it must be transformed from a mechanical process to a computational one.

References

1. B. Salt, *Film Style and Technology: History and Analysis*, 2nd ed., Starword, 1992.
2. I. Fang, "Videotape," *The Encyclopedia of Television*, H. Newcomb, ed., Museum of Broadcast Communications, 1997.
3. J.T. Caldwell, "Video Editing," *The Encyclopedia of Television*, H. Newcomb, ed., Museum of Broadcast Communications, 1997.
4. B.J. Pine, *Mass Customization: The New Frontier in Business Competition*, Harvard Business School Press, 1993.

regularly produce and share video, we need a new mode that isn't based on the current professional or amateur production paradigm.

For the most part, adding computation to video production, both professional and amateur, has merely attempted to assist the current mode of media production. Instead, we should rethink and reinvent the relationship between video and computation to create a new mode of computational media production.

Computational media production

Motion pictures enable us to record and construct sequences of images and sounds. Computation enables us to construct universal machines that—by manipulating representations of processes and objects—can create new processes and objects, and even new machines. Computational media production isn't about the application of computer technology to the existing media production methodology; it's about the reinvention of media production as a computational process. To reimagine movies as programs, we need to make motion pictures computable. That is, rather than producing a single motion-picture artifact, we produce a computer program that can, based on input media and parameters, produce motion pictures. Consequently, media production shifts from the craft-based production of single movies to the production of computer programs that can output mass-customized movies.

To illustrate the idea of mass-customized computational media, consider the famous Budweiser Super Bowl commercial in which a sequence of friends asks each other "Wassup?" on the telephone. Imagine this same commercial personalized to the level of each individual viewer so that when you see the ad on TV you see your friends and yourself saying "Wassup?" and when I see my version of the same commercial I see my friends and myself. Current media production technology and methodology would render this scenario impossible because of the time and money required to custom shoot and edit the commercial for each Super Bowl viewer. With a computational media production process, software programs would assist consumers in the automated capture of the required video assets and automatically integrate them into the final personalized commercial based on a professionally authored template. Rather than merely handcrafting a single commercial that millions of viewers would see, for basically the same amount of effort, computational media production methods could produce millions of personalized and customized versions of the commercial. We can envision similar examples for other commercials, movie trailers, music videos, and video greetings and postcards.

My project teams and I have been working on the concepts and prototypes for computational media production over the last decade. Initially, we conducted research at the MIT Media Lab, then at Interval Research, at Amova (http://www.amova.com), and now at my Garage Cinema Research group (http://garage.sims.berkeley.edu) at UC Berkeley's School of Information Manage-

ment and Systems (SIMS). As the diagram (Figure 1) from our time-based media processing system patent[2] illustrates, computational media processing uses content representation and functional dependency to compute new media content from input media content and stored assets.

We used extensions to this simple computational framework to systematically automate many of the functions human editors perform: audio–video synchronization, audio substitution, cutting on motion, L-cutting for dialogue, and postproduction reframing of shots. By reimagining media editing as a computational process, we also developed a new class of parametric special effects by which we can automatically manipulate video and/or audio as a function of continuous parameters of other media. An example is to automatically rumble video to a music soundtrack by having the video's scale and position be a function of the music's amplitude.[2] Note that these media automation functions (and structures and applications built with them) aren't used to make better or more automated editing tools for consumer video production. Rather, our intent is to remove the editing process as we know it from consumers and replace it with a variety of quick, simple, and satisfying ways that users can create, shape, and share video content. The most complex "editing" process a consumer should ever have to perform should resemble the simple selection and control processes involved in using a TV remote.

To make video computable, we need to represent its content (especially its semantics and syntax). This effort must address the semantic gap between the representations of video content that we can derive automatically and the content and functional knowledge that cinematographers, directors, and editors use in the traditional media production process.[3] We address this semantic gap by

- developing metadata frameworks that represent the semantic and syntactic structures of video content and support computation with and reuse of media and metadata;
- integrating metadata creation and reuse throughout the computational media production process;
- redesigning media capture technology and processes so as to facilitate capturing metadata at the beginning of the production process and proactively capturing highly reusable media assets; and
- inventing new paradigms of media construction that automatically recombine and adapt media assets by using representations of media content, semantics, and syntax.

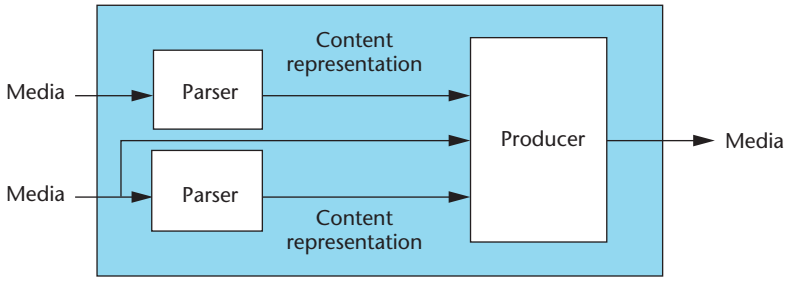

Figure 1. Movies as programs is the central idea of the time-based media processing system.

Metadata and media reuse

Since 1990, my project teams and I have developed a comprehensive system for video metadata creation and reuse: Media Streams.[4,5] Media Streams provides a standardized iconic visual language using composable semantic descriptors for annotating, retrieving, and repurposing video content. Recent standardization efforts in MPEG-7 are also working to provide a uniform and reusable framework for media metadata.[6] To create metadata that describe the semantic and syntactic content of motion pictures, we can draw on semiotic[7,8] and formalist[9,10] film theoretic analysis of the structures and functions of motion pictures. Computational media aesthetics researchers[3] have a key role—not only in digital technology, but in film theory itself—in developing theoretical and practical frameworks for describing computationally usable and generable media. We need to develop frameworks that are not only analytic (designed to parse motion pictures into their structural and functional components) but also synthetic (designed to support the production, combination, and reuse of motion picture content).

By readjusting our focus on the synthesis of motion pictures, we can reenvision the opportunities for creating and using media metadata. While Media Streams supports retrieval-by-composition methods for media synthesis (assembling sequences from disparate database elements in response to user queries), the system's main usage paradigm is for annotating video content (analysis) in postproduction. Our recent work has focused on reinventing the production process to generate Media Streams' metadata throughout

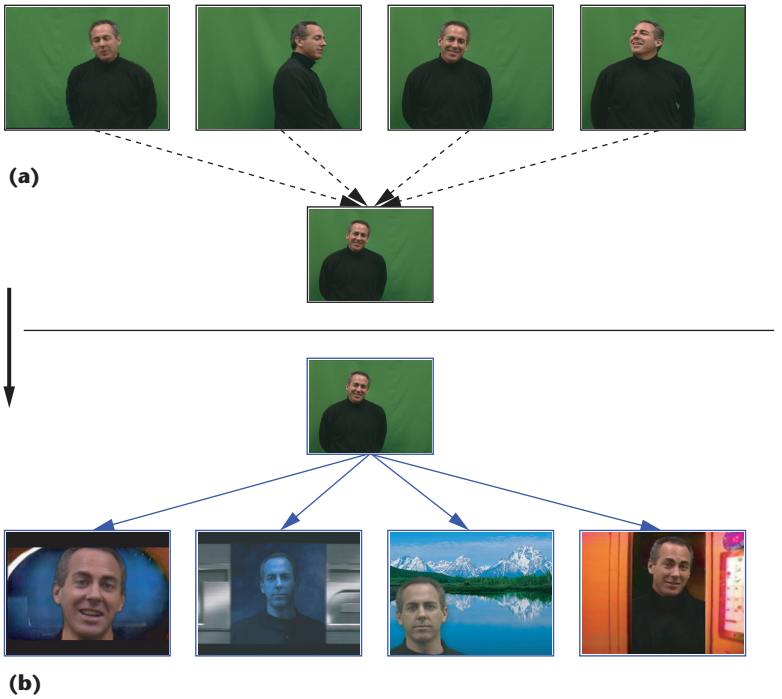

Figure 2. New capture paradigm for creating reusable media assets. (a) The current capture paradigm involves multiple captures to get one good capture. (b) With the new capture paradigm, one good capture drives multiple uses.

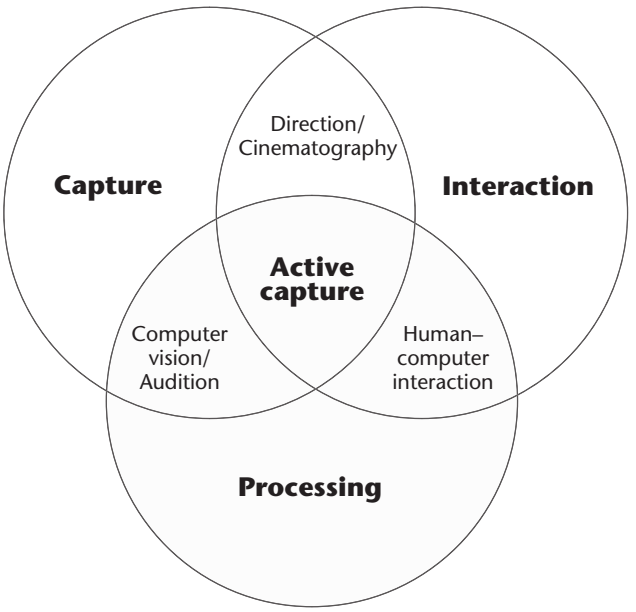

Figure 3. Active capture integrates capture, processing, and interaction.

the computational media production process. We especially focus on the point of capture, where so much of the metadata that analysis seeks to reconstruct is readily available.

Rethinking capture

What's the purpose of video capture? What type of production process is it designed to serve? Earlier in this article I alluded to the inherently wasteful process of current media production: where producers shoot 10 times (or more) the amount of footage than they use in the final edit and rarely reuse captured media assets in future productions.

How can integrating computation and video make the media capture and production process more efficient and yield reusable media assets? Researchers have been exploring creating "data-cameras" that encode camera parameters and, with human assistance, semantic information during capture.[11,12] While these approaches correctly attempt to frontload the annotation effort to the point of capture (and in some cases even earlier to the preproduction phase), they assume that the process and goals of video capture and production remain largely unchanged. The point of capture is the best opportunity to encode much useful metadata (camera settings, temporal and spatial information, and so forth), yet we can capture semantically richer metadata by rethinking the goals and process of video capture.

Our work in computational media production automates direction and cinematography to create a new media capture process whereby the capture device interacts with the user and the world to proactively capture metadata and annotated reusable media assets in real time. The key insight is to invert, as Figure 2 shows, the current media capture process. Rather than capturing 10 (or more) minutes of footage to get one usable minute, we redesign the process to capture 1 minute of (re)usable footage that could be used in creating 10 (or many more) minutes of media content.

To automatically capture a small set of highly reusable video assets, we've developed a new method of video capture that we call *active capture*. Active capture reinvents the media capture and production process by integrating capture, processing, and interaction (see Figure 3) into a control system with feedback connecting the capture device, human agents, and the environment shared by the device and agents. Active capture overcomes the limitations of standard computer vision and audition techniques by using human–computer interaction design to simplify the world and the actions that the vision (and audition) algorithms need to parse.

With active capture, the media capture

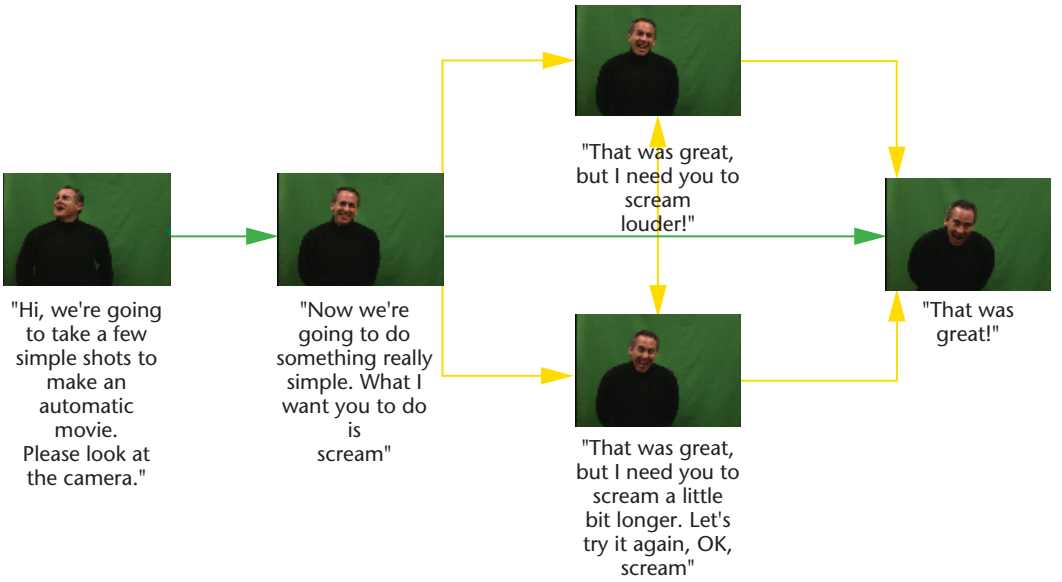

Figure 4. Active capture process for capturing a "scream" shot. Quotes are verbal instructions from the active capture device. The green arrows represent an error-free path. The yellow arrows are error-correction loops.

process can then become an interactive session between a recording device and its subject—much in the same way that a director and cinematographer work with actors today or the ways amateur photographers prompt or instruct their subjects. Importantly, the active capture process requires no special expertise on the part of the subject being captured. This low threshold for the subject of a computational capture process is important for its potential viability as a capture paradigm for consumers.

In portrait photography and motion picture direction humans both prompt and evaluate the capture subjects' responses. In active capture, both prompting and response evaluation can be achieved by the device itself. However, an active capture device uses audio and/or visual cues to prompt the capture subject to perform some desired action (such as smiling and looking at the camera). Through real-time audio and video analysis, an active capture device can determine the fitness of the subject's response in relation to some predetermined capture parameters. If the capture meets these parameters, the capture process is complete. If not, the active capture device can prompt the user again until it achieves a suitable response or the process has timed out.

Real-time audio and video analysis of the subject's responses can enable an active capture device to offer specific suggestions as to how to improve the subject's response. Figure 4 illustrates an example of active capture's control process with feedback. This example depicts the active capture routine for capturing a high-quality and highly reusable shot of a user screaming. To capture a shot of the user screaming, the system prompts the user to look at the camera and scream. The system has a minimal average loudness and overall duration it's looking for, and like a human director, it can prompt the user accordingly (such as scream louder **or** longer) to capture a loud and long enough scream shot.

By using a recording device that can interact with its subject (through verbal, nonspeech audio, and visual cues), we can quickly and easily capture reusable assets of people (reaction shots, focus of attention, and so on) that the system preannotates through the capture process. Metadata creation then is no longer predominantly a postproduction process, but a real-time and interactive process to create metadata and annotated reusable media assets at the point of capture. The active capture process can create an inventory of annotated reusable media assets that can serve as resources for automated media production and mass customization of video.

New paradigms for media construction

In designing and inventing new paradigms for computational media construction, I turned to childhood constructive play experiences that seemed effortless and satisfying: Mad Libs and Lego. These two toys let children quickly and easily construct complex, coherent structures in text and plastic. They leverage simple actions within the context of constrained structures to generate complex and varied outputs. These toys also provide two outstanding paradigms for computational media construction.

Computational Media Aesthetics

One of the _____ hurdles facing _____ media management systems is the semantic _____ between
ADJECTIVE ADJECTIVE NOUN

the _____ meaning that users desire to _____ when they query to _____
ADJECTIVE VERB (PRESENT TENSE) VERB (PRESENT TENSE)

and _____ media and the _____of the content descriptions that we can
VERB (PRESENT TENSE) ADJECTIVE +"NESS"

_____ compute today for media _____ and search.
ADVERB VERB ("ING" ENDING)

Figure 5. A Mad Lib version of the first sentence of the call for papers for this special issue of IEEE MultiMedia.

In 1953, Leonard Stern and Roger Price invented Mad Libs while working as comedy writers for the popular *Honeymooners* TV show. In Mad Libs, an existing syntactic structure (usually a brief narrative) has parts of its structure turned into slots for semantic substitution. To play Mad Libs, a person or persons are prompted to fill in the empty semantic slots, without foreknowledge of the syntactic structure in which they occur. Once all slots are filled, the completed Mad Lib is read aloud. To illustrate the concept, I created a Mad Lib version of the first sentence of the call for papers for this special issue of *IEEE MultiMedia* (see Figure 5).

Lego blocks are known and loved worldwide. Invented by master carpenter and joiner Ole Kirk Christiansen in the 1950s, Lego plastic blocks use a patented stud-and-tube coupling system that let the blocks be easily and securely snapped together and apart in a myriad of configurations. Lego blocks are a supreme example of combinatoric richness from constrained structure—just six 8-studded Lego blocks fit together in 102,981,500 ways.[13]

Before turning to the explanation of applying the Mad Lib and Lego construction paradigms to computational media, it's important to note that a different concept of authorship is embodied in the construction paradigms of Mad Libs and Lego than underlies popular conceptions of traditional media production and authorship. With Mad Libs and Lego, skilled craftspeople construct reusable artifacts whose constrained structures enable children to use them to create new artifacts. While this bricolage[14] process is highly creative, saddled with our inherited 19th century Romantic notions of creativity (in which an artistic genius creates something out of nothing), we can be blind to the actual process of creativity.

In the creative process, we always begin with a repertoire of preexisting cultural and formal materials within which we construct new artifacts and meanings. The continued marketing of cinematic auteurs (even though cinema is our most collectively produced art form) is symptomatic of these centuries-old notions of authorship, craft, copyright, and genius. In the 20th century, both art practice and theory have offered alternative notions of authorship (in which authorship is a process of recombining existing artifacts) that are more aligned with the construction paradigms of Mad Libs and Lego and help provide a framework for conceptualizing and situating new construction paradigms for computational media production.[15]

While Mad Libs are made of text and Lego blocks of plastic, both construction paradigms can be understood in semiotic terms. *Semiotics*[16] is the study of sign systems. *Semiotic systems*, with language as the primary example, are our fundamental means of human communication and sense-making with each other and the world. Semiotic systems are comprised of *signs*. Signs consist of two interrelated parts: the *signifier* (a mental acoustic image like the perception of the sight or sound of the word "tree" or an image of a tree) and the *signified* (the mental concept to which the signifier refers). Signs gain their definition and value by their differences from one another. Semiotic systems also have two major forms of organization that can be understood as orthogonal axes: the *paradigmatic* and *syntagmatic*. Figure 6 illustrates the interrelationship of these axes of organization.

The paradigmatic axis relates to signs that can be selected from a set of signs and thereby substituted for one another based on their function in the syntagmatic axis (semantics is the linguistic form of paradigmatic organization). The syntagmatic axis relates to signs that can be combined in sequential structures (syntax is the linguistic form of syntagmatic organization). For example, in the syntagm "My daughter likes tofu," "son" is an acceptable paradigmatic substitution for "daugh-

ter," while "eats" is not. Mad Libs and Lego blocks offer users differing constraints and opportunities in the paradigmatic and syntagmatic dimensions. In playing with Mad Libs, users provide acceptable paradigmatic substitutes for missing elements in a fixed syntagmatic structure—users don't write the sentences of Mad Libs, they provide their own words to complete them. Playing with Lego blocks involves constructing syntagmatic structures from a fixed set of paradigmatic terms—users don't make their own Lego blocks, they make structures with Lego blocks. These differing construction paradigms offer us compelling user interaction models for computational media production.

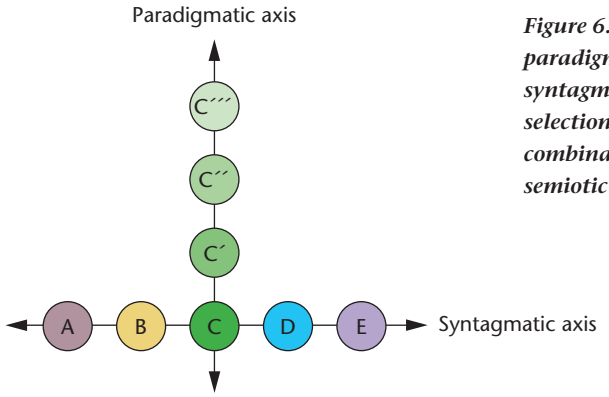

Figure 6. The paradigmatic and syntagmatic axes of selection and combination in semiotic systems.

Video Mad Libs

Based on the construction paradigm and user experience model of Mad Libs, we developed *adaptive media templates* (AMTs) that support automatic paradigmatic substitution and adaptation of new media elements within a set syntagmatic structure. In an AMT, the structure is largely fixed, while the content can be varied. AMTs have template assets (media that are referenced by the functional dependency network) and input media assets (assets that fill empty slots in the AMT and/or can be substituted for existing template assets). AMTs coadapt template and input media assets based on the content of the media assets and a set of functions and parameters to compute unique customized and personalized media results. To date, we've developed systems for authoring AMTs (MediaCalc written in Macintosh Common Lisp and MediaFlow written in C++ for Windows) and specific instances of these templates that use and produce television commercials, movie trailers, music videos, movie scenes, banner ads, and Macromedia Flash animations. Figure 7 shows a storyboard of an AMT for an MCI commercial that incorporates and adapts the scream shot captured in Figure 5. The commercial illustrates the speed of an MCI networking product by showing a rocket car and its pilot. In this case, the AMT version of the commercial processes the scream shot to substitute the subject of the scream shot for the pilot of the rocket car.

Template-based production has an important place in the history of consumer software. While many people may think of programs like Aldus PageMaker when remembering the desktop publishing revolution, for average consumers in the 1980s it was "The Print Shop" distributed by Brøderbund that revolutionized document production (especially of greeting cards, invitations, posters, and banners). The Print Shop's success was due in large part to its simple user experience that leveraged a large library of templates (syntagmatic structures) and clip art designed for use in these templates (paradigmatic elements). Similarly, computational media production methods let us radically simplify the media construction paradigm by offering users prebuilt AMTs and variable template assets, as well as the ability to easily capture input assets for use in templates.

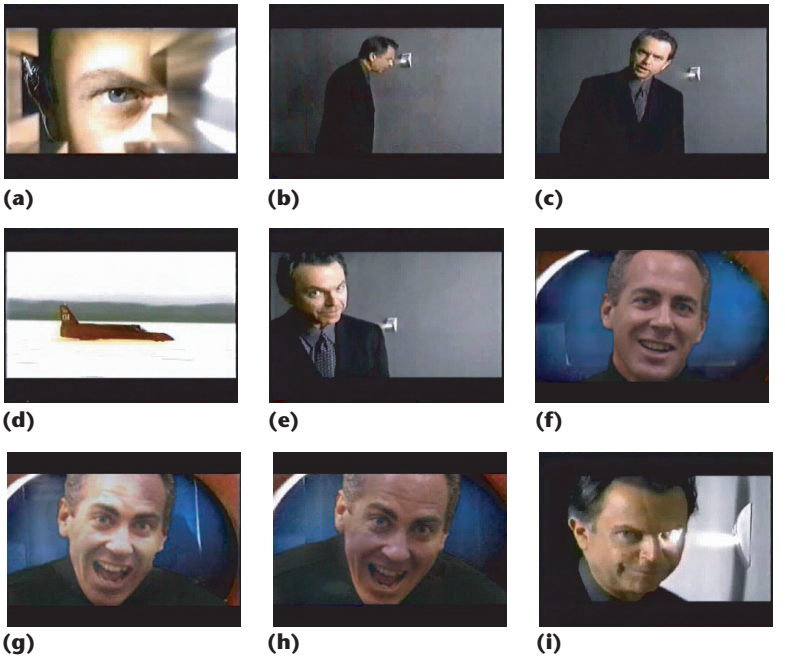

Figure 7. Adaptive media template storyboard for a personalized MCI commercial.

Video Lego

While still in the design stage, Video Lego will use and extend the construction paradigm of Lego blocks. Ideally, Video Lego will be a set of reusable media components that know how to fit together. Unlike the fixed syntagmatic structures of Video

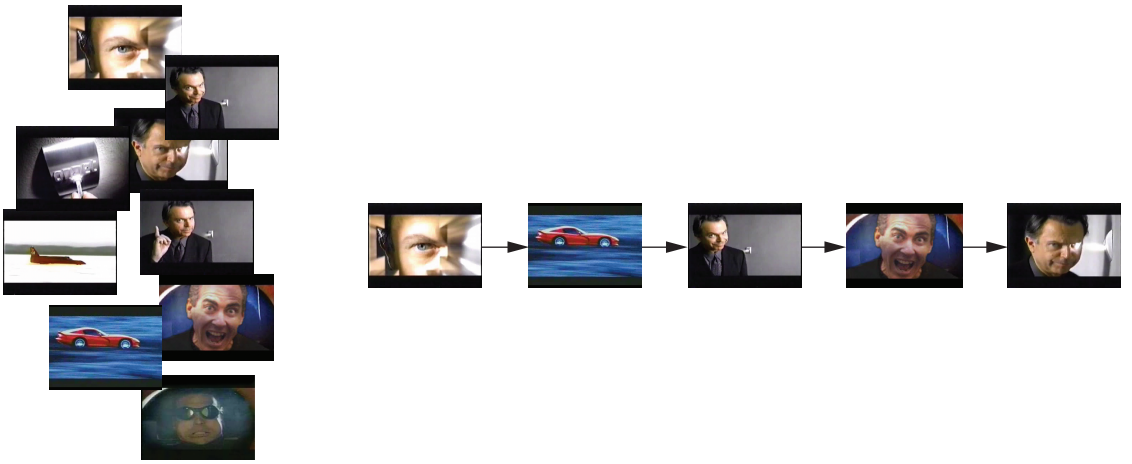

Figure 8. Visualization of MCI ad built with Video Lego.

Mad Libs, Video Lego will facilitate and automate the construction of a variety of syntagmatic structures from paradigmatic media elements (see Figure 8, next page). Furthermore, unlike physical Lego blocks, a Video Lego system will support creating new paradigmatic components. With the creation of Video Lego, Video Mad Libs will become a special case of highly constrained Video Lego structures made out of Video Lego components. The evolution of software design also provides an inspiration for the design of Video Lego components: the invention of object-oriented programming and reusable software components.[17]

Mass customization

Accustomed as we are to the commercial products of a craft-based media production process, it may seem counterintuitive that we could automatically mass customize movies from reusable components. Two assumptions should be made explicit to understand this transformation: the role of craft and the scope of media repurposability. While our technology envisions and enables the automated mass customization of media, it does so by redeploying existing knowledge inherent in craft-based media production in two ways. First, we encapsulate and represent media production and postproduction expert knowledge in our software programs in terms of how we describe media content and structure, and the functions and parameters that guide media adaptation and recombination. Second, the production process of AMTs involves craft-based production, not of single media artifacts, but rather of machines that can mass customize them. As a result, we now use cinematic production knowledge and postproduction to make programs that make movies.

The question of media repurposability is more complex. Work in film theory points to minimal levels of cinematic structure that form the basis of intrashot coherence, intershot montage, prenarrative action sequences, and narrative sequences.[7–10,18,19] While these low-level continuity systems provide an essential framework for repurposing content to create coherent sequences, the work of the avant-garde, music video, and fan video production point toward other possible continuity systems for repurposing. Movies made out of parts of other movies won't be Hollywood feature films (though *Dead Men Don't Wear Plaid* may be an important exception), but they'll usher in new possibilities for human expression in much the same way that digital sampling has begun to reshape popular music.

Existing professional motion picture production practice involves a certain amount of repurposing of elements (from the reuse of sets and props to the incorporation of stock footage). However, it's in the practice of television fans who repurpose program content to make new music videos and television scenes that we see the harbinger of things to come.[20] Television fans repurpose and personalize popular television content to express and share their personal and collective desires. Recombination of video elements happens on a variety of levels including, in addition to traditional cuts editing, the composition of various image planes and the recombination of visual and audio elements (music soundtrack and voice-over narration can be especially powerful repurposing techniques). Finally, the connective tissue of popular culture and intimate social groups can overcome many of the limitations of traditional narrative and continuity principles to engender new media forms that hybridize personal and popular content.

Finally, why not use 3D computer graphics to

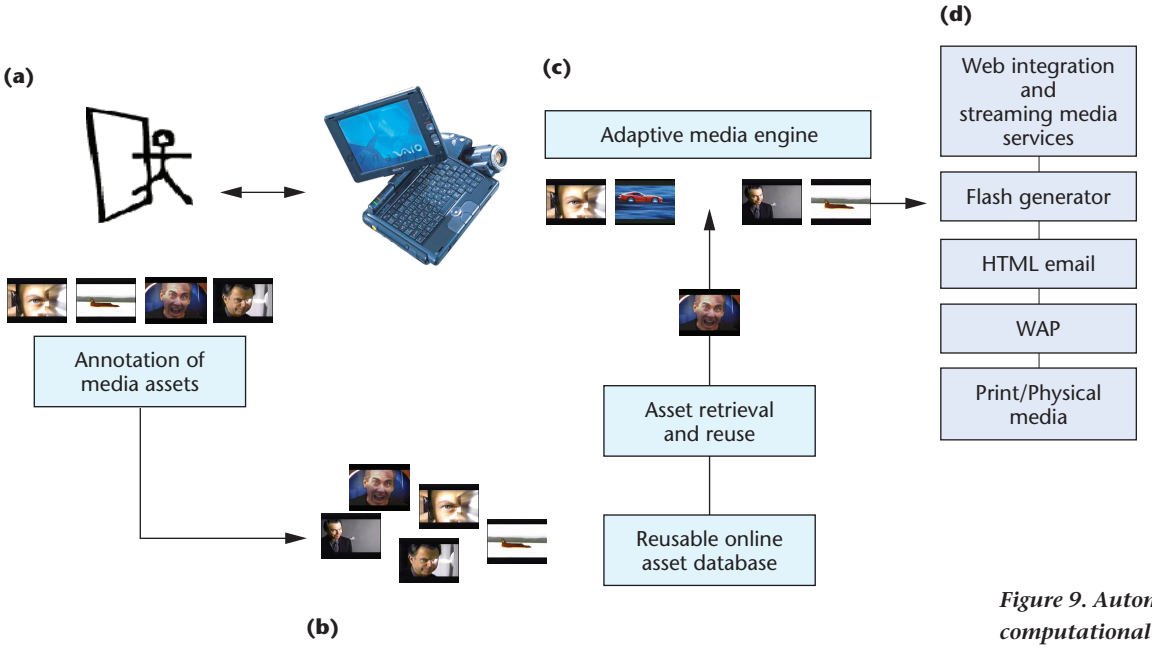

Figure 9. Automated computational media production process. (a) Automated capture, (b) annotation and retrieval, (c) automatic editing, and (d) personalized and customized delivery.

avoid the difficulties of recombination of motion picture elements? The knowledge and technology of paradigmatically and syntagmatically recombining media and subsequently automating the media production process applies as well to 3D computer graphics. Both for video and for computer graphics, we need semiotically informed technologies for computational media production that capture and automate the use of descriptions of media content and structure.

Toward new user experiences

By reinventing media production as a computational process, we can remove the obstacles to consumer participation in motion picture production, sharing, and reuse. Rather than directly manipulating media assets and attempting to cut and paste them into sequences, consumers can use new computational media production paradigms. These systems enable automated capture and automatic assembly by integrating metadata throughout the production process (see Figure 9) and by applying functions that can compute motion pictures according to representations of their content.

With computational media production methods, users can

- interact with a capture device that guides the capture process and automatically outputs reusable shots and edited movies;
- describe the sequence they want to see and view the results computed automatically;
- select media assets and see edited sequences featuring them computed automatically;
- select media sequences and easily vary the media assets within them; and
- play with media asset combinations and easily construct well-formed sequences.

Future work

Technologies and methods for componentized recombinant production can bring great economic and societal benefits.[21] We must therefore work not only to address the technical challenges to making computational media production a reality, but also to transform the barriers that exist in the spheres of law and public policy for digital media copyright and fair use.[22]

The key technical challenges that we must address are the continuing integration of signal processing, human–computer interaction, and media theory and practice to facilitate the creation and use of media metadata to automate media production and reuse.

We can achieve these research goals by crossing the disciplinary divide that separates the various communities of practice needed to design innovative solutions and thereby work together to bridge the semantic gap that separates what our systems can parse from the meaning people ascribe to media content. Through interdisciplinary research in multimedia we'll continue to invent new paradigms of computational media production that will

help us quickly, easily, and pleasurably produce, share, and reuse video content every day. **MM**

Acknowledgments

I want to give my deepest appreciation to the brilliant and energetic colleagues who have worked with me on this research at the MIT Media Laboratory, Interval Research, and Amova: Brian Williams, Golan Levin, Kenneth Haase, Henry Jenkins, Seymour Papert, Warren Sack, David Levitt, Gordon Kotik, Kevin McGee, Dave Becker, Steve Bull, Kevin Ramsey, John Szinger, Ron Levaco, Alan Ruttenberg, Neil Mayle, Malcolm Slaney, Nuno Correia, Rachel Strickland, Dick Shoup, Meg Withgott, Diane Schiano, Mitch Yawitz, and Jim Lanahan.

References

1. M. Davis, "Garage Cinema and the Future of Media Technology," *Comm. ACM* (50th anniversary edition), vol. 40, no. 2, 1997, pp. 42-48.
2. M. Davis and D. Levitt, *Time-Based Media Processing System* (US Patent 6,243,087), Interval Research, 2001.
3. C. Dorai and S. Venkatesh, "Computational Media Aesthetics: Finding Meaning Beautiful," *IEEE MultiMedia*, vol. 8, no. 4, Oct.–Dec. 2001, pp. 10-12.
4. M. Davis, *Media Streams: Representing Video for Retrieval and Repurposing*, doctoral dissertation, Massachusetts Inst. of Technology, 1995.
5. M. Davis, "Media Streams: An Iconic Visual Language for Video Representation," *Readings in Human–Computer Interaction: Toward the Year 2000*, R.M. Baecker et al., eds., 2nd ed., Morgan Kaufmann, 1995, pp. 854-866.
6. MPEG-7 Requirements Group, *Overview of the MPEG-7 Standard* (version 6.0), ISO/IEC JTC1/SC29/WG11 N4509, 2001.
7. C. Metz, *Film Language: A Semiotics of Cinema*, Univ. of Chicago Press, 1974.
8. U. Eco, "Articulations of the Cinematic Code," *Movies and Methods: An Anthology*, B. Nichols, ed., Univ. of California Press, 1976, pp. 590-607.
9. N. Burch, *Theory of Film Practice*, Princeton Univ. Press, 1969.
10. D. Bordwell and K. Thompson, *Film Art: An Introduction*, 6th ed., McGraw Hill, 2001.
11. G. Davenport, T.G. Aguierre-Smith, and N. Pincever, "Cinematic Primitives for Multimedia," *IEEE Computer Graphics and Applications*, vol. 11, no. 4, July 1991, pp. 67-74.
12. F. Nack, "The Future of Media Computing: From Ontology-Based Semiosis to Computational Intelligence," *Media Computing: Computational Media Aesthetics*, C. Dorai and S. Venkatesh, eds., Kluwer Academic, 2002, pp. 159-196.
13. P. Keegan, "Lego: Intellectual Property Is Not a Toy," *Business 2.0,* Oct. 2001, http://www.business2.com/articles/mag/0,1640,16981,00.html.
14. C. Levi-Strauss, *The Savage Mind*, Univ. of Chicago Press, 1966.
15. R. Barthes, *Image Music Text*, Hill and Wang, 1977.
16. F. de Sausure, *Course in General Linguistics,* McGraw-Hill, 1983.
17. L. Barroca, J. Hall, and P. Hall, "An Introduction and History of Software Architectures, Components, and Reuse," *Software Architectures: Advances and Applications*, L. Barroca, J. Hall, and P. Hall, eds., Springer-Verlag, 2000, pp. 1-11.
18. S.M. Eisenstein, *Film Form: Essays in Film Theory*, Harcourt Brace Jovanovich, 1949.
19. R. Barthes, "The Sequences of Actions," *Patterns of Literary Style*, J. Strelka, ed., State Univ. of Pennsylvania Press, 1971.
20. H. Jenkins, *Textual Poachers: Television Fans and Participatory Culture*, Routledge, 1992.
21. H.R. Varian, "The Law of Recombinant Growth," *The Industry Standard*, 6 Mar. 2000, http://www.thestandard.com/article/display/0,1151,11884,00.html.
22. M. Davis, "From Pirates to Patriots: Fair Use for Digital Media," *IEEE MultiMedia*, vol. 9, no. 4, Oct.–Dec, 2002, pp. 4-7.

Marc Davis is an assistant professor at the School of Information Management and Systems (SIMS) at the University of California at Berkeley where he directs Garage Cinema Research (http://garage.sims.berkeley.edu). Davis' work focuses on creating the technology and applications that will enable daily media consumers to become daily media producers. His research and teaching encompass the theory, design, and development of digital media systems for creating and using media metadata to automate media production and reuse.

Readers may contact Marc Davis at the School of Information Management and Systems, Univ. of California at Berkeley, 314 South Hall, Berkeley, CA 94720-4600; email marc@sims.berkeley.edu.