Mindful Documentary

by

Barbara A. Barry
B.F.A Massachusetts College of Art 1991
M.S. Massachusetts Institute of Technology, 2000

Submitted to the Program in Media Arts and Sciences
School of Architecture and Planning,

in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Media Arts and Sciences

at the Massachusetts Institute of Technology
June 2005

Signature of Author
Program in Media Arts and Sciences
May 6, 2005

Certified by
Glorianna Davenport
Principal Research Associate, Interactive Cinema Group
MIT Media Arts and Sciences
Thesis Supervisor

Accepted by
Andy Lippman
Chair, Departmental Committee on Graduate Studies
Program in Media Arts and Sciences

Mindful Documentary

by

Barbara A. Barry

The following people served as advisors and readers for this thesis:

Thesis Advisor
Glorianna Davenport
Principal Research Associate
Director, Media Fabrics Group
MIT Media Lab

Thesis Reader
Walter Bender
Executive Director, Media Laboratory
Senior Research Scientist
Director, Electronic Publishing Group
MIT Media Lab

Thesis Reader
Erik Mueller, Ph.D.
Research Staff Member
IBM Research

Mindful Documentary

by Barbara A. Barry

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning, Massachusetts Institute of Technology,
on May 6, 2005,
in partial fulfillment of the requirements for the degree of Doctor of Philosophy
in Media Arts and Sciences

Abstract

In the practice of documentary creation, a videographer performs an elaborate balancing act between observing the world, deciding what to record, and understanding the implications of the recorded material, all with respect to her primary goal of story construction. This thesis presents *mindful documentary*, a model of a videographer's cyclical process of thinking and constructing during a documentary production. The purpose of this model is to better support documentary creation through systems that assist the documentary videographer in discovering new methods of observation, ways of thinking, and novel stories while recording the world.

Based on the mindful documentary model, a reflective partnership is established between the videographer and a camera with commonsense reasoning abilities during capture and organization of documentary video collections. Knowledge is solicited from the videographer at the point of capture; it is used to generate narrative or contextual shot suggestions, which provide alternative recording path ideas for the videographer. Thus, the system encourages the videographer to reflect on the story possibilities of a documentary collection during real-time capture.

Qualitative results of studies with a group of videographers – including novices and experts – showed a willingness to take suggestions during documentary production and, in some cases, to alter the recording path after reflection on shot possibilities presented by the system. Moreover, suggestions often had increased influence on the recording path if they were not taken as directives but as catalysts, i.e., prompts to expand thinking about the documentary subject rather than explicit shot instructions.

Critical lessons were learned about methodology and system design for documentary production. As a documentary is built, evidence of what the videographer has learned is represented in the documentary. The model, methodology, and system presented in this thesis provide a basis for understanding how videographers think during documentary construction and how machines with commonsense reasoning resources can serve as creative storytelling partners.

Thesis Supervisor: Glorianna Davenport
Title: Principal Research Associate, Media Fabrics, Program in Media Arts and Science, MIT

Table of Contents

List of Figures

List of Charts

Acknowledgements

This thesis was a marathon for the mind. Along my run there were coaches, teams, companions, fans and volunteers who helped me get here – to the finish line. There will be other marathons in my future but there is a certain satisfaction in breaking the first tape, in completing a long journey of ideas in which I trained, stretched, ran, struggled up hills, then cruised carefully down. These are the people who accompanied me. They helped me to be focused, nourished and on the right path.

Glorianna Davenport, my advisor, has been my mentor during my years at the Media Lab. She has taught me how to see as documentary filmmaker and a think as a researcher. This work was a risk. She supported it without hesitation and was the great trainer who knew when to encourage me to press forward with the idea and when to give me space to reflect. She is a pioneer in the field of computational documentary; her filmmaking has been an inspiration and our conversations about stories have brought my thinking to new creative heights. Thank you, Glorianna.

To my readers, Erik Mueller and Walter Bender, I am grateful for their intellectual generosity. Erik Mueller's daydreaming computers showed me the possibilities for imaginative machines. His thoughtful precision advanced this work. Walter Bender, in the midst of running a lab, was never too busy to help me refine an idea or solve a puzzle. His work in community publishing brought a new perspective to my thinking about commonsense knowledge acquisition.

Nicholas Negroponte built the Media Lab as a haven for the researchers on the edges. It is amazing how well the edge dwellers fit together into a working, growing lab. I am convinced that this work flourished in part because of the everyday life at the Media Lab, where there is always an idea being created or debated at the coffee machine, in the elevator, in an office, or during a sponsor meeting. My colleagues at the lab shared ideas, struggles and victories. Thanks to Peter, Michelle, Kelly, Andrea, Erik, Stefan, Kimiko, Joanie, James, Ramesh, Nathan, Fernanda and Lily.

Special thanks to Marvin Minsky and Seymour Papert for sharing big ideas and encouraging students to be original thinkers. Brian Smith influenced this research with his insight and humor. His knowledge of symbolic artificial intelligence and of reflective practice in education was a touchstone for this work. Thanks to Ted Selker for tuning the evaluations of this work for practicality and maximum discovery and to Henry Lieberman for many discussions about interactive commonsense agents. Thanks to Mitchel Resnick, Pattie Maes, Bruce Blumberg and John Maeda for their well placed questions during the journey.

The Interactive Cinema/Media Fabrics Group at the Media Lab is group of colleagues who I would call immediately if I were ever faced with a brainstorming competition. It has been joyful spending this part of my life with you dreaming up the future of cinema. Thanks Aisling, Ali, Hyun, Pengkai, Paul, and Kevin.

The Commonsense Computing Group at the Media Lab was my second academic family. They have the courage to tackle tough problems and the ingenuity to solve them. This work would not have been possible without them. Thanks to Push, Ian, Hugo, Bo and Ryan. I am looking forward to seeing what the world looks like once the commonsense problem is solved!

Thanks to the UROPs who helped me hash out the ideas for this work in the early stages: Grace Lim, Austin Kim and Tina Hsieh.

I have benefited greatly from conversations and support from colleagues in multi-media and computational story community: Frank Nack, Marc Davis, Chitra Dorai, Flavia Sparacino, Carol

Strohecker, Louis Weitzman, and Marina Umaschi Bers. Thank you Jack Driscoll for keeping me connected to the world of news.

I would like to acknowledge the generous support of this work by Motorola Corporation, the Media Lab Europe and Warner Bros. Mike Taylor and Franco Lodato of Motorola encouraged this work in the early stages. Ken Haase of Media Lab Europe brought the commonsense work to MLE and participated in many helpful conversations about commonsense problem. Chuck Dages of Warner Bros. supported the later stages of this work and believed in the possibilities for common sense for interactive narrative.

Linda Peterson and Pat Solakoff were extraordinary administrators who made sure all the mazes of MIT paperwork on time and with ease. Tracy Daniels, the administrative assistant for the Media Fabrics group, took care of all the details with skilled calm.

Stacie Slotnick and Laura Packer were sensitive editors who helped me find the right words, put them in the right order, and remember to use a serial comma.

Tara Rosenberger has been a work ally and a great friend who forged the way, a year ahead, placing road signs that helped me embrace the straight roads and avoid the bumps. Thank you, dear friend!

To my friends who supported and believed in me even when I was out of sight, on one of the lonely stretches of road: Linda & Jim; Fee; Kamille, Mitch & Elena, Katherine and Cydney. Thanks to Julie for cheering me on during the hill climbing. Thanks to the Minsky's, Fredkin's and Singh's for their open doors and kindness.

Thanks to my family for patience, love and encouragement: Mom, Tom, Michele, Michael, Claire, George, Aunt Barb, Uncle Tom, Sandy, Mike, Patrick, Carlton, Catherine.

Push, you have the perfect balance of mind and heart. I am so happy to have accomplished this work side-by-side. Success shared in love is infinitely sweeter.

This work is dedicated to my nephew Michael, who always calls me "aunt-barb-with-the-magic-camera."

1. Introduction

In 1895 the Lumière Brothers made one of their first films, entitled *Arrival of a Train.* It was comprised of a single, fifty-second shot of a train arriving at a station. Yet even this simple film reflects many decisions: to begin recording with the train in the distance with a worker walking across the platform toward the camera, to place the camera at the edge of the platform nearest to the path of the train, to include the passengers watching the train as it approaches then moving toward the train as it comes to a stop, to develop the film and show it to an audience who, as folklore has it, ran from the theater believing the image cast in light was a real locomotive instead of a phantasmagorical train arriving at a mirage of a station. The audience was presented with a situation and created an inference that caused them, at least, to wince (Barnouw, 1993, pp. 7-8). They could not distinguish the filmed image from reality and perhaps inferred the oncoming train might hit them. A train is large, heavy and travels quickly. It is not dangerous to the passengers riding it, but certainly to someone in its path. When we watch moving images in the form of film or video we each create inferences that forge our understanding of them. Some of these inferences are personal and individual, others commonsense products of knowledge we all acquire from living in a shared world. Since the time of the Lumières, viewers have learned to read film language and apply their inferences to the film presentation. They stay in their seats but their minds continue to generate, parse, and discard inferences as they construct their understanding of the documentary. The task of understanding becomes even more complex as documentaries grow from a single short shot, to longer linearly edited films, and contemporarily to digital, non-linear computational documentary. Still videographers observe, infer, decide, and record life resulting in collections of video that set our minds to the task of understanding.

The fundamental role of the documentary videographer is as a fine-tuned observer, inference generator, and designer; and, one who recasts real-world events into new patterns of understanding. A documentary videographer can present us with a way to make inferences about an image or story that we have never before experienced. The process of documentary creation is always subjective; this is its strength. The videographer brings knowledge about the subject and a point of view to the act of documentary construction. Creating a documentary is not merely crafting an appearance of the *real* world but also a learning venture, one that reveals new patterns of discovery, novel stories and new ways of understanding the *lived* world. As the documentary is built, what the videographer learns is reflected in the documentary.[1] The construction of a documentary demonstrates "praxis," individual and social change through a cyclical process of thought and action.[2] The videographer ultimately creates a story that can be used as a catalyst for discussion of and reflect on the documentary subject.

1.1. Purpose of the Investigation

The purpose of the investigation is to better understand the process of documentary creation and how to develop technologies to assist the documentary videographer during the capture phase of videography, with a view toward constructing interesting stories. This thesis presents both a model of how the videographer thinks during documentary construction called *mindful documentary* and the system I developed called the *mindful camera* system. The mindful camera system has commonsense reasoning abilities designed to help the videographer reflect on story possibilities during the recording of a video collection. The system was tested in the field by expert and novice videographers and useful lessons were learned about modeling documentary practice and the use of commonsense reasoning machines as assistants during the creative process of storytelling.

1.2. Problem

Video documentary is a practice with many styles and motivations, both amateur and professional. A video recording device might be used intermittently to record events as notes or points of interest that exist independently, with no motive for constructing a larger context or story. Alternatively, life events might be captured as indices into the lived moment, as pointers to our memory of an event. The resulting video collection may be considered as snapshots of important life events: not stories in themselves, but catalysts around which we tell a story, scrapbooks of life. With less frequency, the motivation of the videographer turns poetic. The medium is used to produce an impressionistic look at the world, often using an aesthetic sensibility and metaphorical relationships between images to present a new view on an everyday object or concept. Finally, and the concern in this thesis, is the investigation of a documentary subject though narrative construction. The videographer captures a video collection of a documentary subject with the goal of presenting a story, an account of what happened. Narrative storytelling is one of the most pervasive modes of documentary construction (Nichols, 2001) and creating such a documentary is a difficult, creative task. Often, as viewers, when we watch a documentary we are only aware of the final product, not the intricacy of thought employed to create the documentary, the decisions and actions that enabled the successful end product. To understand the construction process such that we can create technologies to support it, we must explore how the videographer thinks during construction.

1.2.1 Problems Videographers Face

The capture and editing phases of documentary construction are intricately connected activities. Richard Leacock remarks, "I insist on editing my own films. That's half the fun. You are shooting when you are editing and editing when you are shooting."(Leacock, 2005) However, capture and editing are wrongly considered by many to be separate activities. First, the videographer is constantly generating construction possibilities during observation and deliberating about what to record. This valuable information ultimately helps a videographer to successfully edit a story from a collection of video material: "It is important to understand the relationship between editing and filming. The editor is not going to be able to construct a film successfully from the footage shot unless adequate footage is returned from shooting." (Pinkus, 1972, p.64) Second, there is an assumption often made by videographers that just because the camera is pointed at the world and recording that an account of what happened is realized. The feeling of simultaneously capturing everything and nothing about a documentary subject or event often vexes amateur videographers. Construction decisions are actively made during capture and consequently the videographer is not creating an inference guide that can serve as a way to see the content collection, and worse, the videographer may not be actively learning or discovering a new way to think about the documentary subject.

Documentary videographers also face the problem of incomplete and fragmented collections, without realizing the shortcomings of a collection during the window of opportunity for capture. Sometimes a videographer might take a risk and follow a path of action that does not unfold the way she expects. Alternatively, she might get too attached to one storyline at the expense of other supporting context. As a result, she is left with partial documentaries – collections of content that have beginnings with no endings, details with no context, and actions without reactions. The collection is missing vital building blocks of a story that describe what happened. Understanding the *narrative possibility* of a content collection is knowing what story elements and relationships between those elements create a complete, understandable story. The narrative possibility of a content collection and how it accords with the documentary story grows in complexity as we become able to document entire lifetimes.

The videographer faces a challenge: to choose a recording path that will best support story construction for a single story or multiple stories from a documentary collection. This requires a delicate balancing act in which the videographer must attend to the potential for story in their collection and in the world they are recording.

In summary, the problems documentary videographers face during construction addressed in this thesis are:

- Finding a recording path that will best support story construction
    - *"What shot should I take next?"*
- Reflection on narrative possibility during capture
    - *"Do I have the shots needed to tell a story?"*
- Capturing an inference guide through the development of a content collection
    - *"What would someone think if they saw this clip?"*

1.2.2. Designing Technologies to Support Documentary Videography

Current tools to support documentary construction are limited by their inability to understand the videographer's main goal – story creation. In this thesis the concern is that the videographer be generating story possibilities in mind that will aid in video capture. The resulting video collection will not be edited but will contain elements needed for story creation due to the videographer collecting video clips while guided by story suggestions. There are few tools that support documentary videographer reflection on content collections during capture, and none that encourage reflection on story. The problem is threefold. How can the system interact with the videographer as a storytelling partner? How can the system understand and generate narrative possibility both in the world and in a collection of content? How can video material be annotated to enable story suggestions to be generated for the videographer?

In order to support the videographer we must design machines that understand the videographer's creative process. By carefully studying the process of creation, instead of solely the language of film, we can build systems that can intervene, interact, and encourage videographers to consider story possibilities during capture. Historically creative needs have driven technology innovation, which have allowed new forms of documentary to emerge. In the case of this thesis, the creative need to see multiple story possibilities in real time during capture – particularly as the ability to capture documentary collections that can span entire lifetimes – drives the development of a technology that can think with the videographer about the many stories that exist in her video collection.

A camera with story understanding and generation must be able to reason about complex relationships between events, objects, states, and people – all the building blocks of stories. Much progress has been made in artificial intelligence to read text-based stories and answer questions about them; most often the systems are designed to handle a particular set of stories or stories that come from a very well-defined domain. A system to support documentary videography must have a broader ability to reason across many domains and shift between them efficiently. People employ a vast amount of knowledge and methods simultaneously while storytelling. Knowledge of past events informs the current experience of storytelling or listening. Storytellers know they can omit certain details of a story, trusting the audience will reach a plausible conclusion if given the right inference clues. They know that if they introduce a train in a story a listener might generate knowledge such as "speed," "travel," "industrial revolution," or maybe a specific story of a vacation in which they took a train. The listener is reconstructing the story based on their own knowledge and the storyteller must be aware of this during construction and guide the listener carefully. If the storyteller is not judicial about the information in the story, its elements and relationships, the listener might not be able to make sense of the story. By living in a shared world, people acquire commonsense knowledge they use to construct and understand stories. Commonsense knowledge is a large collection of facts that the mind knows based on experience and the ability to reason with this knowledge imperfectly yet successfully in a world where we need to constantly accommodate the new and the unexpected. To imbue a computer with the mental capacity of common sense is a very difficult problem in artificial intelligence. Marvin Minsky writes, "If we want our computers to understand us we'll need to equip them with adequate knowledge. Only then can they become truly concerned with our human affairs."(Minsky, 2000, p. 68) For a camera to be smart enough to aid the videographer in story construction it needs commonsense reasoning abilities, the ability to reason about the everyday human life that the camera is recording. This requires using known resources of commonsense knowledge and developing new resources to aid in reasoning about real world situations and stories.

In summary, creating technologies to support documentary videography requires:

- A cognitive model of how documentary videographers think during documentary creation;
- Commonsense reasoning in-camera to provide flexible and creative story generation during the capture process;
- A practice of collecting video content that encourages reflection on story construction.

1.2.3 Prior Work Overview

In this section we look at two research areas that impact documentary videography. The first is knowledge used to guide video capture. The camera or system has a set of computational representations that provide capture parameters that control an autonomous camera. The second research area is story-driven navigation and automatic assembly of already-captured video clips from a video database. Canonical examples of prior work in these areas are described in this section. A more detailed overview of related and prior work can be found in Section Three.

1.2.3.a Guiding Capture

Kodak first developed the Brownie camera in 1900. The Brownie revolutionized photography by mass production of affordable cameras for the general public. Special and everyday events of life could be recorded cheaply and conveniently. The idea of the snapshot was born. Kodak innovated not only in their engineering but also in their educational materials. Instruction manuals provided lists of what an amateur photographer might want to take a snapshot of in a variety of life events, such as birthday parties and graduation ceremonies. The lists were prototypical. The details were left to how the event unfolded in detail as the photographer observed it. Unfortunately, Kodak did not produce an analogue instruction manual for their moving-image cameras.

Research in guided capture using video cameras has yielded work on automated recording of tightly scripted, limited-domain documentaries. Pinhanez and Bobick (1996) built a smart camera that can implement the instructions of a human director during the recording of a cooking show. This work is designed to enable framing of shots when the content of the documentary is pre-determined. The director instructs "close-up of chef" and the system has vision routines that will allow for the recognition of the object "chef" in the scene and proceed to take a close-up of the chef. The system can also decompose actions by accessing more detailed knowledge about a shot. For example, an action of mixing involves hands, a bowl and a spoon. This smart camera can track actions over the course of a cooking show as they accord to a set of predetermined scripts, sets of actions represented in a variation of the script representation of Schank (Schank, 1997). The first lesson from this work is that tracking actions over the course of a documentary is possible using a set of scripts. The second lesson is that story understanding by computers or cameras for limited domains is brittle. If any unexpected event occurs the system will ignore it, miss it or, in the worst case, grind to a halt due to the incompatibility of the unexpected event. The mindful documentary work differs because it uses broader commonsense reasoning and a fail-soft approach to enable tolerance to the unexpected. In addition there is a commitment in this work to supporting decisions about what to capture as a priority over how to frame a particular video clip.

Schroeder implemented a system called Ingmar as an automatic movie director (Schroeder, 1987). Ingmar used a set of story scripts to retrieve and order video clips from a collection as a director or editor might. The scripts were close in representation to Schank's scripts but altered to include information about how to frame a shot. Ingmar was designed to generate a sequence of clips about a dinner party. Before collecting the video necessary for the system, the computational scripts were created to represent a dinner party. The video footage of a dinner party was recorded. The scripts' elements were dynamically presented to the cinematographer or videographer during capture. The author of the system then looked at the captured footage, broke down the shots, and annotated the constraints between shots. For example, if a shot had a plate full of spaghetti, a constraint for automatic generation of the next clip would be that the server could not be filling a plate. The plate was already full. In the case of the mindful camera the documentaries are observational; the videographer is portraying a story that is unfolding before her eyes. The scripts cannot be predetermined nor can the videography view all the footage and accord it with story structures in order to assemble the story. The system must dynamically understand the story possibilities during real-time capture.

1.2.3.b Computational Documentary

Systems for documentary production and presentation have included hypermedia systems (Lippman, 1980), random-access techniques (Davenport, 1987), editors-in-software (Davis, 1995; Murtaugh, 1996), story-generation systems to guide video sequencing (Bloch 1988; Mateas, 2000), and case-based reasoning systems for video retrieval (Schank, 1998; Burke & Kass, 1995). These systems have greatly contributed to the understanding of the challenges of video representation, methods for sequencing video clips from closed content collections, and the understanding of how to preserve continuity of user experience in different ways, from continuity using a single variable, such as time, to continuity using multiple story variables, such as events, locations, and character. A few of these systems have laid the groundwork for automatic or interactive documentary creation. Story continuity between clips was implemented in random-access and editor-in-software systems by retrieving clips that had similar keywords as previously viewed or retrieved clips. This enabled character and location continuity. If a user viewed one clip from a collection described with the location keywords "Boston," the algorithm for next-clip selection would retrieve clips described with the same keywords using a "next-in-list" approach (Davenport, 1987). More complex algorithms weighted multiple keywords to influence retrieval of similar clips using spreading activation (Murtaugh, 1996; Maes, 1990) the clips with the highest score for keyword similarity would be presented to the viewer. All these systems provided continuity of viewing by providing "sensible" transitions between clips. Other systems used story structures to organize clip retrieval and viewing. Chua & Ha matched keywords with script elements for retrieval and sequencing of video clips (Chua & Ruan, 2000). These systems focus on the content of the video clips rather than the rhetorical form of stories. In an alternate approach, Brooks used structural narrative primitives combined with specifications of relationships between video clips to drive recombination of elements from a video database. (Brooks, 1999) A video clip could be marked as having a causal relationship with another clip in the database and be described as filling a particular narrative role, such as the introduction of a character.

While computational documentary systems have provided a basis for understanding how to create stories using information retrieval from existing video databases they ignore computation possibilities during the capture phase. The mindful documentary work is a necessary counterpart to the work in retrieval by guiding capture so as to enable the collection of video clip collections that contain sets of story elements.

1.3 Mindful Documentary

Solving the narrative puzzle of documentary creation involves recording events from the real world and understanding how their shape allows composition with other existing shots a given collection. It is a matter of building the whole from the parts. During video capture, the filmmaker uses observations of each recorded moment to establish mental models of possible stories that might be constructed. As recording occurs, the videographer adjusts the models in accord with the real world. Mental models continue to be recalled and revised as capture continues; story patterns begin to develop, as knowledge about the composability or relations between clips are aggregated and organized. The result of this mental process can be thought of as story models, possible stories observed over the course of documentary construction. Of course, not all the stories in mind will be able to be told from the collection gathered. The challenge is to understand what stories can be told and develop them while there is still an opportunity to collect video pieces. Trying to understand the ways a video collection can express a story involves a type of puzzle solving; as in all puzzle solving, sometimes we get stuck because we have an incorrect image of the target solution; perhaps we are making an incorrect analogy to a different puzzle we have previously solved successfully; yet we cling to our decision when evidence tells us to move on. Sometimes we overcome these impasses by exhausting all the incorrect solutions; other times by using scaffolding created by someone who has found a solution; sometimes we merely need a break to free our minds from current views of the problem. Hopefully, we persevere. The lesson is this: in order to reconfigure the world as a documentary we must solve a narrative puzzle which requires reflection; sometimes, in order to solve the puzzle we need to *stop* and *think*.

Mindful documentary is a model of a videographer's cyclical process of thinking and constructing during a documentary production. I developed the model as a way of better understanding the process of documentary creation in order to build a camera that could participate in the story construction with the videographer during real-time capture.

It is a cognitive model for documentary creation that expresses the complexity of thought that happens during documentary capture. During documentary construction, the videographer is observing the real world, deciding what to record, recording, and then reflecting on the implications of the recorded material. The videographer is constantly moving between attending to the real world, attending to her own thinking about the possibilities for capture and attending to the possibilities for story expression in her captured collection.

1.3.1 The Construction Cycle

The main component of the mindful documentary model is called a *construction cycle* (Figure 1). The construction cycle consists of a *shot action,* a *reflection window,* and a *shot decision.* The shot action is the act of starting and stopping recording which yields a single video clip. The reflection window is the thinking the videographer does about the implications of the recorded shot to inform the next shot decision. The reflection window will be discussed in more detail in a moment. The shot decision is the goal for a particular future video shot. This cycle is repeated as shots are taken and as a result the videographer incrementally builds knowledge about the documentary subject and about the inferences that enable narrative construction from shots in her collection. The total of all construction cycles over the course of a documentary constitutes the videographer's recording path.

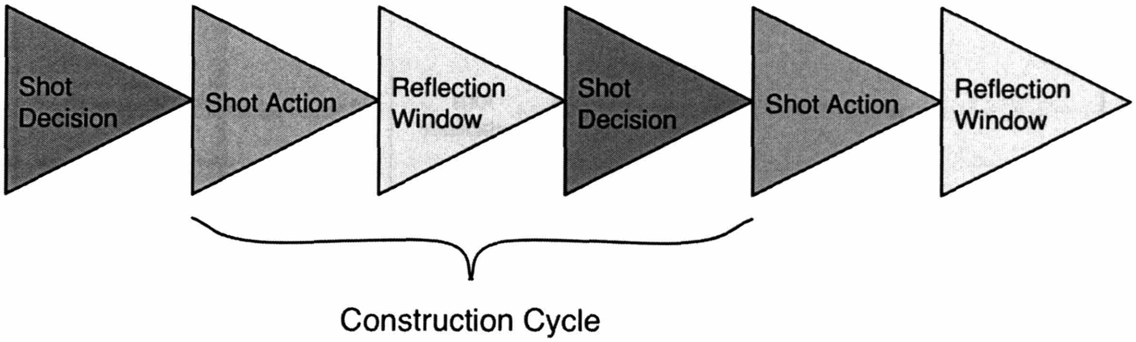


Figure 1: The mindful documentary model of documentary video construction

The mindful documentary model inherits ideas from three existing models of how people think: Donald Schön's ideas about reflection during learning; John Dewey's general model of how people think; and Marvin Minsky's six-level model of mental activity (Schön 1983; Dewey 1933; Minsky 2005). The construction cycle is a version of Schön's decide-act-reflect cycle specified for the documentarian. Schön's important idea is that during an activity we have the opportunity to stop and think, to reassess our progress – in the videographer's case on solving the narrative puzzle, and making changes that will positively impact our ability to solve the problem. The construction cycle also inherits from John Dewey's five-step model of reflection. The steps of his model are: detect, define, suggest, implicate, and decide. In the mindful documentary model each shot action creates a set of constraints and possibilities for story creation; this is the detected problem, what Dewey would call a "felt difficulty," which sets off a chain of five reflective reasoning steps. In the mindful documentary model these five steps occur in the reflection window and are specific to the goal of story construction. The concentrated thinking about documentary construction happens during the reflection window. Marvin Minksy's theories about levels of mental activities in people and computers contribute the idea of reflecting on our predictions or imaginings about what might happen next in the world.

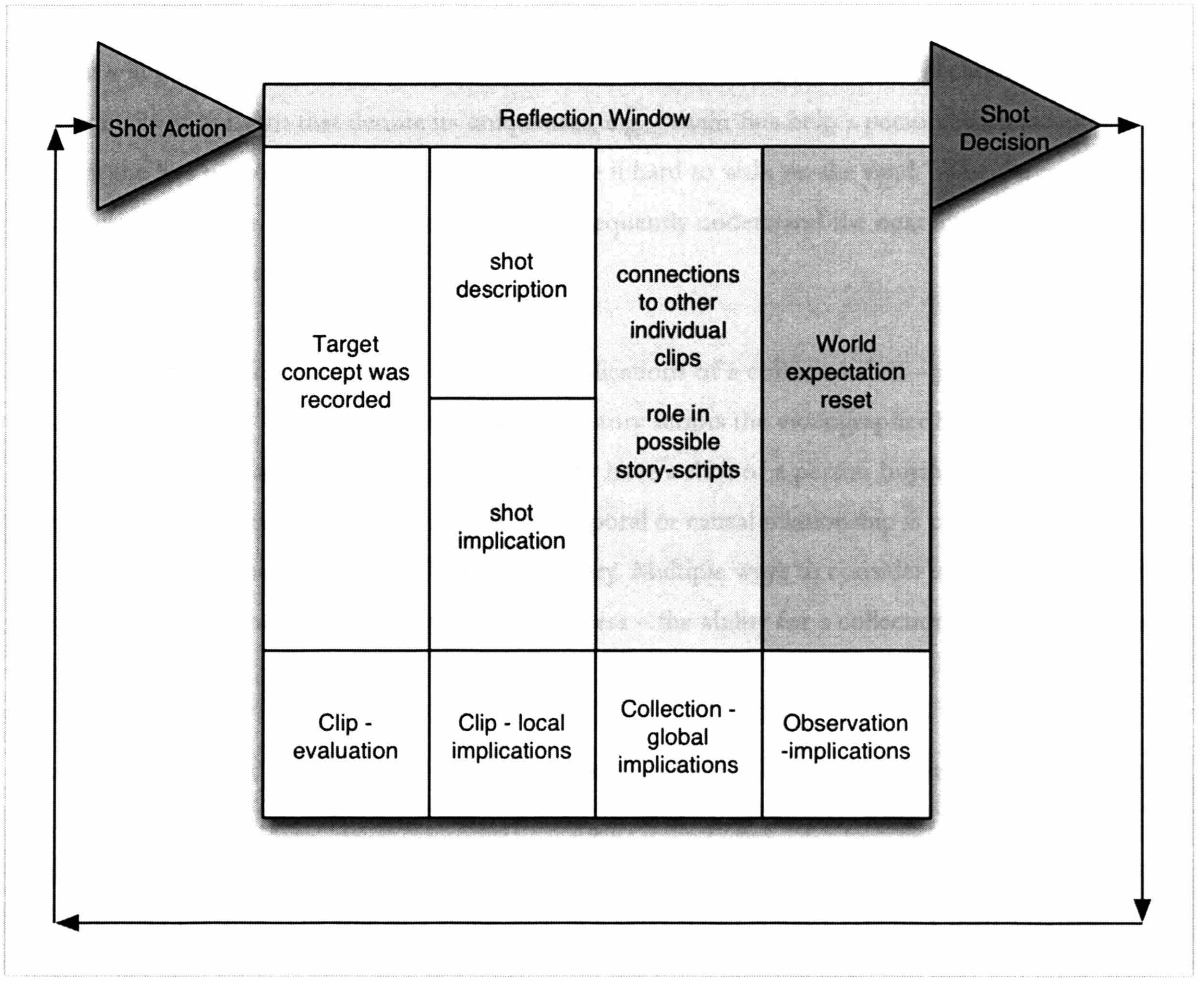

Shot Action
Reflection Window
Shot Decision
Target concept was recorded
shot description
shot implication
connections to other individual clips
role in possible story-scripts
World expectation reset
Clip - evaluation
Clip - local implications
Collection - global implications
Observation -implications

Figure 2: Detailed model of the reflection window

A detailed model of the reflection window is shown in Figure 2. During the moment of reflection after the capture of each shot the videographer: 1) evaluates the clip; 2) generates a description and local implications of the clip; 3) generates and evaluates the global implications of a clip to the collection; and 4) sets an expectation of what might be observed next in the world. After recording a shot the videographer generates in mind the implications – the possibilities, constraints, and impossibilities for narrative construction.

After a shot has been captured the videographer evaluates the clip by deciding whether the target concept was recorded successfully. This can be expressed simply: the shot was captured or it was not. In the case of failure, the videographer would immediately make another attempt or skip the shot.

If the shot is captured, the videographer has a description in mind of what the shot depicts. In a documentary about a day at the beach, a videographer might capture a shot such as "a person is putting on swim fins." The immediate shot implications are a simple, one-step inference such as "the person will go in the water" or "the person will swim." In addition, the shot implications might also mark details of the shot that denote its uniqueness, e.g., "swim fins help a person swim faster" or present the less obvious inference "swim fins make it hard to walk on the sand." The description of a clip drives our ability to create inferences and subsequently understand the next immediate or global implications of a video clip.

The videographer also thinks about the global implications of a collected shot – possible connections to other single shots in the collection or expected story scripts the videographer has in mind for the documentary; for example, the videographer might have a shot of a person buying swim fins or stealing swim fins from another beachgoer. A temporal or causal relationship is created between pairs of clips. These are the first building blocks of a story. Multiple ways to consider each video clip's role in a collection increase the possibility of story success – the ability for a collection to produce at least one coherent story.

We all have different experiences of going to the beach and each has both unique characteristics and a canonical events. These experiences can be thought of as *story-scripts*, mental models of our experiences. Each videographer brings her experiences, a story-script collection, to the act of documentary creation. The role of a single shot in a story-script is more complex than a binary relation between clips. A clip's role in a story-script has more dependencies and may not make sense if another clip is discarded or added to the story. In a story about someone going on a trip to the beach there are multiple ways to begin: showing a person driving up to a parking lot, showing a person leaving home with beach gear to embark on a journey, or showing a person already laid out on a beach towel looking out at the water. If a person is attacked by a shark while swimming, there are many possible outcomes to the story, which can be viewed as next steps in unique story-scripts.

The swimmer might be rescued by a lifeguard, swim to the safety of a dinghy or be eaten by the shark. Videographers hold story-scripts in mind as possibilities about what can reasonably happen in the world they are recording. The production of story-scripts strongly influences the recording path the videographer decides to take. Later in this document we will look at representing story-scripts to a computer, to give it the capacity to reason using stories.

During the last phase of the reflection window, the videographer resets her world expectation. Ideally, the next plausible shot in the real world that occurs and best completes one or more story-scripts is the target next shot, the one the videographer will decide to shoot. The shot decision is the result of deliberation by the videographer involving the opportunity in the world and the constraints of her collection so far. Obviously, not all shots captured are useful and not all shots are successfully captured. The videographer can always abandon an idea if it does not accord well with opportunity in the real world or if, creatively, it looks like a part of the recording path that would be better left unexplored.

1.3.2 Inference Guides

The fundamental goal of the videographer is to design narrative *inference guides*. This work is done intensively during the reflection window, after each shot has been taken. Construction decisions guide how the audience understands the motivations and actions of social actors, dynamics of events, and recognition of themes in the documentary. Not everything in a documentary is stated explicitly. The audience uses world knowledge to "fill in the blanks."[3] It is the work of the videographer to predict the inferences and conclusions the audience will likely draw at each moment of the documentary.

Historically, observational documentary utilizes temporal and causal inference, with each event building to a "defining moment" in which the goals of the characters are realized or thwarted. *Primary*, by Robert Drew and Drew Associates, follows the Wisconsin campaign and vote in 1960 presidential primary between Hubert Humphrey and John F. Kennedy (Drew, 1960).[4] Since observational documentary relies heavily on narrative structure to deliver to the audience a version of "what took place in the presence of the camera when the camera was turned on," it would be inconsistent to violate or simply ignore expectations of event order (Davenport, 1980).[5] In "Primary," the voting action follows campaigning action. The announcement of a winner follows the counting of votes. Observational documentary also relies on delivering details to the audience to reveal something new or unexpected about the documentary subject (e.g. close-up of Jacqueline

Kennedy's white gloved and wringing hands behind her back shows her stress while greeting a crowd, even though her face shows composure).

The problem of documentary creation is ultimately one of content collection for story creation. In this thesis a model of how the videographer thinks during a documentary is presented as the guide for designing a partnership between a videographer and a camera with commonsense reasoning. The mindful documentary model provides a way to understand how videographers generate and use the building blocks of story to solve the puzzle of creating documentary stories from their observations.

1.4 Thesis Statement

A reflective partnership between a videographer and a camera with commonsense reasoning abilities can aid story construction during documentary videography by providing useful suggestions to help the videographer identify narrative possibility during the documentary creation.

There are a series of claims that can be generated from this hypothesis. The claims are addressed in the implementation, testing, and evaluation of the mindful camera system.

- Documentary videographers are open to suggestions during the reflection window of the construction cycle.
    - Reflection on content collection during capture can aid future capture decisions and influence the videographer's recording path.

- Commonsense knowledge can be used for story suggestion.
    - A camera with commonsense reasoning can help the videographer understand narrative potential during real time capture;
    - A camera with commonsense reasoning can tolerate domain shifts;
    - Natural language annotation submitted in the reflection window allows the machine to generate commonsense inferences.

- Commonsense reasoning can be used to generate the narrative potential to guide video collection.

1.5 Approach

This thesis introduces a novel approach to documentary video capture called mindful documentary in which a partnership between the videographer and the camera supports the story construction goals of the videographer. The term "mindful" imparts a dual meaning: the camera is mindful because it uses the annotation of video content to identify possible next events and context – it is mindful of the development of a content collection; and the human is mindful because suggestions by the camera help the videographer attend to the construction process while observing, recording and participating in everyday life.

The partnership is established by creating a methodology during documentary creation that encourages the videographer to reflect after each shot by submitting a text annotation and considering commonsense shot suggestions generated by the mindful camera. The mindful camera intervenes during the reflection window and offers alternative narrative inferences, based on a video clip's annotation that can be taken as shot suggestions by the videographer (Figure 3). The yellow blocks in Figure 3 show the places of intervention and possible influence on the videographer's thinking about narrative possibility during the reflection window.

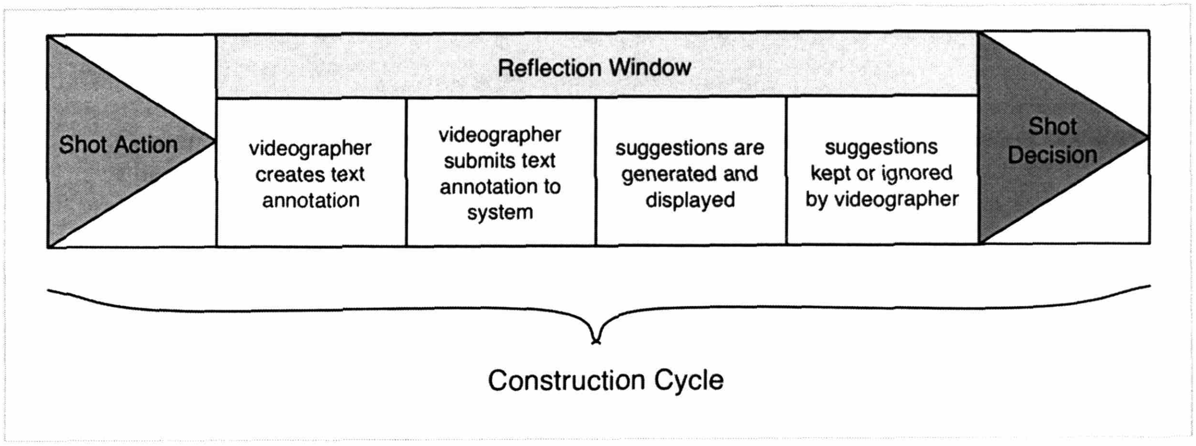

Reflection Window
Shot Action
videographer creates text annotation
videographer submits text annotation to system
suggestions are generated and displayed
suggestions kept or ignored by videographer
Shot Decision
Construction Cycle

Figure 3: Diagram of how the mindful camera interacts with the videographer during the reflection window of the construction cycle

The camera has the following abilities:

- Solicits videographer annotation just after the moment of video capture;
- Expands annotation of each acquired clip by generating related commonsense knowledge;
- Generates commonsense suggestions based on inference about next possible events and event context;
- Suggests the content of future shots to record;
- Logs videographer capture decisions.

Using observations by the videographer input as annotations, the mindful camera is able to generate future story-element possibilities; predict what might happen next; and understand implications of a recorded video clip. The abilities of the camera provide a means for the videographer to reflect on the work-in-progress.

Mindful documentary utilizes existing commonsense resources to generate narrative shot suggestions[6]. There are three resources used by the mindful camera: ConceptNet, LifeNet and StoryNet. These resources can be thought of as systems for reasoning about everyday life. They consist of large collections of commonsense assertions, collected by a diverse population of non-expert web users, and mechanisms for reasoning (Singh, 2002). ConceptNet, LifeNet, and StoryNet take unconventional approaches to acquiring, representing, and reasoning with large quantities of commonsense knowledge, expressed in English and contributed by non-expert web users. Each adopts a different approach: ConceptNet (Figure 4) is a large-scale semantic network, LifeNet (Figure 5) is a first-person probabilistic graphical model, and StoryNet (Figure 6) is a database of story-scripts. Here are diagrams and a simple reasoning example from each commonsense resource:

- ConceptNet expresses a broad collection of related concepts. If the documentary occurred at a beach the system could reason that a seagull might be present.

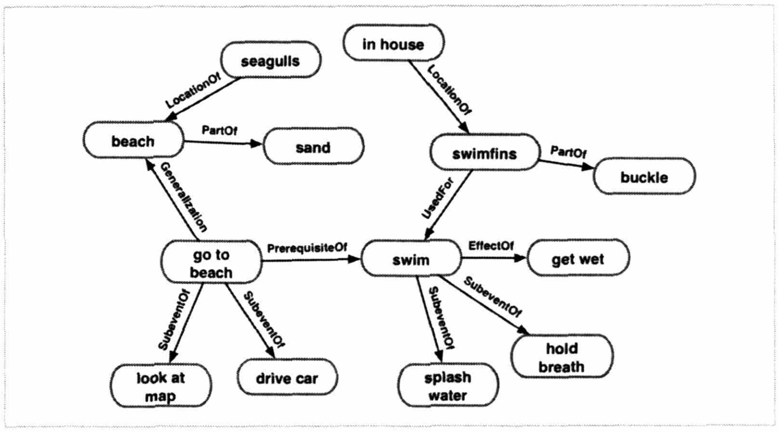


Figure 4: Simple subset of ConceptNet

- LifeNet represents time slices and probability that given the truth of an event (aqua filled nodes) another event is likely to occur or not. If a person arrived at a beach a likely next event would be they would swim in the ocean.

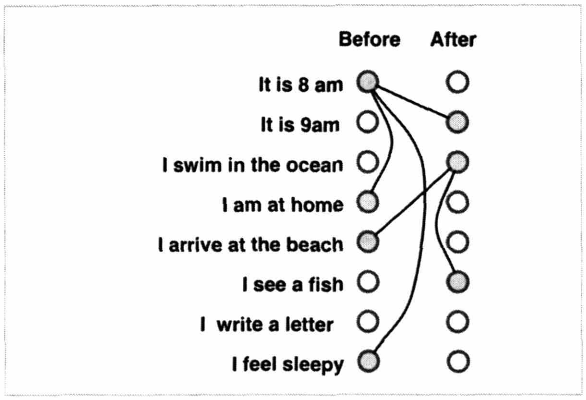


Figure 5: Simple subset of LifeNet

- StoryNet story-scripts represent the elements of an episode ordered temporally. During one instance of a trip to the beach a person swims in the surf and gets bitten in the leg by a shark.

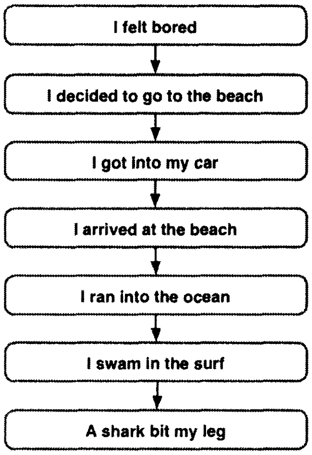


Figure 6: Example of one StoryNet story-script

These three systems can be considered distinct "ways of thinking" which can be used during story construction.[7] ConceptNet can help the videographer goal of capturing and expanding context for a recorded clip by providing info for what might be true or very likely at the current or following time step in a story. StoryNet helps identify more detailed story elements and their place in a sequence or story episode. In documentary videography, the goal of commonsense inference is not full story understanding by a computer but a step toward it by generating the temporal and causal event relationships that will help the videographer create the story of what happened. Since suggestions may not be correct or immediately applicable to the documentary subject each system offers suggestions to the videographer in a "fail-soft" manner.[8] Useful suggestions can be accepted and unwanted ones can be rejected. The videographer's patterns of taking suggestions may be evidence of a particular documentary style. The observational documentary maker might rely more on LifeNet for temporal event progressions while a poetic documentary maker might rely on the associative nature of ConceptNet. Detailed information about the three commonsense systems and the reasoning they support is provided in Chapter Four of this thesis and in Singh, Barry & Liu, 2004. Lieberman et al. provide an overview of applications that use ConceptNet (Lieberman et al., 2004).

The methodology of mindful documentary proposes that reasoning about story during the capture can support the story construction goals of the videographer. An artifact of this method is a shift in the production process, reuniting capture and editing phases of documentary videography. It requires that the videographer engage in annotation tasks, reflection, and decision-making about alternative recording paths while in the field. In mindful documentary the endgame for the videographer is to create a collection of video material that is richly represented.

1.6 Criteria for Success

The criterion for success of this investigation is to confirm that a system can be a synergistic partner to a human videographer during the documentary construction process. Evidence will be collected regarding how the system can help videographers reflect during the creation of documentary shoots and ideally help them find recording paths that support story construction. The success of the system depends on the value of the commonsense suggestions to the videographer. In addition, the evaluations should provide detailed information that will inform future design of creative commonsense assistants.

1.7 Results Preview

Studies showed that videographers were willing to take suggestions from a camera during the reflection window. Further, videographers reported that the suggestions helped them to think differently about the narrative possibilities resulting changes in recording path. Videographers took suggestions, thought about them, and acted by taking shots that were catalyzed by commonsense suggestions.

1.8 Points of Novelty

This thesis presents three novel contributions to the fields of computational documentary: the first cognitive model of how the videographer thinks during documentary construction; the first video camera that provides shot suggestions to the videographer; and the first video camera with commonsense reasoning abilities.

1.9 Roadmap

We have looked at the main goals for the videographer during documentary creation, the problems they face, and a model and approach to encouraging narrative reflection during the capture of a documentary collection. Chapter Two provides a more detailed look at theory and rationale, including its grounding and relationship to related work in the fields of computational documentary, commonsense computing and technologies for reflective practice. Chapter Three describes the implementation of the mindful camera system and its commonsense reasoning resources. In Chapter Four reports on the use of the system by a group of novice and expert videographers, showing detailed examples of the influence of the system on documentary practice in order to test the claims put forth in the thesis statement. Chapter Five presents recommendations for future work. Chapter Six is a conclusion. A glossary of terms is provided after Chapter Six to help the reader with the new (italicized) and referenced terms presented in this thesis. References and Appendix follow. The Appendix is a still-image catalog of each documentary created during the evaluation of this work, with a representative still from each video clip. In addition, the video collections can be viewed at http://mf.media.mit.edu/mindfulcamera/

2 Theory and Rationale

This chapter surveys related work in the theory of the nature and function of storytelling; story understanding by computers; commonsense reasoning; computational documentary; reflective practice; and reflective tutoring systems.

2.1 Story

A story is a type of representation we use to organize our lives and make sense of the world. Webster's Dictionary defines story as "the telling of a happening or connected series of happenings, whether true or fictitious; an account; a narration." (Webster, 1983) Within the scope of this thesis *story* is a term used to mean an account of an incident or happening made by selecting, adapting, and composing observations for presentation to oneself or others. *Story construction* is the act of creating a story by producing and making decisions about story elements and their relations. A story's composition is the shape or pattern within which events or series of events can be organized and understood (Livo & Rietz, 1986). Folklorists, philosophers, cognitive scientists, and computer scientists agree that stories are complex patterns that can be used for remembering, organizing, reconstructing, and learning from the vast amount of information that we process every day (Bartlett, 1932; Bruner, 1991; Minsky, 1974; Schank, 1977; Kearney, 2001). When we retell a story to another person, we rarely tell the story the same way twice (Bartlett, 1932). We have a model in mind that is used to guide the reconstruction for delivery to an audience, be it our friend in conversation, an auditorium full of people or even ourselves. The final model we choose for our telling provides a guideline for audience inference that successfully prompts the audience to infer what it is not explicitly stated in the story. Inference design in story construction is a delicate process because a story is not just a collection of elements. It is a collection of elements that have relations, conditions, and consequences. We engage the commonsense knowledge that is needed to understand the implications of each choice in story construction. Therefore, stories are 'implicit contexts' for knowledge, which have many of the advantages of explicit contexts. A good story about a day at the beach, for example, relates knowledge about the effects of an action (chasing a bird will cause it to fly away), problems such knowledge helps you solve (food must be protected from competing animals), situations in which such knowledge may be relevant (a picnic at the beach), and so forth. Stories rely on the shared commonsense knowledge of the maker and the listener.

The cognitive revolution declared that we could represent and share our perceptions of the world with others using stories. The process of creative storytelling used in documentary construction is similar to other realizations of story. As Jerome Bruner argues,

> "*... we organize our experience and our memory of human happenings mainly in the form of narrative – stories, excuse, myths, reasons for doing and not doing, and so on. Narrative is a conventional form, transmitted culturally and constrained by each individual's level of mastery and by his conglomerate of prosthetic devices, colleagues and mentors.*" (Bruner, 1991, p.4)

The videographer is creating narrative possibility by generating stories in her mind, which is represented (or not) in her clip collection depending on her success capturing video of story elements. Bruner identifies ten features of constructed narratives:

1. Narrative diachronicity – the temporal sequence of events in a story ;
2. Particularity – attention to detail in a story;
3. Intentional state entailment – realistic action by characters in a story;
4. Hermeneutic composability – the whole of the story is more than the sum of all its parts;
5. Canonicity and breach – the unexpected story grounded in the familiar one;
6. Referentiality – verisimilitude is more important to a story than verifiability;
7. Genericness – kinds of narratives that can be used to demonstrate life's situations;
8. Normativeness – the legitimacy of the narrative in a time, place and culture;
9. Context sensitivity and negotiability – the ability of a story to be understood across cultures with respect to narrative intention and background knowledge;
10. Narrative accrual- story collections become histories, cultures, systems, traditions.

The first two categories are concerns of this thesis and I would argue that without the first feature, the ability to make sense of events occurring over time, stories fall apart. Mindful documentary also involves the last feature. Narrative accrual in the mindful documentary work has an unusual role. It can be thought of as the commonsense knowledge collection. An existing story collection is used to actively intervene in the construction of new stories.[9]

2.2 Story Understanding

Story understanding by computers is the ability for a program to take a text-based story as input, parse it, and produce reasonable inferences that can enable question answering, story identification, and story summarization. Story understanding research in artificial intelligence began in the early 1970s. Case-based, rule-based, and connectionist models were developed to generate and guide inferences to support the generation of information that was not explicitly stated in the input text. Research in story understanding reached a plateau in the mid 1980s. Many researchers transitioned to work on information retrieval research such as message understanding. This was, in part, a response to the high overhead in designing systems able to answer questions and summarize only a few pre-determined stories. Researchers who persevered were faced with heavy overhead of hand-coding knowledge structures. Practically, a machine must have many capabilities in order to understand even a simple children's story (Mueller, 2002). Advances in technology, information retrieval, and availability of extensive knowledge bases create a climate ripe for a revival of interest in this subject. Some have proposed that researchers in story understanding and information reunite and tackle the problem together using methods generated in each field over the last fifteen years (Riloff, 1999).

This thesis is not an attempt at full story understanding and generation by a video camera. Instead, the goal is to provide story suggestions that are akin to the type of inferences possible in early story understanding systems. The methods of creating the inferences in StoryNet is akin to the frame-based methods of inference generation developed by Roger Schank and his students at Yale in the 1970s and 1980s (Schank et al., 1977). Their frame-based method used scripts' canonical representations for generic events. A script from a collection would be activated when its header matched a parsed version of the input text and inferences about causal relationships of events; object or prop necessities; and actor states. Scripts were invented as an inference direction and restriction method. When trying to understand a story, a person or computer does not want to keep generating all the possibilities at every granularity of expression and continue chaining together rules to create a story without any end. The mindful camera software creates story inference of events that have temporal and causal relationships. Issues of conflicting events and guiding inference are handled by user's interaction with the knowledge. The system generates possibilities and the user prunes the possibilities for sense and relevance to the situation. The system presents options and the videographer uses the 'scripts' in her mind to direct the inference and decide what is applicable not only to a generic story but to the specific story that she expects and is directly experiencing. The inferences created by the mindful camera system made by ConcpetNet and StoryNet are local. The annotation is considered the current story event and inferences are next possible story steps. Inference at each step is not influenced by all previous annotations. Each submitted annotation and inference can be considered to be bridging inference as defined by Clark in his work on the psychology of human inference (Clark, 1977). Bridging inference occurs when a listener or reader constructs implications between two given phrases. A single predicate in ConceptNet can be thought of as two concepts with a bridge (EffectOf "sunbathing" "get sunburn"). While LifeNet is also comprised of such predicates, the probabilistic interference is the result of all previously submitted annotations during the course of the documentary. Pairs of temporally linked events in StoryNet can also be considered bridges with the inference relations of occurring before or after.

Research in story understanding by computers has produced many other knowledge representations used to encode stories and perform inference including goals, plans, themes, space, and time. A good overview is provided in Mueller (2002) and will be discussed further in Chapter Four. In order to understand even a simple story such as "Ellis decided to run in a marathon. He came in second place" we need to know hundreds or even thousands of facts about people, marathons and competitions such as "people move when they run," "running is usually faster than walking," and "resting is an activity that usually follows running." A large repository of such knowledge and techniques for reasoning with it enable us to be flexible in the world, ask questions, and make inferences about the world as we observe it. Recent advances in commonsense reasoning support the understanding of broader story domains approximating the real world. (McCarthy et al., 2002)

2.3 Common sense

Common sense is the collection of knowledge and methods for reasoning we use to make sense of the everyday world. Although we make use of common sense during our daily life, in conversations, actions and activities, this knowledge is rarely made explicit. Research in commonsense reasoning by computers has two distinct approaches: logical and analogical, also called neat and scruffy. In 1959 John McCarthy published a landmark paper in AI called "Programs with Commonsense" that presented the first commonsense program called Advice Taker. The system could make deductions by manipulating sentences. In the paper McCarthy gives us a description of how a machine demonstrates common sense.

> *"... a program has common sense if it automatically deduces for itself a sufficiently wide class of immediate consequences of anything it is told and what it already knows."*
> (McCarthy, 1959, p.2)

Marvin Minsky, who is a long time advocate of a symbolic approach to commonsense reasoning, has more recently argued for using multiple representations and inference methods for commonsense reasoning.

> *"Human versatility must emerge from a large scale architecture in which each of several different representations can help overcome the deficiencies of the other ones. To do this, each formally neat type of inference must be complemented with scruffier kinds of machinery that embody the heuristic connecting between the knowledge itself and what we hope to do with it." (Minsky, 2002, p.70)*

Although the their theories and approaches differ, each are addressing the challenge of creating computers that have the kind of knowledge and reasoning abilities needed to understand the world. The mindful documentary work uses the analogical, scruffy approach to commonsense reasoning, but we acknowledge here that the studies with the camera will hopefully provide insight into which commonsense reasoning methods might be appropriate for the different challenges the videographer faces during story construction.

Commonsense reasoning is a different kind of intelligence than has been successfully demonstrated by expert systems which encode all the rules of a very small domain and reason to solve problems. As I discussed in the introduction, these systems are proficient within their intended domain. In documentary videography even a defined subject is very unpredictable and demands more breadth of knowledge than typical expert systems. In order to have a computer generate the kinds of inferences that humans do during storytelling and documentary creation, the computer needs common sense.

The largest scale project in commonsense reasoning is Cyc.[10] It was started in 1984 by Doug Lenat with the original goal of giving a computer enough common sense to read an encyclopedia. The Cyc knowledge representation is comprehensive and growing. The Cyc approach to knowledge engineering is to hire experts – knowledge engineers and philosophers – to enter knowledge using a special language called cyc-l. The Cyc upper ontology for representing the world is extensive and knowledge engineers must know the special representation language in order to contribute, search, and perform inference using commonsense knowledge. Initially, Cyc was considered as a resource for the mindful camera work. Ultimately, it was impractical because its translation from natural language into cyc-l and back again is not yet reliable yet. In other words, the videographer would have to learn cyc-l and the Cyc ontology to encode and submit annotations then read the suggestions. They would need to become experts in representing commonsense knowledge in cyc-l.

None of the existing large-scale semantic knowledge bases contain enough story knowledge to support broad reasoning during documentary construction. The breadth and depth of knowledge needed exceeds commonsense story knowledge available in existing systems. Mueller compared the story knowledge of several candidate systems (Cyc (Lenat, 1995), FrameNet (Johnson, 1998), Gordon's Expectation Packages (Gordon, 1999), ThoughtTreasure (Mueller, 1998), and WordNet 1.6 (Fellbaum, 1998)). While Gordon's database contains 768 scripts – more than any other system – the representation does not specify the arguments of script events. FrameNet and WordNet share this restriction, making inference about the ordering of events or instantiation requirements for any script very difficult. Cyc has the representational capacity to represent scripts as collections of assertions and has a very comprehensive ontology for representing events. The major drawback of Cyc, aside from the difficulty of knowledge entry, is an average of only two sub-events per script. Mueller's ThoughtTreasure script representation, the most comprehensive of the systems, contains eight sub-events on average per script and offers other useful representational elements that the other systems lack, such as duration and emotions and locations. Initially during the development of the mindful camera software, I anticipated acquiring a very large database of story-scripts for StoryNet. The ThoughtTreasure system contains a total of 100 scripts, initially thought to be not nearly enough scope to satisfy the story goals of the mindful camera. Ultimately, a smaller and domain-directed version of StoryNet was used in the evaluations for this thesis, but the aforementioned resources and their representational innovations informed the work.

Recently, with the advent of the WWW, rapid collective aggregation of large knowledge bases has become possible. Websites such as Wikipedia,[11] the Internet Movie Database,[12]and, more recently, Flickr[13] rely on large groups of enthusiasts, of varying levels of expertise, to submit knowledge to create an encyclopedia, review movies, and describe still photographs with still images, respectively. In 2000, Push Singh saw an opportunity to use a website to gather commonsense knowledge, in English, from everyday web users. Over the course of five years 750,000 commonsense facts were submitted to the knowledge base he created, Open Mind Common Sense, via filling in arguments to simple templates. ConceptNet and LifeNet were built from the Open Mind Common Sense corpus and their toolkits have been used recently in a number of commonsense-based applications. Most notably Aria (2002) is a photo annotation and retrieval tool that uses commonsense to expand the contextual keywords of a digital photo in order to retrieve photos relevant to the textual content of a user's email. Before ConceptNet was built, Liu & Singh (2000) made a story generator called MAKEBELIEVE in which users collaborated with the system to build a short, fictional, and text-based story. The user enters a sentence that catalyzes a chain of reasoning using assertions from the Open Mind Common Sense corpus of the "Effect Of" type. When the system gets stuck the user can enter another sentence to help the story along. The chaining is evaluated by a manager that checks for and eliminates loops and contradictions in the story.

There have been few commonsense-based applications that use commonsense resources other than Open Mind Common Sense. Gordon (2001) used his expectation packages, to improve browsing of subject terms in an image collection. The Steamer project used commonsense knowledge in a different way than context expansion or chaining binary predicates. Cyc-l was used to encode information about interface design. The representation of graphical knowledge and rules for design combinations helped the system guess the designer's intention and assist in the creation of a user interface for a copy machine (Members of the Human Interface Lab, 1898). These are just a few examples of commonsense-based applications. Research work in this area of commonsense applications is just beginning to gain momentum because commonsense resources are becoming more comprehensive and accessible.

2.4 Computational Documentary

The introduction of the computer into documentary videography has had a useful effect: video content can be accompanied by an information track, a machine-readable description of the video content and its uses (Negroponte, 1977). Design of the information track and heuristics for selection provide the means to dynamically assemble and present sequences from video collections. For the documentary maker, this means collecting, representing and organizing a set of possible stories, instead of a single linear narrative. The audience becomes an investigator and co-constructor of video material instead of a passive viewer. In this section we look at the following challenges in computational documentary: representing video, creating continuity, structuring story, and guided capture. We look at examples of systems that address these concerns and lessons learned that influenced the mindful documentary development.

The preservation of coherence during navigation has been explored in different ways. In some cases the navigation paradigm is closely associated with the real world, such as the navigation of a sequence of images of city streets (Lippman, 1980). Other systems develop methods for viewing that use computational models to determine if a video clip should be viewed next by measuring its relevance to clips previously shown, using methods in artificial intelligence such as spreading activation (Murtaugh, 1996). This creates associations between video clips based on shared keywords. Other systems, primarily using fixed databases of fictional content, use context-free, cinematic grammars as sequencing guides. In each of these systems the burden on the videographer is to constrain content capture to fit with the design of the system, such as shooting to satisfy a keyword representation-design of an information track.

In Aspen Movie Map, navigation was closely associated with the real world (Lippman, 1980). Using a touch screen, users could, in a first-person point of view, navigate seamlessly through images depicting the streets of Aspen, Colorado. Users "drive" through images of the city that are organized sequentially on videodisks. Jump cuts are implemented at intersections, hotspots, or season changes by switching between multiple videodisks. The significant computation in the system was the detection of user-activated hotspots, locations in the system that, when selected, would show a short documentary sequence. For example, selecting the police station hotspot would display a documentary sequence shot inside the station, of policemen during a typical day. System design and image capture were tightly coupled in the Aspen project. Using navigation of the physical world as the interaction pattern preserves the continuity of experience. Freedom to depict a story was relegated to hotspot locations. The model is a powerful one – using place as an organizing principle for video sequencing. Each journey through Aspen was unique, depending on how the user navigated and which hotspots were chosen. Spatial organization preserved continuity of the participant's experience. Narrative could be contained in the hotspot documentary footage but was not assembled by the system.

Computational documentary work inspired by observational documentary first strove for interaction and retrieval of video clips from a pre-recorded collection that could preserve story continuity for the user. Contour, a system developed by Michael Murtaugh, uses a method in artificial intelligence called spreading activation as an editor in software to present a documentary about urban development in Boston's North End (Davenport & Murtaugh, 1997). The documentary maker describes video clips using the keyword categories of person, location and theme. When the viewer chooses one clip, the system prioritizes clips with similar keywords and highlights them as candidates to be viewed next. The system preserves strong continuity along the axes of keyword categories. There are advantages to this system: the content database is extensible; knowledge of keyword categories can guide capture over a long period of time; and added user capability to select positive or negative weightings of video clips enables breadth-first or depth-first presentation of the content collection. Disadvantages are the limitations of sequencing by keyword and the spreading activation method itself. System action selects clips based on shared attributes, which conceivably could be played in any order. Sequences are not built on story relations between actors, events, objects, and places. This can be thought of as a telling, a sujet, in which the teller, here the computational system, has limited access to the content, or fabula, in the form of keywords. The Contour system did not aim to organize content into stories. It was successful at its goal of providing a next level of viewing continuity beyond temporal ordering of clips.

There has been significant research progress in areas such as representing video databases to a computer for non-linear editing and presentation of video footage; organizing video clips for recall; describing video content with annotation; and using annotation for reasoning about relationships between clips (Davenport et al., 1991; Davis, 1995; Kankanhalli & Chua, 2000; Bloch, 1988; Nack & Putz, 2001; Mateas et al., 2000). These methods focus almost exclusively on the post-production and presentation aspects of documentary creation. A notable exception is the work on providing video clip annotation during the moment of capture by Nack (Nack & Putz, 2001). The mindful documentary approach seeks to leverage human observation and commonsense computing in the field to better understand story possibilities as they begin and unfold. There are two major distinctions between the work reviewed in this section and mindful documentary. First, annotation and suggestions are brought into the capture process giving the filmmaker a way to reflect on the narrative potential of a collection while in the field. Second, the videographer and an intelligent camera work cooperatively to represent the documentary, by sharing story ideas and story representations.

2.5 Reflective practice

The pragmatist movement in philosophy pioneered by C.S. Pierce and William James incorporated the scientific method into the study of human thought (Menand, 1999). For Pierce the logician beliefs, meaning what we hold to be true, must be tested by observation in the real world. Pierce was objective in his theory of testing beliefs; a belief can be tested with the same method and evaluation as a scientific hypothesis. There is a proof! James is widely credited for defining the pragmatist movement, and in Pierce's opinion ruining it. James brought subjective knowledge into the deliberation process, and maintained that our mental interests, hypotheses, and beliefs are what cause us to act in the world. Therefore, they influence our decisions about what to hold as true. Observations about the world are not immune to the imprint of our current beliefs.

John Dewey was a pragmatist who incorporated the views of both Pierce and James into his philosophy of thought. Dewey is the champion of informal education and the grandfather of education theory in the United States. His theories about thought and action are mutually supporting. Thinking is decision-making supported by evidence gathered during action. Action is the process used to test and redefine our beliefs. Dewey defines three types of thought: everything that goes through our heads, anything beyond direct perception through our senses, and beliefs that rest on examined or unexamined evidence. Reflection is the examination of the basis for a belief (Dewey, 1933). If there were no reflection we would be bound to a life where our decisions were all based on unexamined beliefs or worse any thought running through our minds. Radically, he equates the ability to think reflectively as vital to learning and therefore freedom (Dewey, 1963).

Donald Schön's work called on professionals to use reflective thinking in their practice. Reflective practice is the methodology he developed to help professionals be more explicit, accountable, and revisionary in their professional lives (Schön, 1987). The word practice here has a dual meaning of a range of professional strategies (e.g., a lawyer's practice) and preparation for performing (e.g., practicing the piano). Both senses connote repetitive action that Schön would argue kills attentiveness and puts the practitioner in the danger of "overlearning." A habit is formed that hinders decision-making when a discrepancy is experienced. The default schema is used without deliberation over the possible necessity of accommodation. The professional has no trace of the decision-making process preventing any learning or formation of new schemas to accommodate an anomalous situation. Reflection-in-action is Schön's major contribution to theories of reflection. It inherits strongly from Dewey's coupling of action and thought. We can think about something and improve on performance while in the midst of the action. A baseball player can make adjustments to his position while pitching. Here reflection is about finding your groove by observing your own habits while trying to correct what does not work and preserve what does work. Reflection-in-action applies to more than just physical activities but also to any type of problem solving by practitioners – doctors making diagnoses, editors revising texts and teachers guiding students. Reflection-in-action must take place during what Schön calls an "action present." This is the period of time during which a change of action can make a difference in the situation. Another type of reflection defined by Schön is reflection-in-practice. This can involve refection after a situation has passed and requires the practitioner to make tacit understandings explicit and subject them to criticism.

Freire's critical reflection asks us to look at historical reality, and to diagnose problems in order to take action that will inevitably lead to social change by the disruption (or dismantling) of oppressive structures in society (Friere, 1977). Freire is relevant here because of his explicit conviction that dialogue is the medium of reflective practice. The process of understanding one's situation and acting on the possibilities is called praxis: action and reflection enabled by dialogue. Mindful documentary encourages reflection by an individual in conversation with a commonsense resource, which can be considered a representation of a community. The community is available to the videographer and offers suggestions about how to consider different possibilities. The commonsense resources are cultural construction built by volunteer contributors.

2.6 Reflective tutoring systems

Intelligent agents have been developed that make suggestions to users involved in a variety of tasks such as shopping, email management and perhaps, most infamously by "Clippy" the Microsoft paperclip, in word processing. The argument against these systems is that they presume too much about the user and usurp control. The direct manipulation advocates direct observation and control of all information in the system. Mindful camera employs techniques from both approaches in order to achieve maximum freedom for the user (Wexelblat & Maes, 1997). This freedom comes from the ability of the camera to better understand the goals of the user, an ability that we address using commonsense reasoning. This work is also informed by case-based tutoring systems (Schank, 1998) and systems that support learner reflection (Lin et al., 1999).

Computers are often used in educational contexts to scaffold the learning process. The computer, or technology, is used to help students explain known phenomena, organize their knowledge, realize knowledge gaps, and even recognize problems in their thinking methods. Computers can help students monitor, evaluate, and modify their thinking (Lin et al, 1999). Refection can be individual introspection (constructivist) or the reflection within social group (socio-cultural theory). These two categories of reflective learning are not absolute. Education technology designers are encouraged to mix and match depending on the task at hand and the goals of the learners (Scardamailia & Bereiter, 1996). By analyzing a set of technologies for reflective thought Lin, Hmelo et al. have identified four design approaches for implementing technology to aid a student or students in the reflective process. The four approaches are: process displays, process prompts, process models and a forum for reflective social discourse. Process displays are technologies for making evidence of the learning process visible by capturing and playing back the learner's actions for presentation as an artifact to be reconsidered during the learning process. The goal is to show students their process of solving a problem or successfully completing a task. In some cases a space is provided for students to note their thought processes during as the task is implemented (e.g., Geometry tutor (Anderson et al., (1985)). Process prompts encourage students to articulate their thoughts and make visible the reasoning behind their problem solving. Prompts are usually modeled after questions experts would ask in a similar situation as encountered in the learning task. The goal is student attention to their thought processes during learning. Not only is an object created to represent their actions and thinking but also students are prompted to unveil explanations that may not have been obvious during the action at hand. Process prompts might be employed at particular points in a task or during a state of activity of the learner (e.g., Isopod Sim (Lin & Lehman, 1999)). Process models model expert behavior for presentation to students during a task in order to scaffold students in the learning of a new domain. Students can compare their knowledge and thinking processes to those of the expert and reflect, revise, and retry. The goal is for the student to understand expert techniques for problem solving and compare their own techniques and products to the expert's. In forums for reflective social discourse, individuals reflect on their own efforts, other's efforts, and how these actions and thoughts contribute to the realization of group goals. Feedback from the group drives the revision of practice and guides learning. There are three benefits to reflective social discourse: multiple perspectives, audience encouragement for reflection, and group learning process as an artifact for reflection. The goal is to provide many points of view on problems and problems solving through social discourse (e.g., CSILE (Scardamailia & Bereiter (1996)).

The mindful documentary process and system uses process prompts but cannot be mapped exactly to the existing categories for reflective tutoring systems. It could be considered reflective community practice except that the contributors of the commonsense knowledge are not participating actively in the videography construction, nor would they want to. The goals of the knowledge acquisition community and the videographer are different.

The mindful documentary work shares a kinship with video systems that encourage students and learners to reflect on their own ideas and use video as a medium to construct new hypotheses and histories. Roy Pea and his students have created systems that encourage students to explore and repurpose video records to inspire discussion about topics (Pea, 1993). Pea has also recognized the importance of scaffolding in tutoring and learning systems. The software should not serve the answer to the learner but act as a coach that guides the student to discover solutions (Pea & Mills, 2004). Smith & Reiser (1997) and created a multi-media system that helps students make hypotheses about the predatory behavior of animals by guiding students though annotating and discussing video clips from nature television shows. This work is significant because the system *and* a methodology for interaction was designed to help students reflect on their beliefs and gather video evidence to support or debunk their understanding of animal behavior and evolution. Other video systems use video to query a user during a process and help her take next steps by offering expert advice. Schank (1995) created a system that gave advice to a plumber. Video clips of an assistant asking a series of questions was used to determine a plumber's problem then video clips of an expert plumber offering a solution would be shown to the user. The system used the case-based technique discusses in Section 2.2.

3 System Implementation

The mindful camera software is a set of client and server applications built in Python 2.0. The server is an XML-RPC server implemented in Python that has method calls accessing functions in the three commonsense resources: ConceptNet, LifeNet and StoryNet. The software runs locally on a Sony PCG/GT3-K laptop/video camera hybrid (Figure 7). The machine has a 512 MhZ processor, 128MB of RAM, and a 40GB hard drive. A separate application called "Sony Smart Capture" is used for video recording, as it is optimized for MPEG compression and designed to run on this specialized laptop, with its limited RAM capacity. This machine was chosen because it was the only device available that had the computing power needed for performing searches using large commonsense resources, a quality lens, and optical zooming capabilities. In addition, the two modes of operation are analogous to a camera and a laptop allowing novices easier mappings than in the original wearable idea for the system which would have required learning to use a heads-up display, controlling a wearable camera and typing into a chording keyboard.

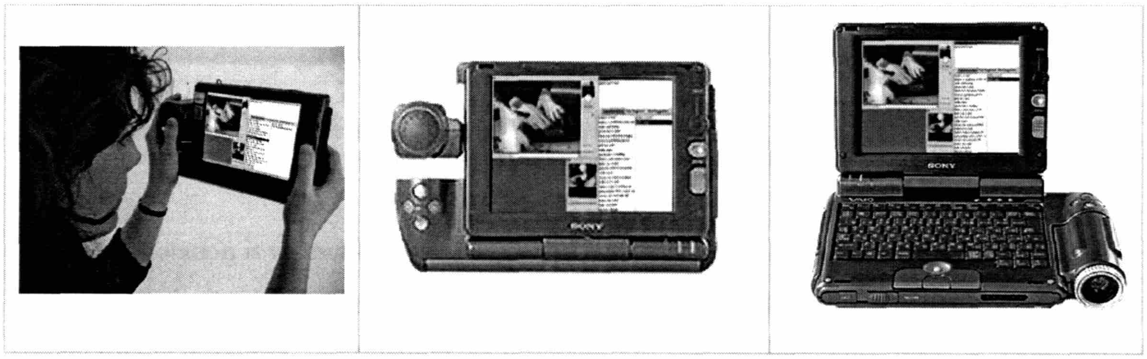
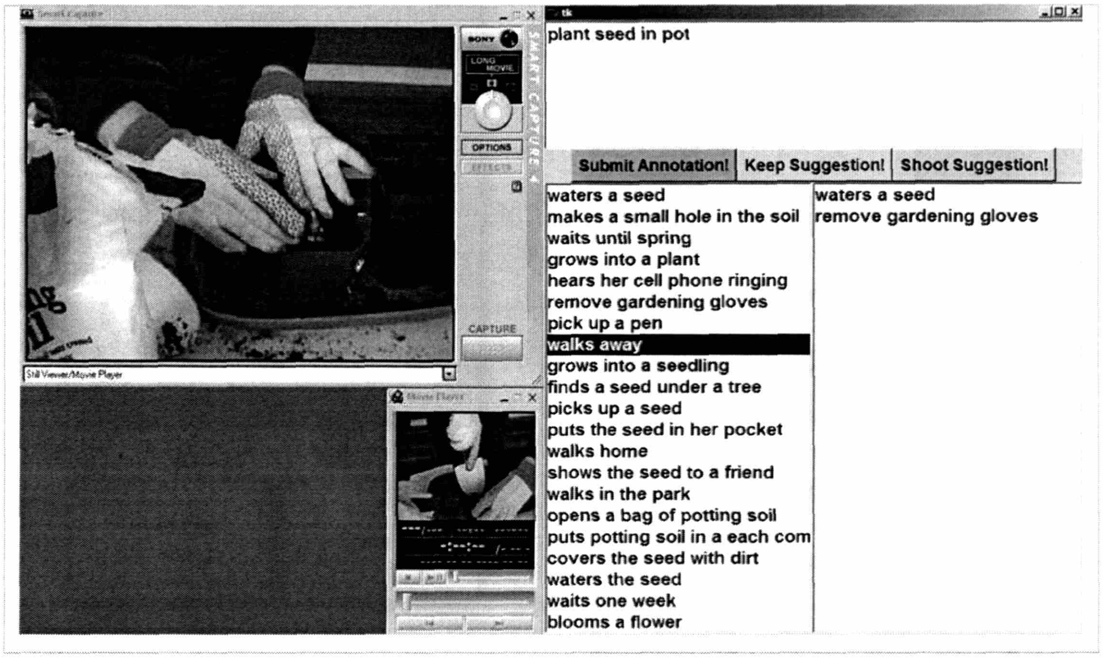


Figure 7: The camera platform running the mindful camera and video recording software. The picture on the left shows the camera being used in recording position. The picture in the middle shows the camera in recording position. The screen is turned and lies flat across the keyboard and the lens turns. The picture on the right shows the camera platform in the position for typing annotation and selecting suggestions. After each shot the videographer twists the screen from recording position to the open laptop position in order to type and select suggestions, then turns it back to recording position to shoot.

3.1 Mindful camera system

The interface has three major components (Figure 8): the video capture window, the media player window and the mindful camera software. To capture a clip, the videographer presses a physical button on the camera and begins recording. The same button is pressed to stop recording. Evidence of the recorded clip is shown in the media player as the first still of the captured clip and text string of the recorded video clip's time stamp is sent to the server and written to the trace file. The videographer then turns the camera into laptop position and types a text annotation into the top field of the mindful camera software window. The annotation is submitted to the server application by hitting the "Submit Annotation!" button.



plant seed in pot

waters a seed
makes a small hole in the soil
waits until spring
grows into a plant
hears her cell phone ringing
remove gardening gloves
pick up a pen
walks away
grows into a seedling
finds a seed under a tree
picks up a seed
puts the seed in her pocket
walks home
shows the seed to a friend
walks in the park
opens a bag of potting soil
puts potting soil in a each com
covers the seed with dirt
waters the seed
waits one week
blooms a flower

waters a seed
remove gardening gloves

Figure 8: Interface shows the video capture software, clip viewer, and the StoryNet version of the mindful camera interface.

The text annotation is written to a text file, parsed, and stopwords are eliminated; the remaining words are lemmatized and submitted as arguments to a function call which accesses one of the commonsense resources. The function call is designed to find the knowledge node or frame item that has the most matching words with the submitted annotation. The inference procedures are run. The resulting list of commonsense inferences is returned, passed to the client software, and up to twenty top inferences returned from the commonsense resources are presented to the videographer in the current suggestion list, the left hand window of the interface. With each submitted annotation the current suggestion list updates, erasing the previous suggestions and presenting the new ones. The system is designed to be fail-soft; appropriate suggestions can be selected using a check box then acted upon by the videographer, inappropriate ones ignored. Using the track-point mouse or the arrow keys the videographer can select an annotation and press the "Keep Suggestion!" button to add the annotation to her kept suggestion list, the rightmost column in the interface. The kept suggestion is sent to the trace file. The kept suggestion list grows over the course of the documentary shoot. The videographer does not have the ability to delete previously kept suggestions. During the course of a shoot the videographer may select a specific, previously kept suggestion from the list and press the "Shoot Suggestion!" button. This sends the text of the suggestion to the trace file indicating only that the videographer has noted that he or she would now attempt to record a shot based on that particular suggestion.

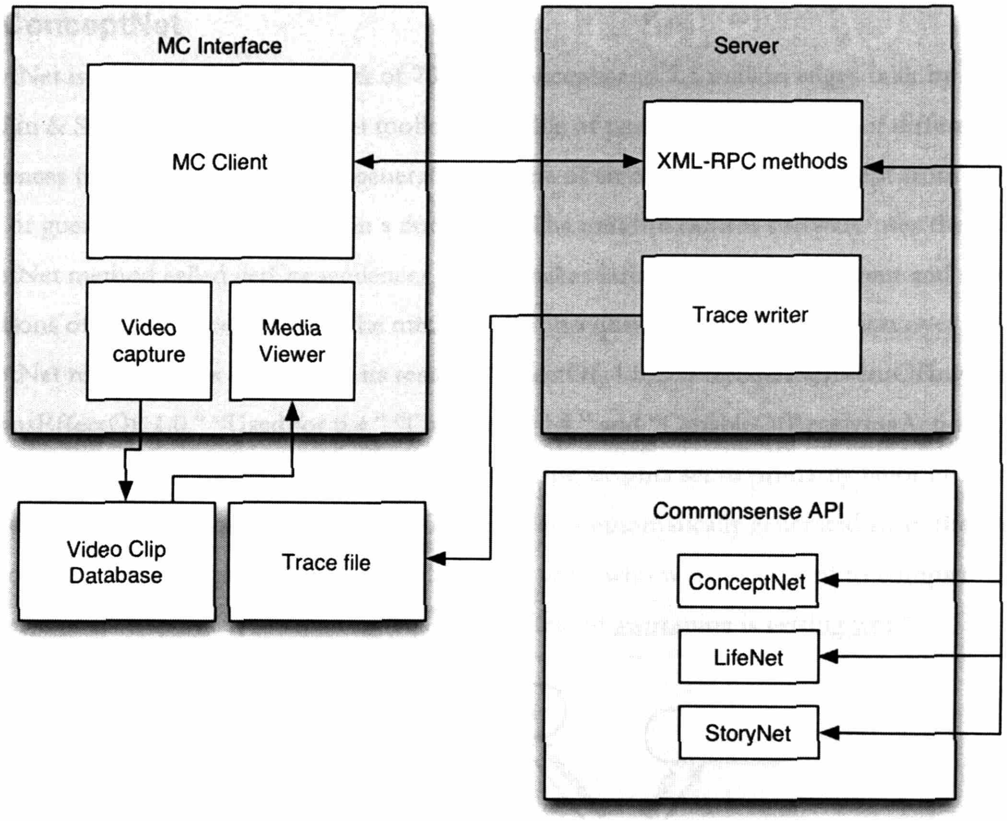


Figure 9: Diagram of the mindful camera architecture.

A system diagram is shown in Figure 9. There are five versions of the software, each a server and client package. The no-annotation version has a server that contains only a trace writer for annotation. ConceptNet, LifeNet, and StoryNet versions connect to their respective commonsense resources in the commonsense API.

3.2 Commonsense Inference

ConceptNet, LifeNet and StoryNet are the commonsense resources are that described in this section. As described in Section 1.5, the commonsense inference using each toolkit has been focused on predicting next events and generating a context for a situation. The three resources are in different stages of development. ConceptNet is the most developed toolkit and has been used in more than fifteen research applications (Liu, 2002; Lieberman et al, 2005). LifeNet is more recent and has only been used in one other application besides the mindful camera software, a voice recognition topic-spotter (Eagle and Singh, 2004). StoryNet is the youngest of the resources and still under its initial development cycle.

3.2.1 ConceptNet

ConceptNet is a large semantic network of 700,000 concepts and 1.5 million edges built by Liu and Singh (Liu & Singh, 2004). ConceptNet toolkit is capable of performing a number of different types of inferences from a given text. It can generate the parts of an object, guess a concept from a block of text, or guess the mood of the text in a document. The mindful camera software uses the ConceptNet method called getConsequences, which makes inferences about the event and object implications of a given input phrase. The method getConsequences spreads activation over six ConceptNet relation-types to generate its results: "EffectOf, 1.0," "PrerequisiteEventOfInverse 1.0," "DesirousEffectOf, 1.0," "UsedFor 0.4," "CapableOf 0.4," and "CapableOfReceivingAction, 0.3." The numbers following the relation types are the link type weights set to primarily favor events over objects during the inference of suggestions. ConceptNet is automatically generated from the Open Mind Common Sense corpus, which was built by web users who were prompted to complete English templates that corresponded to relations (e.g., "The effect of swimming is getting wet").

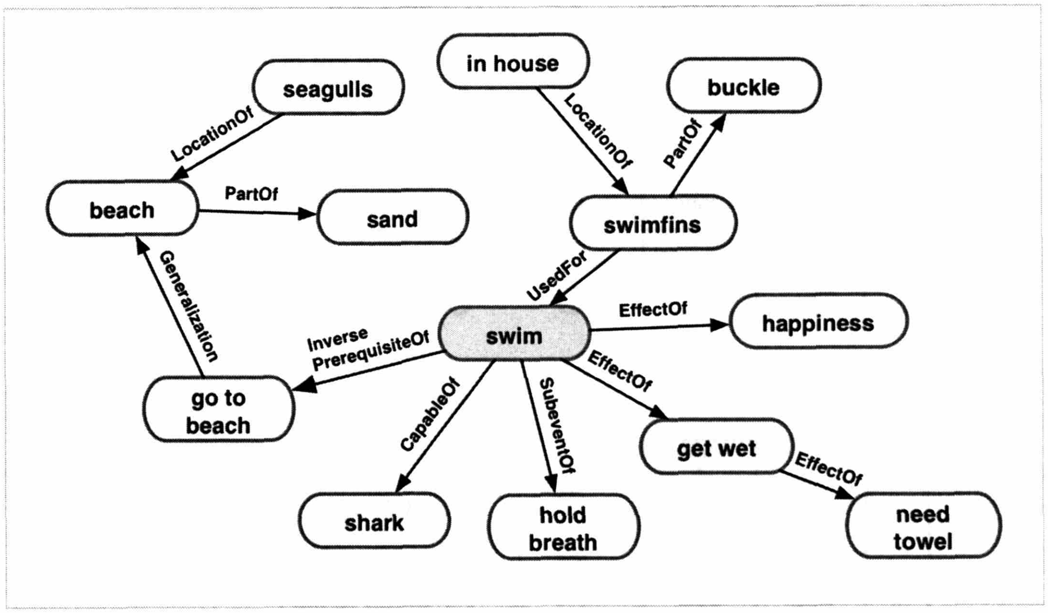
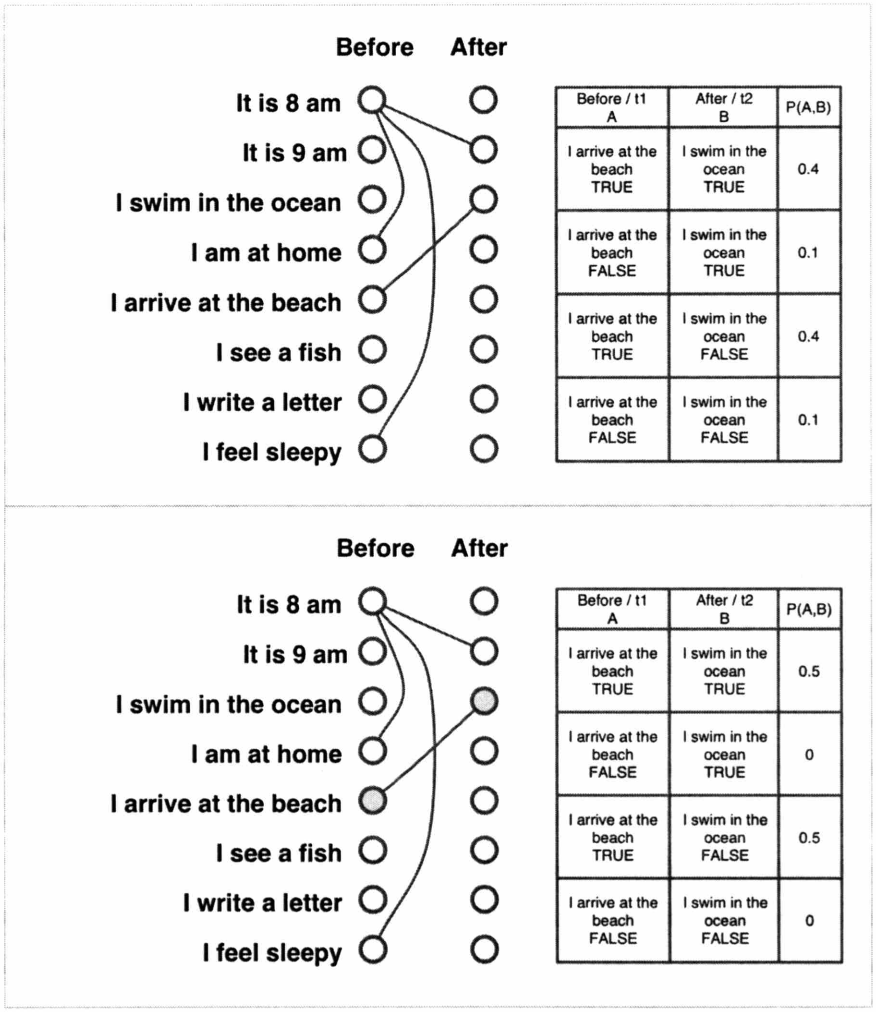


Figure 10: Visualization of inference using the ConceptNet function getConsequences ("swim")

Since knowledge entry for a given relation was not restricted to a part of speech there are sometimes states generated by the inference (e.g., "The effect of swimming is happiness"). When an annotation is submitted to the system by the videographer the ConceptNet resource spreads activation to related nodes in the network using the specified link type relations and weights. Concepts from the twenty top-scoring inference nodes are returned and displayed to the videographer. A simplified example of spreading activation in response to the annotation "swim" is shown in Figure 10.

3.2.2 LifeNet

LifeNet is a first-person probabilistic model of human activity. It is generated from a subset of the Open Mind Common Sense corpus using knowledge about: actions, objects, states, locations, and time. The network contains 80,000 nodes and 300,000 links between nodes. LifeNet has joint-probability tables for node pairs that are used for inference. Detailed information on the first version of LifeNet can be found in Singh & Williams (2004), and work on the current version can be found in Morgan (2005). The LifeNet version of the mindful camera software creates three separate LifeNet graphs to perform inference, each containing two time slices (An example of one graph is shown in Figure 11).



Before / t1 A	After / t2 B	P(A,B)
I arrive at the beach TRUE	I swim in the ocean TRUE	0.4
I arrive at the beach FALSE	I swim in the ocean TRUE	0.1
I arrive at the beach TRUE	I swim in the ocean FALSE	0.4
I arrive at the beach FALSE	I swim in the ocean FALSE	0.1

Before / t1 A	After / t2 B	P(A,B)
I arrive at the beach TRUE	I swim in the ocean TRUE	0.5
I arrive at the beach FALSE	I swim in the ocean TRUE	0
I arrive at the beach TRUE	I swim in the ocean FALSE	0.5
I arrive at the beach FALSE	I swim in the ocean FALSE	0

Figure 11: Diagrams of LifeNet graphs, each showing two time slices. The tables at the right of each diagram show the joint probability table between the nodes "I arrive at the beach" and "I swim in the

ocean." The top graph shows the joint probability table between the nodes "I arrive at the beach" and "I swim in the ocean" before evidence is created and belief propagation has occurred. The second graph shows the graph and conditioned probability table after the "I arrive at the beach" node has been set to true and belief propagation has occurred.

The first graph has all nodes set to their default probability values from 0 to 1, false or true. The second graph contains the nodes of evidence, nodes set to true, based on the annotation submissions. The third graph holds a version of the knowledge with updated inference. When an annotation is submitted values from the evidence are copied to the inference graph and belief propagation is performed to determine probabilities of the remaining nodes. For example, when a node "I arrive at the beach" is set to true its probability table is updated so that probability of "arrive at the beach" and "swim in the ocean" both being true is 0.5. The probability of "arrive at the beach" being true and "swim in the ocean" being false is 0.5. The inference graph can be queried to return the top twenty most likely nodes after belief propagation has occurred. The default graph is used for optimization only. The mindful camera software LifeNet version one activates true nodes as evidence that match annotation submissions over the course of the entire documentary. The LifeNet version two subtracts the former state of the network each time belief-propagation is performed after annotation submission. The second version makes the inference from the last submission more influential on the inference. In both cases the top twenty most probable inferences are returned as suggestions.

3.2.3 StoryNet

The goal of StoryNet is to eventually collect at least one million story-scripts from casual web users as a corpus for commonsense reasoning. A few trail acquisition sites have been built and are described in Singh & Barry (2003) and in Singh, Barry and Liu (2004) which explored different story-script representations and acquisition techniques. The work is ongoing. The mindful camera system uses a smaller, domain-specific version of StoryNet that contains 30 story-scripts totaling 231 story events. I built the StoryNet database by using story-scripts submitted by users of the StoryNet acquisition website prototype and hand coding story-scripts from text stories submitted to the Open Mind Common Sense story building activity, asking people to tell me stories about planting seeds and looking at gardening websites. The story-scripts submitted in the prototype acquisition systems were simplified to the frame representation shown in Figure 12. Each event in a case is associated with a slot in the frame that specifies the actor who performs the action. In the studies of the mindful camera system described in this thesis actor to action bindings were not used in suggestion generation.

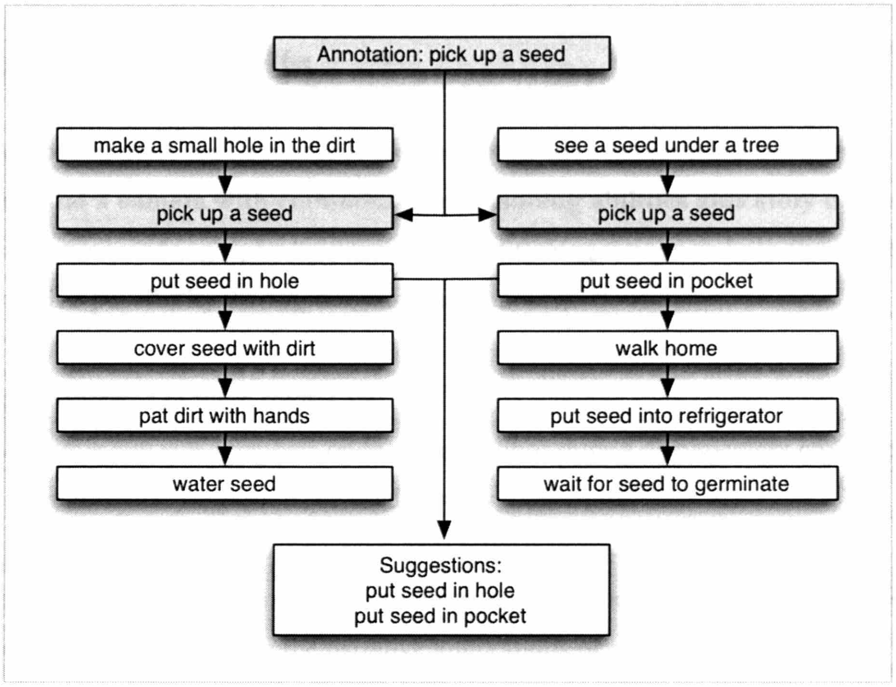


Figure 12: Visualization of inference in StoryNet..

Submitted annotations are matched to slots in story-script frames and the next event in each matching story-script is returned as a suggestion. The shown suggestions are not limited to twenty results. All relevant suggestions are shown to the user.

StoryNet Inference:

```
For each accepted event E,
    for all story-scripts S containing event E,
        output the event that follows E in S.
```

4 Evaluation and Results

The main purpose of these evaluations is to see if **the reflective partnership between a videographer and a camera with commonsense reasoning abilities aids story construction during documentary videography by providing useful suggestions to help the videographer identify narrative possibility during documentary construction.**

The questions addressed in these studies are:

1. Are documentary videographers open to suggestions during documentary construction?
2. Can commonsense suggestions be used as story suggestions?
3. Can a camera with commonsense reasoning tolerate domain shifts?
4. How do the different commonsense resources support documentary construction?
5. Do videographers reflect on the narrative implications of suggestions?
6. Does reflection on suggestions change the videographer's recording path?

To address these objectives, versions of the claims stated in Section 1.4, the system was tested in the field using methods previously based on intelligent tutoring system evaluations described in (Iqbal et al., 1999). The evaluations of this work consist of two studies of the mindful camera system. In the first study a videographer, this author, used the camera to document four day-long marathons. The second study had four videographers, two experts and two novices, who created short documentaries of a gardener planting seeds, which lasted less than ten minutes each. Each videographer shot a documentary without suggestions and then multiple documentaries each with a different commonsense source of suggestions. Individual marathon shoots were recorded using ConcpetNet suggestions and two different versions of LifeNet for suggestions. Individual seeds shoots were documented using ConceptNet then StoryNet suggestions. In summary:

1. Marathon studies
    a. One videographer
    b. Four documentaries shot, each of a different marathon
        i. No annotation
        ii. ConceptNet suggestions
        iii. LifeNet suggestions (version 1)
        iv. LifeNet suggestions (version 2)

2. Seed planting studies
    a. Four videographers
        i. Two experts
        ii. Two novices
    b. Each videographer shot three documentaries
        i. No annotation
        ii. ConceptNet suggestions
        iii. StoryNet suggestions

During each documentary shoot the videographer recorded individual shots and annotated each clip with text-based annotations. Suggestions (when generated) were accepted by being placed in a shot list or ignored. Data was collected by tracing videographer activity during the shoot, through questionnaires and post-shoot feedback from the videographers. The system created a trace after each shot decision of the date and time the shot was taken, the submitted annotations, the generated suggestions, the suggestions accepted into the shot list, and suggestions in the shot list that were taken explicitly as a direction for the next shot. The questionnaires provided feedback about interactivity and the perceived quality and impact of the suggestions. Post-shoot discussions offered videographers' observations and insights about the system, the methodology, and their experience as a documentary maker.

4.1 Examples

This section presents example traces of videographer interaction with the system. The left column of each shot table shows a representative still frame of a video clip and the adjacent columns show the trace of activity for that shot. These examples are from marathon documentaries in which LifeNet was used for suggestion generation.

4.1.1 Taken Suggestions

A taken suggestion is one that the videographer selects from the suggestion list and places in her suggestions list. (Figure 13) The videographer may choose to follow-up immediately on the suggestion or wait until another time during the shoot. Videographers are not given any specific criteria to judge the suggestions. They are told at the beginning of the study that they are under no obligation to take suggestions. The can take them if they like or ignore them.

Annotation: "runner threatens to fight other runner"

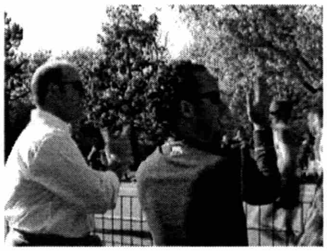
Suggestions:
make a battle plan
increase a interest rate
lower price
raise interest rates
increase in price
reduce interest rates
attack a leader
represent my country in battle
shoot another man
throw things
feel mad
spend less
throw a punch
make fun of him
protect myself from a enemy
gather information about my enemy
plan an assault
fall in price
punch somebody
throw my drink on him

Kept Suggestion: **attack a leader**

Figure 13: Example of a clip with annotation, suggestions and one taken suggestion.

4.1.2 Ignored Suggestions

Ignored suggestions are simply left in the suggestion list and overwritten at the time of the next shot action. Figure 14 shows an example of a set of suggestions that were ignored. Suggestions can be ignored for a variety of reasons. A suggestion might not make sense as a concept due to the scruffy nature of the commonsense knowledge. A suggestion might not be applicable to the domain or have already occurred previously as a taken suggestion.



Annotation: "encourage contestants"

Suggestion List:
exercise fun human interaction
move a white knight
move a white pawn
remember my childhood
improve physical fitness
plan strategy
lose to another team
make an opening move
take a break to drink water
move at a fast pace

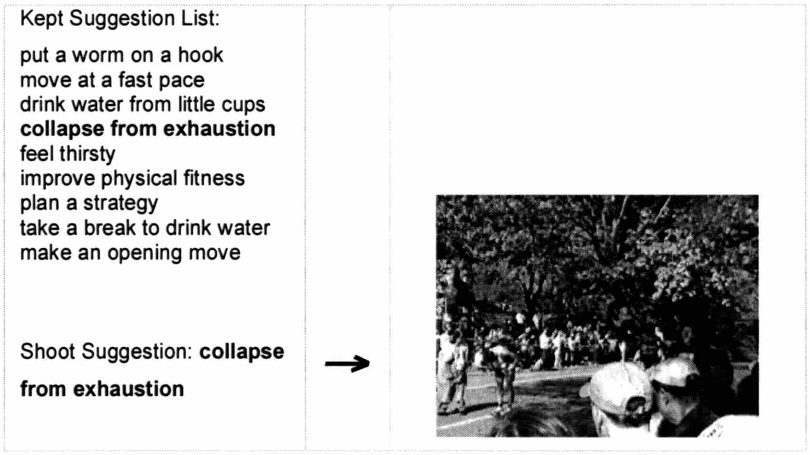
Kept Suggestions: None

Figure 14: Example of a clip with annotation, suggestions and all suggestions ignored.

In addition, the videographer is taking or ignoring suggestions based on how she thinks the suggestion will accord with possible events in the world. If a suggestion seems possible but unlikely the videographer might choose to ignore it. Timing is also an issue. If the videographer is shooting many clips in a short time period she may not have the time or the desire to read all of the presented suggestions.

4.1.3 Shot Suggestions

A videographer can attempt to capture a shot from her kept suggestion list at any time during the shoot. Videographers were encouraged to hit the "Shoot Suggestion!" button in the mindful camera interface after they had chosen to look at the kept suggestion list and attempted to make a shot based on a suggestion. This was not always the case. In reports by videographers, they often shot suggestions immediately after accepting them into the kept suggestion list or simply shot one from the list without using the interface to trace their intention. An example of a shot suggestion is shown below in Figure 15.



Kept Suggestion List:

put a worm on a hook
move at a fast pace
drink water from little cups
collapse from exhaustion
feel thirsty
improve physical fitness
plan a strategy
take a break to drink water
make an opening move

Shoot Suggestion: **collapse from exhaustion**

→

Figure 15: Example of a previously taken suggestion being selected from the "Kept Suggestion List" and used by the videographer to direct the next taken shot.

4.2 Suggestion Data

4.2.1 Marathons

System Location	Shots	Suggestions Shown	Suggestions Kept	Suggestions Shot	Kept/Shown	Shot/Kept
Annotation Only Portland	85	**n/a**	**n/a**	**n/a**	**n/a**	**n/a**
ConceptNet Falmouth	110	1100	62	7	5.64%	11.29%
LifeNet v1 New York City	92	920	16	5	1.74%	31.25%
LifeNet v2 Philadelphia	51	1020	12	1	1.18%	8.33%

Chart 1: Marathon data totals from each of four documentaries.

Chart 1 shows the data from all marathon documentary shoots. The videographer took suggestions and saved them in the cumulative taken suggestion list in each of the marathon shoots in which suggestions were offered. In addition, in each of the documentaries some taken suggestions were used as shot suggestions, directives to take a particular shot. During the course of each marathon the videographer was presented with a large number of suggestions. It would be impossible for any videographer to consider, accept, and shoot all suggestions that are offered after an individual shot action. The number of suggestions offered was to ensure that some would be relevant to the subject and to the situation. The number of suggestions offered also tested the videographer's threshold for judging individual suggestions. During the documentary shoot it was helpful to have larger lists with some suggestions not considered rather than shorter lists with nonsense or irrelevant suggestions.

4.1.2 Seed Shoots

System version	Shots	Suggestions Shown	Suggestions Kept	Suggestions Shot	Kept/Shown	Shot/Kept
Annotation Only	30	n/a	n/a	n/a	n/a	n/a
ConceptNet	20	199	8	4	4.02%	50%
StoryNet	23	198	27	8	13.6%	29.6%

Chart 2: Seeds shoot data totals from four videographers.

Chart 2 shows the collective data from all videographers in each of the seed-planting shoots. These documentaries were shot in a very shot time frame: each individual shoot took less than ten minutes and generated a collection of between four and ten video clips. Again, it would be impossible for the videographers to take and act on the number of suggestions that were offered. Videographers took suggestions in each of the seed shoots in which they were offered. Feedback from experts and novices differed in that experts took more suggestions from StoryNet and novices took more suggestions from ConceptNet.

4.3 Commonsense Story Suggestions

The suggestions generated, accepted, and ignored were examined to understand the narrative implications of suggestions. How does a suggestion help the videographer assemble the pieces of her narrative puzzle and collect new pieces that increase the narrative possibility of her collection? Figure x shows the types of narrative suggestions that can be seen in the Falmouth shoot. The categorizations were determined after the shoot as part of the analysis of the data. There were also ignored suggestion types. These are not narrative categories but reasons why a suggestion might be ignored. The rejected suggestion types are shown in Figure x.

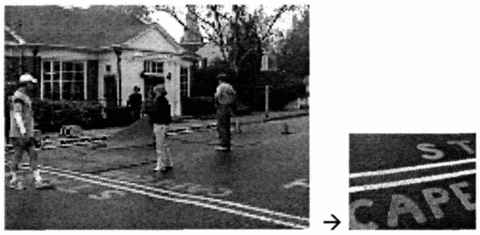
4.3.1 Story Categories for Accepted Commonsense Suggestions

It was the intention in designing the mindful camera system to focus suggestions on events and objects in a story. However, events and objects have commonsense dependencies. Two events might occur simultaneously in a story but not be interconnected or causal. In future systems, the following story categories for suggestions can be used to give the videographer and system increased specificity of suggestions; for example, if a videographer knows that the role of an actor is to perform a particular task at a given location the videographer might want to follow all possible effects of that event. In addition, an actor's actions determine if he or she is making progress on or abandoning a goal. These ideas call for more complex story structures than predicting next events based on common sense. However, there are a few promising implications from the categorization work. First, even though suggestions are based on temporal relations, some of the commonsense knowledge was not exclusively verb phrases describing events. States were often returned as effect of actions during inference using ConceptNet. This was somewhat expected, since the templates for knowledge collection were did not restrict part of speech. However, it was not expected that some of these suggestions would be taken at a more thematic level of description and acted upon. It was expected that a videographer would be more likely follow a suggestion that contained an action rather than thematic concept. Second, one of the accepted suggestion types "get details" could be considered not only a content suggestion but a formal suggestion about how to frame the shot. In other words, the language of film can be derived from the common sense of how we look at the world. In the early stages of this work there was much debate about the focus on the content suggestions without formal shot suggestions such as "close-up" or "establishing shot" as cues for how the videographer should frame the action of the shot. Initially, I intended to interview filmmakers and develop a set of formal suggestion types (Barry, 2002). It quickly became obvious that the formal decisions for a shot are dependent on the content and the situation of the videographer. Hard-coded rules that direct a videographer to take an establishing shot as the first of her documentary enforce a particular narrative filmmaking style. One can always find contradictions to these rules. Following is a list that shows an example of each suggestion types.

4.3.1.a List of Accepted Commonsense Suggestions

Get Detail	Suggestion zooms in on one detail of the situation
Role of Actor	Property of an Actor in a situation
Actor Action	Next action, short or long term
Action in location	Where an event would occur
Thematic Pointer	Identifies a possible theme present
Script Potential	Subset of suggestions can be arranged to form a simple script
Characterization of Action	Generalized description of an action
Effect of Action	State of an actor or world as the result of an action
Object in Location	Objects that contextualize a location
Prior Event	An event that could have happened before the last recorded action
Actor State	Identifies a possible mental state or expression of an actor during an action

4.3.1.b Examples of Accepted Commonsense Suggestions

A. Get Detail Suggestion zooms in on one detail of the situation  Example: Falmouth Clip #1 Annotation: "The volunteers set up the starting line" Suggestions: "volunteer, up, starting, sets, the, line, be in charge of project, raise, join committee, get in line" Taken suggestions: 1) "get in line" 2) "be in charge or project," 3) "line" Shot suggestion: 1) "line"	B. Role of Actor Suggestion identifies a property of the actor in the situation Example: Falmouth Clip #1 Annotation: "The volunteers set up the starting line" Suggestions: "volunteer, up, starting, sets, the, line, be in charge of project, raise, join committee, get in line" Taken suggestions: 1) "get in line," 2) "be in charge or project," 3) "line" Shot suggestion: 1) "be in charge of project"
C. Actor Action Suggestion identifies a possible action by an actor in the situation	D. Action in location Suggestion is taken because it predicts a future event

Example: Falmouth Clip #10

Annotation: "The police are on bicycles"

Suggestions: "police, on, bicycles, move car, wear, have gas, wear uniform, report, issue,

Taken suggestions: 1) "arrest"

Example: Falmouth Clip #1

Annotation: "The volunteers set up the starting line"

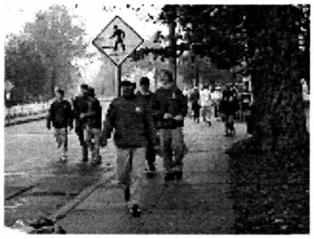
Suggestions: "volunteer, up, starting, sets, the, line, be in charge of project, raise, join committee, get in line"

Taken suggestions: 1) "get in line," 2) "be in charge or project," 3) "line"

Shot suggestion: 1) "get in line"

E. Thematic pointer

Suggestion identifies a present theme



Example: Falmouth Clip #4

Annotation: "The volunteers arrive"

Suggestions: "volunteers, go to school, drive car, go for walk, go to market, go for drive, curiosity, take car for drive, accident"

Taken suggestions: 1) "curiosity"

F. Script potential

Subset of suggestions could be arranged to form a simple script

Example: Falmouth Clip #7

Annotation: "Runner stretches"

Suggestions: "runner, stretch, run marathon, win, start, pass, get tired, compete, run in marathon, tiredness"

Taken suggestions: 1) "run marathon", 2) "win", 3) "compete", 4) "tiredness

G. Characterize action

Suggestion characterizes the action of the annotation

Example: Falmouth Clip #3

Annotation: "Putting mats on the street"

Suggestions: "street, putting, the, mats, on, curb, take bus, cross, help, passage"

Taken suggestion: 1) "help"

Shot suggestion: None

H. Effect of Action

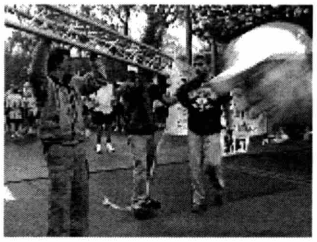
Suggestion is the effect of an action.

Example: Falmouth Clip #24 (note as video clip progresses we see runner run by woman with flag)

Annotation: "The runners run by pier while volunteer waves a flag"

Suggestions: "pier, volunteer, run, flag, runners, by, while, waves, cramp, raise"

Taken suggestion: 1) "cramp"

I. Object in Location Suggestion identifies an object at a location  Example: Falmouth Clip #13 Annotation: "Putting up the starting line" Suggestions: "put, up, the, starting. Line, club, enjoy person's food, help, equipment, newspaper" Taken suggestion: 1) "help", 2) "equipment"	J. Prior event Suggestion identifies an event that occurred before the event of the suggestion. Sovereign Bank Cape Cod Mar DUNKIN' DONUTS Marathon Relay START Example: Falmouth Clip #15 Annotation: "The runners line up at the start" Suggestions: "start, up, the, runners, at, line, wake up in morning, get marry, start family, pass sentence Taken suggestion: "wake up in morning" Unless the event is recurring it would be impossible to shoot it since it is now history. It does present an opportunity to learn for the next shoot.
K. Actor State Suggestion identifies a possible mental state of an actor. Example: Falmouth Clip #18 Annotation: "singer greets the runners" Suggestions: "singer, runners, greets, play guitar, sing, wear, happiness, sing song, take drug, carry" Taken Suggestion: "sing song", "happiness"	

These suggestion types can be thought of as primitives for building story patterns. Chaining suggestion types can give the videographer more direction about how to tell the story. If a suggestion about an actor's state is accepted by the videographer, the system could know to filter the next set of suggestions to show actions that are usually the result of that state. An accepted suggestion of "happiness" might be followed by "smile." In this case, chains of suggestions would be offered from a single annotation instead of immediately moving on to the next clip. In the case of ConceptNet, a way to recognize the narrative suggestion categories would be necessary to add this chaining of suggestions. Another approach would be to track patterns of accepted suggestions to try to understand how larger story structures are being built as the videographer collects clips. Lehnert created a story understanding system for detecting thematic patterns in text-based stories called plot-units (Lehnert et al., 1983). The system parsed a text-based story into conceptual dependency representations; graphs were constructed based on classification and relational rules; and the resulting graphs were matched to a set of pre-determined patterns, each identified by Lehnert as a story theme. It would be possible to design combinatory rules for relating different commonsense suggestion types to better understand how to reflect story possibilities back to the videographer.

4.3.2 Categories of Ignored Commonsense Suggestions

At first I was calling suggestions that were not taken "rejected suggestions," but after thinking further and talking to the other videographers I realized that there were different and varied reasons for not keeping a suggestion. The videographer might be occupied with observation or want to take a next shot and decide to skip reviewing suggestions or only review the first few. A videographer might consider a suggestion but not accept it because she thinks it will not happen in the world and would therefore not be a shot that could be captured. In some cases, frustration in waiting for suggestions caused videographers to take a break from looking at the list, decide to ignore it or collect multiple clips before annotating. The list below is a list and descriptions of reported reasons suggestions were ignored during shooting.

4.3.2.a. List of Ignored Suggestions

Echoing	No suggestions taken due to nearness of returned suggestions to original annotation
Repeater	Suggestion already taken and added to shot list earlier in the shoot
Alternative Context	Suggestions are generated for a similar event but in a different context
Nonsense	Suggestions have no reasonable, commonsensical relationship to the submitted annotation
Explicit	Suggestions were already demonstrated in the previously captured video shot
Not Shootable	Suggestions are impossible or very difficult to transform into a segment of video

4.3.2.b Descriptions of Ignored Suggestion

A. Echoing

No suggestions taken due to nearness of returned suggestions to original annotation

Example: Falmouth Clip #6

Annotation: "The manager of the race is on a cell phone giving instructions"

Suggestions: "race, cell, manager, phone, on, of, giving, instructions, call, start"

Taken suggestions: None

B. Repeater

Suggestion already taken from previous shot

Example: Falmouth Shot #2

Annotation: "starting line"

Suggestions: starting, line, get in line, indicate melody, see movie, express yourself, pass, carry along street, construction, take car for drive

Taken suggestions: None

C. Alternative Context

Suggestions are generated for a similar event but in a different context

Example: Falmouth Shot #8

Annotation: "the runners arrive at the race"

Suggestions: race, at, arrive, runners, start, bicycle, ride horse, go to school, horse, drive car

Taken suggestions: None

It is worth noting that after the shoot was over it became obvious that some of the suggestions generated were about horse races, auto races or bicycle races. If the system could have alternative contexts identified analogies could be useful in future story generation. Articulating the differences between kinds of races could lead to suggestions meant to give the audience a new way to look at marathons. [microtheory needed]ß

D. Nonsense

Suggestions have no reasonable connections to the annotation

Example: Falmouth Shot #9

Annotation: "runners stretch"

Suggestions: stretch, runners, turtle, machine, one nightly ritual, high rise, do have tooth, pump, magnetic field within sun, not domesticate animal

Taken suggestions: None

D. Explicit

Suggestions that are obviously already contained in the shot whose annotation was the source of the suggestion

E. Not Shootable

Suggestions that state something true but are too difficult (abstract) to transform into a shot suggestion.

 Example: Falmouth Shot #12 Annotation: "tie shoe" Suggestions: tie, shoe, tie person's shoelace, ribbon, tell story, yellow ribbon, chord, knot, rope, wear" Taken suggestions: None	 Example: Falmouth Shot #17 Annotation: "Runners dressed for Halloween" Suggestions: "halloween, change person's appearance, runners, dressed, for, change, mirror, confusion, look, gain weight" Taken Suggestions: None

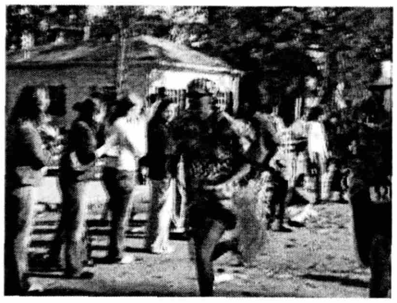
Looking at ignored suggestions provided some insights. If suggestions are considered on a scale from "general" to "specific," videographers are likely to ignore suggestions that are on the extreme ends of the spectrum. If suggestions are also considered on a scale from "expected" to "unexpected" videographers are also less willing to take suggestions that are on the extremes of the spectrum. They used the terms of the spectrum to describe particular suggestions that they rejected. This presents a challenge for future system design. How can we know where a suggestion falls on a spectrum? It would provide better support for the videographer if the system behaved like a companion that did not contribute obvious or unrelated suggestions. This disconnects the videographer from the interaction and breaches the trust of the system.

4.4 Domain Shifts and Fatigue

Documentaries often situate the viewer in more than one context. A documentary about one subject can quickly turn into a documentary about something else when one finds a volleyball tournament at the beach or when a riot breaks out at a marathon. The camera needs to be able to shift domains to provide narrative possibility beyond one story. ConceptNet and LifeNet demonstrated this ability. Examples are shown in Figures 16 – 19.

	"ocean haze" haze, ocean, surf, ride horse, sailor, open bag of chip, revolution, sea, large body of water, telephone book TS: surf

Figure 16: Marathon along the beach

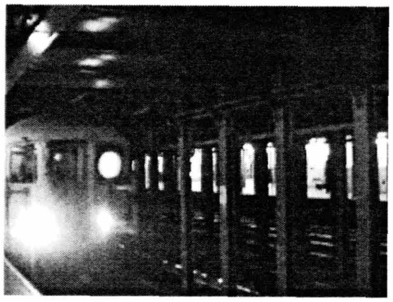


Annotation: "runners in halloween costumes"

Suggestions: halloween, costume, runner, change person 's appearance, in, change, mirror, run marathon, win, confusion

TS: change person 's appearance

Figure 17: Marathon on Halloween

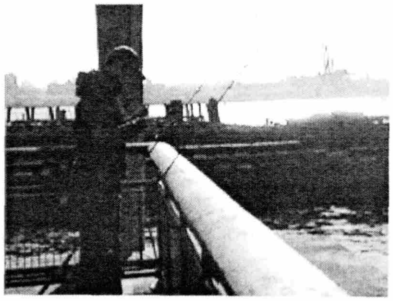


Annotation: "city"

Suggestions:
travel by bus, travel on public transport, step over something walk instead of driving, help with a heavy work, entertain a cat, break a record, give up my civil rights, see environmental scanning, take away sickness

Kept Suggestion: travel on public transport
Kept Suggestion: walk instead of driving

Figure 18: Marathon in the city



Annotation: "fishermen fish"

Suggestions: make a family happy, spend time with nature, gather wood for fuel, put a worm on a hook watch someone grow old, smell up a house, gather firewood. run out of gasoline

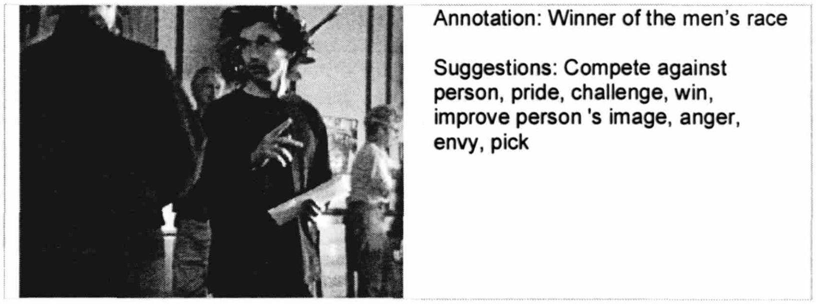
Kept Suggestion: I put a worm on a hook

Figure 19: Fishing near the marathon starting line

4.5 Comparing Commonsense Resources

The videographers who shot the seed planting documentaries had different reactions to the commonsense resources. The novice videographers preferred suggestions from ConceptNet, even though they thought the StoryNet suggestions demonstrated more intelligence. They reported that the StoryNet suggestions were too specific and directed. The experts reported that the StoryNet suggestions were generally more useful. It is worth noting here that although each group reported favoring a suggestion source, the impact of suggestions on their recording path did not agree with their preferences for suggestions. Each videogapher during the seed shoots experienced one recording path change. Of the experts one change was in response to a ConceptNet suggestion and the other in response to a StoryNet suggestion. The novices were also split; one had a recording path change due to a ConceptNet suggestion and the other due to a StoryNet suggestion. I was the only videographer to shoot with the LifeNet version of the system and chose not to have the seed documentaries shot with LifeNet because of the short length of the shoot. The cumulative updating of the network would not have worked during a shoot with so few annotated video clips. In addition, the LifeNet process of inference was the slowest of the three systems and impractical on a shoot of a short sequence. The videographer would have missed too much of the action of planting a seed while waiting for the suggestions to be returned.

4.5.1. ConceptNet



Annotation: Winner of the men's race

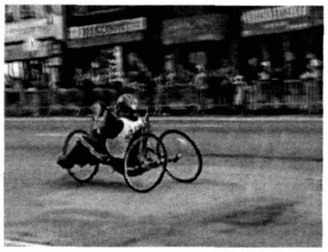
Suggestions: Compete against person, pride, challenge, win, improve person 's image, anger, envy, pick

Figure 20: Example of ConceptNet suggestions

ConceptNet suggestions gave the videographer more room to generate her own connections between the annotations, the suggestions served, and possible events in the world. One videographer reported "...even the nonsense gave a language-based way to create metaphor or let my mind make sense [of it]." This brings up the issue of the inference abilities of people. Could a videographer make sense out of any random suggestion? ConceptNet suggestions were reported to be more random than StoryNet suggestions by the videographers. The novice videographers said they would like to use the ConceptNet suggestions in a shoot over the course of a day instead of in such a constrained shoot with and assigned subject. They thought the system might encourage them to think differently and creatively about their observations over the course of a day. I found the ConceptNet relations helped me release my thinking from a declarative and descriptive mode to a more loose and poetic mode. It reminded me of the productive thinking one does on the edge of sleep when ideas are arranged in new ways. In future tests with the system it would be worth having a group of videographers who are served random suggestions to contrast with another group who are served suggestions based on ConceptNet relations. It would be useful to compare the acceptance, impact, and reporting on the thinking of the videographers during documentary construction.

4.5.2. LifeNet

During the LifeNet shoots when annotations were submitted the probability updating did not always change the most likely 20 nodes in LifeNet. It was frustrating as a user to input to the system with no observable changes and identical suggestion output. This was due to the size of LifeNet and the matching techniques for suggestions to LifeNet nodes. Often a submitted annotation would not have a matching node in the database or a matching node was not relevant enough to other nodes for the updating to show an effect. The example in Figure 21 shows the turnover of suggestions, a sampling from the 20 annotations that did not produce a turnover, then the example of the annotation that produced the change. Notice that the sets of suggestions that bracket the static run of suggestions are highly successful.

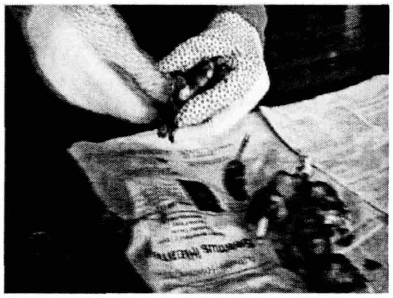
	query is: wheelchair contestant Suggestions: check my diet. reduce my heart rate, hold paper together, dispose of a body, serve vodka, see first class,play more music, travel to a destination via train, plug in a light, suck up houses
	query is: ride in a wheelchair Suggestions: have money to buy things, move at a fast pace, throw my bat, learn to walk, feel thirsty, wear high tops, choose a location, collapse from exhaustion, work too hard, drink water from little cups
	... (suggestions persist in this state over 20 clips...)

	query is: run in park Suggestions: move at a fast pace, have money to buy things, throw my bat, learn to walk, drink water from little cups, work too hard, wear high tops, choose a location, supply a baseball team, rest after work
	query is: spectators watch Suggestions: move at a fast pace, have money to buy things, throw my bat, learn to walk, drink water from little cups, work too hard, wear high tops, choose a location, supply a baseball team, rest after work Note: spectators line up along course,
	query is: play Suggestions: exercise fun human interaction, move a white pawn, improve physical fitness, move a white knight, plan strategy, take a break to drink water, lose to another team, make an opening move, remember my childhood, move at a fast pace Kept suggestion: improve physical fitness Kept suggestion: plan strategy Kept suggestion: take a break to drink water Kept suggestion: make an opening move

Figure 21: Example of turnover of stuck suggestion list from LifeNet.

The experience of shooting with LifeNet was frustrating due to the time lag between annotation and suggestions and, at times, the "stuck" state of the suggestions. However, during these shoots some of the suggestions did help me to reengage with the documentary subject. When I began shooting the second LifeNet marathon in Philadelphia I noticed I had domain fatigue; I felt distracted by a hardening of my expectations, or the feeling that some shots were redundant from the previous marathons, finding novelty during the shoot felt impossible. I was bored with marathons. Each shoot provides experience, even suggestions, for the following shoot. If I noticed something during the annotation shoot, it would look less surprising during the second one. An example of this is the electronic tracking devices that runners wear at most marathons. The devices look like wristwatches around the ankles of each of the runners. The devices are not very obvious, except in their ubiquity and at the end of the race when directly after finishing runners are usually asked by volunteers to return the devices. Noticing this detail at the first marathon, spurred my curiosity: "What are those devices for?" and "How do they work?" After a few marathons the devices are not surprising and may be taken for granted as something any viewer of a marathon would know. In this case unpacking tacit knowledge can help the user. The danger is in reminding the videographer of the obvious or unpacking to a level of detail that is in conflict with the expressive capabilities of the video medium. If the videographer is too concerned with detail she can miss the broader story, either focusing on one sequence or a set of details that are missing a connection to the larger story. LifeNet would be greatly helped by more knowledge and faster inference algorithms.

4.5.3 StoryNet



Annotation: "ais plants bulb"

Suggestions: waters a seed, waits until spring, waits one week, grows into a plant, blooms a flower, hears her cell phone ringing, puts plant in sunlight, clean up spilled dirt, remove gardening gloves, pick up pen

Figure 22: Example of a StoryNet suggestion

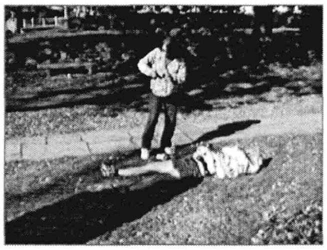
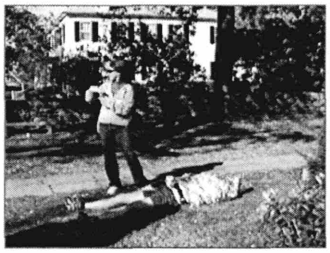
StoryNet was the system that the expert videographers found to deliver more useful suggestions. This could be because expert videographers learn though experience to predict not only the next possible event but many event steps into the future. This helps them to be ready for sudden and subtle changes in shooting opportunity. The videographers may have appreciated that each suggestion was as complete story event, instead of a single word or phrase that did not communicate an action. The seed shoot was constrained to shooting a short sequence in a controlled environment and the StoryNet resource was created to contain only story-scripts that were about seed shooting, even though some seed shooting story-scripts contained events that were not directly about seed shooting but occurred in one instance of a seed shooting experience. StoryNet could have been used differently as a suggestion device, presenting entire story cases as suggestions instead of individual events from each case matching the annotations. If this resource was used in a less controlled shoot it may not have been as effective. StoryNet worked well with a set of only 30 story-scripts. Acquisition of a larger and more diverse StoryNet would enable an ability to shift domains more broadly while shooting with the StoryNet resource. Alternatively, a huge number of story scripts might not be needed. It might be effective to quickly build small, targeted StoryNets about specific subjects just before a documentary shoot as a pre-production activity.

4.6 Recording Path Changes

A change in recording paths is a reported change in shot decisions and actions due to suggestions. Videographers reported on how suggestions changed their thinking about the documentary subject and the story they wanted to tell. Recording path changes demonstrate the videographer thinking about narrative possibility in a different way and acting on her new insights. A suggestion can directly change a recording path. The videographer takes a suggestion and tries to record a shot that depicts the concept in the suggestion. It can also indirectly change a recording path by instigating a new inference by the videographer. Sometimes a recording path change is successful. The videographer is able to get the shots needed to create the change in story direction. This section presents examples of recording path changes and discussions based on the videographer reporting. I have named the recording path examples to emphasize their unique narrative characteristics. The strongest evidence of reflection in the studies can be seen in these examples.

4.6.1 Thematic Reflection

A thematic reflection causes the videographer to shift from recording and observing actions to noticing events that demonstrate themes. This can be thought of as moving from capturing specific events that drive the story forward, plot element, to capturing strong motifs that are characteristic of the documentary subject. Some events in a marathon story are: lining up at the starting line, running to the front of the pack, and crossing the finish line. Themes present in a marathon are competition, overcoming adversity, and perseverance. The following example (Figure 23) demonstrates a recording path change during a marathon documentary. During recording at the finish line of the Falmouth marathon I was recording clips of runners getting medals, turning in the sensor tags which mark their finishing time, and people milling around in the finish line area. I recorded a clip of a runner getting a cup of water from a volunteer. ConceptNet suggestions presented were general marathon events, such as pass and start, but also states of the runners. I took two suggestions, both about the concept "tired" and decided to follow up on that theme. When I took the suggestions I realized that at the end of the marathon exhaustion is a strong theme, one worth portraying well in the documentary. The inference that I made in mind surprised me: tired + get tired = exhaustion. Of course the end of a marathon is about exhaustion! I now had a clear shot decision in mind and tried seven times to record a shot that demonstrated this theme. Finally, the seventh shot of a woman who could barely walk being carried to a seating area by her family communicated the theme.

	Annotation: “runner gets water” Suggestions: run marathon, win, start, pass, get tired, compete, run in marathon, tiredness Kept suggestion: get tired Kept suggestion: tiredness Shot suggestion: get tired
	Annotation: “exhausted” Suggestions: None Note: get better exhausted shot
	Annotation: “exhaustion” Suggestions: fatigue, run marathon, run out of steam, nothing, tiredness, sweat, blister, anger, prove person 's physical endurance
	Annotation: “exhaustion” Suggestions: fatigue, run marathon, run out of steam, nothing, tiredness, sweat, blister, anger, prove person 's physical endurance TS: tiredness TS: run out of steam TS: blister TS: prove person 's physical endurance Note: good prev shot inference Note: exhaustion is a strong theme present in the scene - follow it

	Annotation: "run out of steam" Suggestions: cooking, cramp, run after ball, injury, sweat, go for jog TS: cramp TS: injury Note: Best exhaustion shot!
	Annotation: "leaving the race area" Suggestions: race, area, leaving, start, bicycle, ride bike, ride horse, horse, social animal, clipper

Figure 23: Thematic Reflection

Taking multiple shots to portray a theme produced a larger variety of actions that demonstrated the theme. The actions contained in the seven video clips include runners lying down, a runner sitting on a chair with head in his hands, a runner sleeping while lying in the park, a runner being helped to walk, and a runner limping to a staircase and using a railing to slowly lower himself to sit on a step. Being satisfied with the seventh exhaustion shot ("Best exhaustion shot!" was a note taken during the shoot) I moved on to other observations but returned later in the day to that theme after the race was completely over and the runners were sitting at the awards ceremony. This was not a result of a suggestion at that time but my own storytelling knowledge that themes often persist over long periods of time during a documentary.

4.6.2 Script Extension Reflection

Expert videographers can be accomplished predictors of the events that might unfold during a documentary subject. They know the set of events that must be captured. They have an internal shot list. In this example, the videographer kept the suggestion 'get dirty' which was generated by the ConceptNet function "used for." The common sense of this suggestion is that a person could use "dirt" to "get dirty." The concept "get dirty" prompted the videographer to make a causal inference. If the gardener 'gets dirty" she will need to clean up. The videographer decided to extend the boundary of her shoot, originally reported to be after the seed was planted, to include the subsequent cleaning up. After this realization the videographer extend the set of shot actions to include a shot of the gardener cleaning up.

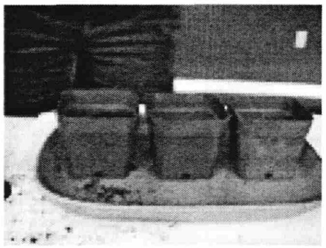
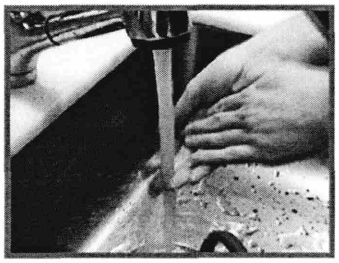
	Annotation: ais gets three pots and puts dirt in Suggestions: have bath, be dirty, become more clean, get clean, grow vegetable, sweat, get dirty, bathe, clean, cleaning, clean clothing, wash clothes, take shower, wash, do housework Kept Suggestions: get dirty
	Annotation: aisling put dirt in three pots Suggestions: grow, buy, break, use light up our night, make light, plant in garden, contain filament, contain, bread, dollar, future, play frisbee, bench, palm tree Kept Suggestions: None
	Annotation: "ais gets hands dirty" Suggestions: victim, snake, do not drink, tie shoelace, love, snack, corinthian column, not drink, think about food, pass, eat cereal, tease, blow, live and hunt in meadow, raise, pick Kept Suggestions: None
	Annotation: "pots get water and ais washes hands" Suggestions: None

Figure 24: Script extension recording path change. The videographer extends the activity of planting the seed to include an additional shot of cleaning up to the end of the sequence.

Figure 24 shows the sequence of shots and trace of suggestion activity. The suggestion marked in red is the one that instigated the recording path change. Notice that although the videographer kept the suggestion "get dirty," nine out of fifteen suggestions served after the first clip either contain the word "clean" or are about activities associated with cleaning. In post-shooting conversations, the videographer did not report being influenced by suggestions about cleaning, only by the suggestion "get dirty." The recording path change was instigated during the reflection on the first kept suggestion and then enacted in the fourth clip (red border) that shows the gardener washing her hands. Although this videographer did not report following up specifically on the "get dirty," but on her own internal inference of cleaning up, the third clip in the series is annotated "ais gets hands dirty." Script extension suggestions happen when the recording path is extended to include events that temporally follow the videographer's expected story or sequence ending.

4.6.3 Component reflection

This novice videographer took the same suggestion – "get dirty" – as the expert in the script extension example, but to different effect. This videographer reported that the suggestion caused her to "change orientation" during the shoot and focus on other objects and elements of planting other than the seed.

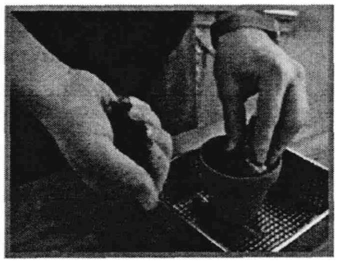
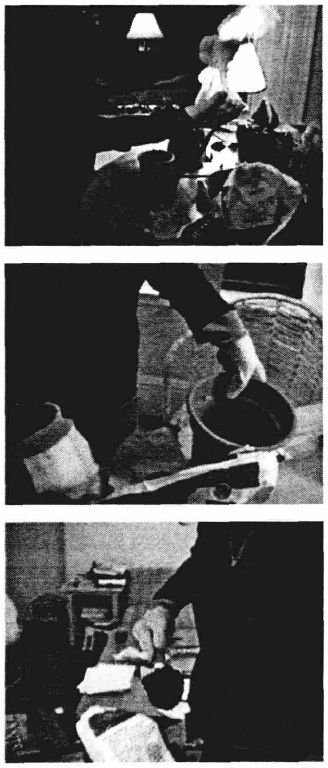
	Annotation: "putting dirt in pot" Suggestions: have bath, be dirty, become more clean, wash clothes, grow, get clean, grow vegetable, bathe, cleaning, clean clothing, get dirty, take shower, raise, cook food, cover, stir Kept Suggestion: get dirty Kept Suggestion: grow Shot Suggestion: get dirty
	Annotation: "putting seed in dirt" Suggestions: have bath, be dirty, produce, become more clean, grow, get clean, grow vegetable, bathe, cleaning, clean clothing, wash clothes, get dirty, take shower, spread, wash dirt from between person 's toe, stay healthy
	Annotation: "watering seed" Suggestions: grow, produce, sprout, pick, buy, spread, carry, become food, become tree, hold, swallow, grow flower, weed, danger, lie, plant in garden, factory, wind instrument Kept Suggestion: plant in garden

Figure 25: Component reflection example. The videographer breaks down the elements of the action to show more of the details of the sequence.

She reported that the suggestions helped broaden her ideas for shots, to record not only "planting a seed" but also other elements such as hands and dirt. The shot of the hands was the one in this collection that the videographer liked the most and "felt more like a director after making a decision based on a suggestion."

4.6.4 Focus reflection

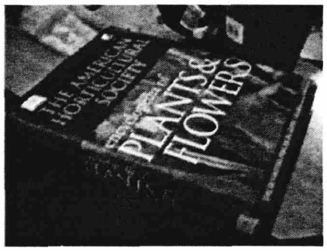
During a focus reflection the videographer decides to zoom in on one or more details of a scene in response to suggestions. The suggestions in this example are from StoryNet. This videographer innovated in his annotations. He took multiple shots for the annotation "sense of filling pot multiple clips" This novice videographer reported that he did shot suggestion directly, but the suggestion of "make hole in soil" made him think about getting a close-up of the seed going in the hole. The next shot after the suggestion was taken shows a close-up of the seed going into the hole. His goal for the shot was accomplished.

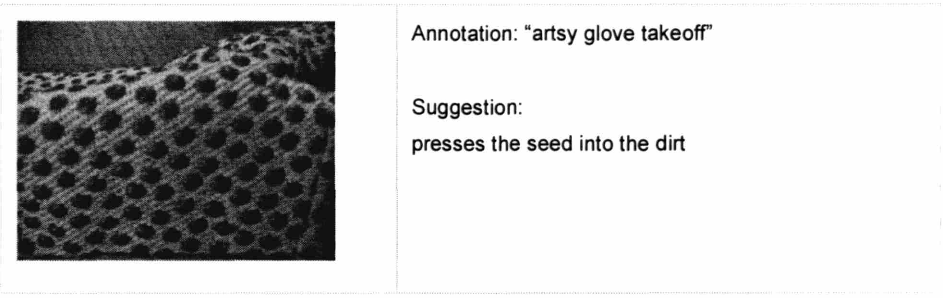


Annotation: "sense of filling pot multiple clips"

Suggestions:
opens bag of potting soil,puts potting soil in each compartment of a seed tray, makes small hole in the soil,
walk over to seed tray

Kept suggestion: walk over to seed tray
Kept suggestion: makes a small hole in the soil

	Annotation: “solo shot seed into pot” Suggestions: makes hole in the soil, waters a seed, walks away, grows into a seedling, finds seed under a tree, picks up a seed, puts the seed in her pocket, walks home, shows the seed to a friend, walks in the park, opens a bag of potting soil, puts the potting soil in each compartment of a seed tray, covers the seed with dirt, waters the seed, waits until spring, grows into a plant, takes one seed out of her pocket, presses the seed into the dirt, hears the telephone ring, hears her cell phone ring, create a hole in the dirt, pat the soil, drops seed, picks up seed, plant a seed, remove gardening gloves, pick up a pen, pick up spade, walk to sink, water the seed, empty the watering can Kept suggestion: covers the seed with dirt
	No Annotation
	Annotation: “finishing shot montage pan” Suggestions: remove gardening gloves TS: remove gardening gloves
	No annotation



Annotation: "artsy glove takeoff"

Suggestion:
presses the seed into the dirt

Figure 26: Focus example. The videographer breaks down the elements of the action to show more of the details

The shot action motivated by "makes a small hole in the soil" was recorded three shots after the suggestion was taken. This videographer also continued shooting more close-ups of different elements and actions in the rest of the shots of his collection ending with a very extreme close-up of the texture on the gardening gloves. The taking of close-ups regardless of content is a formal decision about how to shoot a subject. This videographer made a formal decision from a commonsense relationship between two elements. In other words, sometimes a commonsense inference makes a videographer frame an object or action more tightly. Two out of four videographers described moving closer to the subject or gathering details as a response to the taken commonsense suggestions. This videographer remarked about how the system encouraged reflection by saying "Even though it seems annoying to take time out after a shot to log it in there is a nice time when you can pull back and think about how you might do it differently, rather than just shooting, shooting, shooting." This videographer also thought of the system as pedagogy for teaching the skill of documentary videography, particularly for helping the videographer pay attention to building narrative possibility in his mind as he observed and collected video material. He recalled a lecture by a well-known photographer, "You can't wait for something to happen it has to part of your story arc to work. You have to have something in mind to have an angle on whatever you are taking pictures of. Even if you are refuting the [mindful camera] suggestions you are still thinking about the situation."

4.6.5 Script-detour reflection

The script detour reflection is when there is a detour from recording the events of the documentary subject to include secondary stories that branch from, and then return to, the main story progression. This expert videographer took the first shot of her documentary, annotated it, and received suggestions from StoryNet. She reported that when she saw the suggestion "hears her cell phone ringing" she reacted as if it were an incorrect suggestion because it was not about seed planting.

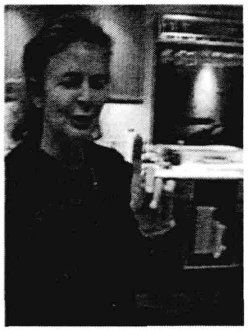
	"planting bulbs for valentine's day" Suggestions: waters a seed, waits until spring, waits one week, grows into a plant, blooms a flower, hears her cell phone ring, waters the seed, put plant in sunlight, clean up some spilled dirt, remove gardening gloves, pick up a pen Kept Suggestion: clean up spilled dirt Kept Suggestion: remove gardening gloves Kept Suggestion: hears her cell phone ringing
	Annotation:" she pats the soil in the pots" Suggestions: opens a bag of potting soil, puts potting soil in each compartment of a seed tray, makes a small hole in the soil, thoroughly water dirt, puts seed in hole, put soil in flowerpot Kept Suggestion: makes a small hole in the soil Kept Suggestion: thoroughly water dirt Shot Suggestion: hears her cell phone ringing Shot Suggestion: makes a small hole in the soil
	Annotation: "making a hole for the bulb" Suggestions: puts the seed in the hole, covers the seed with dirt, pat the soil, put seed in hole, pick up a spade, water the seed Kept Suggestion: pat the soil Shot suggestion: pat the soil

Figure 27: Script-detour reflection

She kept the suggestion but remarked later that she didn't consider it part of a seed story. She kept the suggestion anyway as a curiosity. She was surprised when a moment later the cell phone rang and the gardener went in another room to answer it. The videographer recorded a shot of the gardener returning from answering the phone but missed the shot of her leaving to answer the phone, resulting in one half of a pair the concepts needed to show a complete alternative story thread, leaving to and returning from answering the phone, that would branch from the seed shoot and return in two shots. The videographer remarked that she made a rule from that experience to "shoot all the interruptions." The interruption could be considered an unexpected event that does not fit the most expected script, but is plausible and represented in one of the StoryNet story-scripts. This is an example of a recording path change that was not successfully implemented because the videographer did not consider the suggestion a likely event and the window of time in which the videographer needed to act was very shot.

5 Recommendations

5.1 Story suggestion generation

There is room for improvement in the way the system generates story suggestions. As I mentioned earlier, the goal of the mindful camera software was not to attempt full story understanding. The existing event generation technique can be built upon to better server the videographer. Suggestions would be more relevant if a filtering technique was applied to the results. If the system understood more about event duration, frequency, and temporal position in a story the suggestions could be given with more precision and relevance to the opportunity for recording. The gun firing at the start of a marathon only happens once at the beginning of the race and its duration is extremely short. It does the videographer little good to get this shot suggestion at the end of the race. It would be best to suggest it early in the shoot and let the videographer know what events are likely to precede it; for example, runners would be lining up at the starting line before the starting gun is fired. The start of the race is important event to capture. Filtering of suggestions could also take into account the most expected events in a story by citing the most frequent events in a set of story-scripts. If eighty percent of story-scripts mentioning a marathon also mention firing a gun at the starting line it should be suggested and captured. This kind of filtering could enable the system to be more effective in helping the videographer capture both the canonical and the unusual events or details during a documentary shoot. The classification of story events into usual and unusual could also be aided by LifeNet's probabilistic inference. Ideally, the system could begin to understand the expectation breach during a story and encourage the videographer to create stories that include this narrative characteristic. Currently, the system does not bind actions to any particular action. The videographer has to follow a particular actor or decide if a character is a major or minor one. The design of the studies did not address this issue because the marathon shoot was recorded as a documentary depicting a process. The main character of the documentary was the marathon itself. The engineering of a simplified scene for the seed shoots narrowed the documentary to include only one actor, the gardener. As was mentioned in the Section 4.3.1, a logical next step for this work is to identify larger story patterns as the documentary is being constructed. A suggestion could then be based on a broader view of how an actor's plans progress toward a goal or how configurations of events can express story themes instead of only the immediate implications of the last captured shot.

5.2 Method and System

Presently, the mindful documentary methodology enforces a shooting technique on the videographer; the videographer takes a shot, annotates, waits for suggestions, and then reviews them. This imposition of a structured shooting practice has advantages and disadvantages. An advantage is encouraging the videographer to step back and think after each shot, a practice most effective for novice videographers. The disadvantage is that the videographer's attention is taken from the world. One videographer reported "Loss of control of shooting schedule was the most frustrating part – chopping up the shoot and waiting for the suggestions, not just the waiting time of machine processing the suggestions but the human activity of processing the suggestions." The videographer is not in a position to observe, but is instead focused on the machine and interface. The system could be improved to let the videographer annotate multiple clips at one time, or ask to explicitly for suggestions, although I suspect this would not be as effective in encouraging reflection. When the system has better story understanding it might be able to offer suggestions at more appropriate story junctures, such as when a new character is mentioned, a domain change occurs, or an unlikely event has been annotated. A model of the videographer's position in the story would also help the system not only know when it is useful to intervene, but also with suggestion filtering. If the system knows the videographer is at the beginning of capturing a new story idea just after a recording path change, the system would allow the videographer to explore the new thread over a few shots and annotate the shots collectively. Voice annotations with speech recognition would also help with ease of use. The videographer could keep her eyes focused on the world and the subject she is recording instead of on the mindful camera interface.

5.3 Common sense

The commonsense reasoning tools used in the mindful documentary work are still young. More commonsense knowledge is needed to provide breadth and depth for all possible documentary subjects. The commonsense used in the mindful camera is collected from a large community of users, not any individual videographer. The commonsense of the system inherits the biases of its authors but also the diversity of their experience. Therefore, the commonsense can provide an inference model of a diverse audience, enabling suggestions that the videographer might not have considered. This helps the videographer keep mindful during capture. Mindful documentary, unlike other computational tutors, takes into account how thinking happens, not just how to deliver knowledge to solve a problem in a specific domain. When we exchange knowledge with another person we are aware of what they know and how they think. We decide what we might want to learn from them and whether we trust them as a knowledge source. During a few of the marathon shoots when suggestions were returned I could guess who might have submitted the knowledge. When the word "test" was submitted as an annotation referring to the physical test of running a marathon the returned some of the returned suggestions were about exams a person would taken in school. I guessed it was the commonsense of a high school or college student. A knowledge source can be a friend, mentor, a stranger or, as in the case of Open Mind Common Sense, a collection of ideas from people who are anonymous. The more diverse the knowledge, the more story possibilities are supported in the partnership between the videographer and the system. Videographers may want to define the commonsense knowledgebase that is used for suggestions to embody a specific culture, gender, age, or even a single person. This would enable videographers to observe from different perspectives than what they already know. Alternatively, if videographers were willing to spend significant time contributing commonsense knowledge, they could use their own sense to drive suggestions.

The mindful documentary method in itself is a knowledge acquisition enterprise. The studies yielded sets of annotated footage. As the system is used, it would be fruitful to automatically update the commonsense knowledge base with information learned on any documentary shoot. Videographers could then pass their collected knowledge along to another person who would later attempt to shoot the same documentary subject. In addition, accepting and ignoring suggestions could be tracked in order for the system to learn which suggestions have been highly successful.

The commonsense tools were used individually during the documentary shoots. Each offers a distinct way of thinking that can influence story construction. It would be beneficial if the system could know when each type of thinking was most useful to the videographer. As in life, sometimes we need to free our minds from a problem and think associatively in order to find another path to a solution; ConceptNet makes associative inferences. Sometimes we need to create a plan by knowing the order of next steps while also being on the lookout for unlikely events that might occur; LifeNet can assist with this ways of thinking. At other times, we recall a similar situation and understand the similarities and differences to our current situation – we think using stories. StoryNet inference emulates, in a simplified way, this kind of thinking. Cameras today have optical zooming capabilities that enable a videographer to alter her view of the world and capture the new view. A zooming capability that uses commonsense reasoning instead of optics would let the videographer adjust the thinking of the system dynamically from associative, to temporal or story-based reasoning during documentary creation.[15]

6 Conclusion

This thesis begins with a hypothetical model of the videographer's decision-making process during documentary construction. The main component of the mindful documentary model is called a *construction cycle,* which consists of a *shot decision*, a *shot action,* and *reflection window.* The videographer observes the world; decides what to record; records a shot; and then reflects on the influence of the shot on her creation of a documentary story. The reflection window indicates the time between when the videographer has captured an image and when the videographer makes her next decision to capture a shot. This reflection time can vary in length, but we hypothesize that the videographer always goes through the following assessment regarding the relevance of the recently captured shot to the *narrative possibility* of a video clip collection:

- Was the intended shot captured?
- What is the shot about?
- What are the immediate inferences one would make about the concept captured in the shot?
- How does the shot relate to other shots in the collection?
- How does this influence possibility for the next shot?

I believe this reflection window provides an opportune time for an intelligent system to intervene as a partner, assisting the videographer during the process of capture, in reflecting on narrative possibilities for her video collection. The decision to have the system intervene at all can has two main merits: first, it could allow the videographer to annotate shots in the field (circumventing the problem, common today, where no content annotation can be applied until much later); and second, it could allow a machine to make suggestions about what to capture next, thus acting – in some sense – like a mentor or a friend to help the videographer reason about the current situation.

The hypothesis of the mindful documentary work is that a reflective partnership between a videographer and a camera with commonsense reasoning abilities can aid story construction during documentary videography by providing useful suggestions to help the videographer identify narrative possibility during the documentary creation. In order to test this hypothesis, I created the mindful camera system, a camera with the commonsense reasoning abilities that can act by suggesting shot ideas to the videographer during capture. I used two existing commonsense resources, ConceptNet and LifeNet, and co-developed a new commonsense resource of story-scripts called StoryNet. Each offers a different way of creating story inferences. These resources were used to present the videographer with suggestions about next events or the contextual elements of a documentary subject based on the videographer's text annotation of a shot. A set of field trials (videographers creating documentaries) were performed to test our hypothesis. These trials showed that commonsense suggestions can help a videographer get "unstuck" or think about an alternative way to tell the story of the documentary subject.

Suggestions are reflections on opportunity. They encourage the videographer to reflect on the immediate and potential long-term relevance of a recorded clip. Over the long term, during the shooting of one or many documentaries, the suggestions can help the novice videographer develop what Jean Rouch called "greis" or grace. It is the ability to make decisions during capture that leads to a seemingly effortless recording path; the recording path is effortless because the videographer can see clearly how the opportunities for content collection accord with the story models they are building during documentary construction. Novices found the system to be a unique kind of teacher that encouraged them to step back and think about their shot decisions and expand their thinking about other narrative possibilities. Thus the mindful camera interaction and system can be considered a pedagogical framework for teaching documentary videography. The documentary videographers each reported a change in their minds and in their recording paths due to reflection on commonsense suggestions. We cannot say if these changes made the videographers more productive or helped them to create a better story, but their reports indicate a level of engagement and explanation of their story thinking that is favorable.

The mindful documentary work has implications outside of the documentary domain toward broader collaborative storytelling between computers and machines. The story creator is encouraged to be an active constructor rather than a passive observer of story. In both the real and virtual worlds, we are presented with a vast amount of information; we forge our way through it by creating stories that crystallize our experiences. By understanding and creating models of how people think and act to create stories, we can play with this process and challenge ourselves to see the world using different knowledge collections and ways of thinking, perhaps even borrowed from other people. In the future, we will have more detailed models of creativity, of how people think when they solve puzzles, create artworks or tell the story of their day. The mindful documentary work is a step toward this, and striving toward intelligent machines that can impact our creative practices and serve as imaginative storytelling partners.

Glossary

Annotation: Annotation is the act of creating data that describes or represents a media element.

Construction: In the context of this thesis, construction is defined by the epistemological term "constructionism." Constructionism is a concept developed by Seymour Papert of M.I.T. It extends Constructivist theory, which states that all children construct their own knowledge. Constructionism expands on this concept by claiming that people have many of their best learning experiences when they are actively engaged in making a product or artifact that is meaningful to them or others (Papert, 1991). In the constructionist experience, the environment responds to the builder, giving her feedback during the process of learning. An example of constructionist learning is a child building a scale out of Lego bricks to learn about weight, balance and gravity -- building a flexible object to learn about a concept or idea. Constructionism refers to the creation of all types of artifacts, not only physical objects but also images and stories.

Construction Cycle: The construction is a model of how the videographer thinks and acts during the creation of a documentary. It is composed of a shot action, a reflection window, and a shot decision.

Inference guide: Inference guides are the result of the decisions a documentary videographer during construction. It can be thought of as a map that the viewer can use to generate inferences about the content.

Information track: Information track is a data track that accompanies a video track that can be used as the input to a computational process. The term was coined by Nicholas Negroponte in his work on the impact of the videodisk on filmmaking. The absolute data is the unchanging information that will always be true, such as the location depicted in a clip, whereas the relative information can be thought of as supporting the needs of the filmmaker, such as pointers to other sources of information. It can be thought of as storage repository for different types of annotation associated with video.

Mindful Documentary: My model of how the videographer thinks during documentary construction. The model consists of a construction cycle in which the videographer cyclically takes a shot, reflects on the implications of the shot and then decides the best shot to attempt next.

Mindful camera: The system I built that has commonsense reasoning abilities designed to help the videographer reflect on story possibilities during the recording of a video clip collection. The software provides a way for the videographer to annotate clips in the field. Annotations are used to generate shot suggestions, which the videographer can keep or ignore.

Narrative potential: Narrative potential of a single video clip is its relations to other clips in a video collection. The narrative potential of a video collection is the possible stories that one could build from the video collection.

Recording Path: The recording path is the sequence of shot decisions a videographer makes during documentary production. It can be thought of as the set decision-points in the world that denote where and when the videographer decided then recorded a shot. The recording path can be thought of as the total of all the individual construction cycles of a documentary shoot.

Reflection Window: A model of the process of thinking that the videographer engages in after each collected shot. It purpose is the assessment of the shot with regard to the videographer's goal for the individual shot, a description of what the shot depicts and inference production to understand the possible implications of the shot itself and its relationship to other clips in the collection. The videographer also assesses and decides what is likely to happen next in the world, integrating the collected information into his or her expectation of what might happen next in the world.

Shot action: The action of recording a video clip.

Shot decision: The videogrpaher's goal for the capture of a clip based on their assessment of their video collection and the possibilities for capture.

Story: Webster's Dictionary defines story as "the telling of a happening or connected series of happenings, whether true or fictitious; an account; a narration." Within the scope of this thesis *story* will be a general term used to describe a *narrative.* In Webster's, the definitions of story and narrative are interchangeable.

Story-script: A mental model of a story experience. Its main feature is a temporal ordering of events but can also contain other relationships between story elements such as event dependencies, actor goals, event dependencies, etc. Story-script is also the term we use for the computational story representation of stories in StoryNet.

Footnotes

1. Robert Coles (1998) traces the etymology of the word documentary back to its Latin root "docere" which means to teach. The documentarian has the responsibility of creating a document from which the audience can learn or use to support their understanding of a documentary subject.

2. Paulo Friere (1970) discusses action and reflection as a process for social change by a balance of thought and action. The application here to documentary practice is a nod to John Grierson (1947) coining the term documentary as a human document that had the ability to catalyze social change.

3. Phrase borrowed from Erik Mueller's description of how people understand text (Mueller, 2002).

4. A "defining moment" is a term used in observational documentary. It is the point in the plot at which problems or tensions will be resolved.

5. Glorianna Davenport, Richard Leacock and members of the MIT Film Video department provide a description of a documentary. It is meant to emphasize the intimacy of shooting with a portable camera and the role of the filmmaker as silent observer.

6. Visit the website of the Commonsense Computing Group at the Media Lab for more information. http://commonsense.media.mit.edu

7. Minsky (2005) discusses "ways of thinking" as having more types than are present in the mindful camera system's associative, probabilistic and case-based ways people create inferences about everyday life.

8. Henry Lieberman uses this term as an interaction style for intelligent agent systems. The software agent's failure does not impede the user's experience. In some cases the failure is a cue for the user to offer some knowledge that will improve the system's performance.

9. From a discussion with Luc Courchesne about information available to the artist at all times as a resource or intervention into a creative process.

10. http://www.cyc.com

11. http://en.wikipedia.org

12. http://www.imdb.com

13. http://www.flickr.com

14. http://commonsense.media.mit.edu

15. Thanks to Walter Bender for the optical/knowledge zoom analogy.

References

Anderson, J.R., Boyle, C.F., & Yost, G. (1985). The geometry tutor. In *Proceedings of the International Joint Conference on Artificial Intelligence* (pp. 1-5), Los Angeles, CA.

Barnouw, E. (1993). *Documentary: a history of the non-fiction film*. New York: Oxford University Press.

Barry, B. (2003). The Mindful Camera: Common Sense for Documentary Videography. In *Proceedings of ACM Multimedia Conference*. Berkeley, California, USA.

Bartlett, F. (1932). *Remembering: a study in experimental and social psychology*. Cambridge, England: The University Press.

Bloch, G.R. (1988). From Concepts to Film Sequences. RIAO '88. pp. 761-77.

Brooks, K. (1999). Metalinear Cinematic Narrative: Theory, Process, and Tool. Massachusetts Institute of Technology. Ph.D. Thesis.

Bruner, J. (1991). The Narrative Construction of Reality. *Critical Inquiry, 18* (Autumn 1991), pp. 1-21.

Burke, R. & Kass, A. (1995). Supporting Learning through Active Retrieval of Video Stories. *Expert Systems with Applications. 9*(5).

Chua, T. & Ruan, L. (1995). A Video Retrieval and Sequencing System. *ACM Transaction of Information Systems. 13*(4). October 1995, pp. 373-407.

Clark, H. H. (1977). Bridging. In P.N. Johnson-Laird and P.C. Wason (Eds.), *Thinking: Readings in Cognitive Science*. Cambridge: Cambridge University Press.

Coles, Robert (1997) *Doing Documentary Work*. New York: Oxford University Press.

Davenport, G. (1987). New Orleans in Transition, 1983-1987: The interactive Delivery of a Cinematic Case Study. In Proceedings of *The International Congress for Design and Planning Theory*. Boston, Ma. May 1987.

Davenport, G., Smith, T.A., & Pincever, N. (1991). Cinematic Primitives for Multimedia. IEEE Computer Graphics & Applications, vol.11, no.4, July 1991, pp.67-74.

Davenport, G. & Murtaugh, M. (1997). Automatist storyteller systems and the shifting sands of story. *IBM Systems Journal*, vol.36, no.3, 1997, pp. 446-56.

Davenport, G. (1980). Diaries 1979-1980. Unpublished.

Davis, M. (1995). Media Streams: Representing Video for Retrieval and Repurposing. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1995.

Dewey, J. (1933). *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*. New York: D.C. Heath and Company.

Dewey, J. (1963). *Experience and education.* New York: Collier Books.

Dorai, C. & Venkatesh, S. (2001). Computational Media Aesthetics: Finding Meaning Beautiful! ACM article. *IEEE Multimedia, The Media Impact Column. 8*(4). pp 10-12.

Drew, R. (1960). *Primary*. Drew Associates.

Eagle, N. & Singh, P. (2004). Context sensing using speech and common sense. In *Proceedings of the NAACL/HLT 2004 workshop on Higher-Level Linguistic and Other Knowledge for Automatic Speech Processing.*

Fellbaum, C. (Ed.). (1998). *WordNet: An electronic lexical database.* Cambridge, MA: MIT Press.

Freire, P. (1970). *Pedagogy of the Oppressed.* New York: Herder and Herder.

Gordon, A.S. (2001) Browsing Image Collections with Representations of Commonsense Activities. *Journal of the American Society for Information Science and Technology*, 52(11):925-929..

Grierson, J. (1947). *Grierson on documentary.* New York: Harcourt, Brace.

Iqbal, A., Oppermann, R., Patel, A. & Kinshuk (1999). A Classification of Evaluation Methods for Intelligent Tutoring Systems. In U. Arend, E. Eberleh & K. Pitschke (Eds.) *Software Ergonomie '99 - Design von Informationswelten*, Leipzig: B. G. Teubner Stuttgart, 169-181.

Johnson, C. (1998). *Syntactic and semantic principles of FrameNet annotation.* University of California, Berkeley.

Kearney, Richard. (2002). *On Stories.* New York: Routledge.

Kankanhalli, M. and Chua, T. (2000). Video Modeling Using Strata-based Annotation. IEEE Multimedia. 7(1) p. 68-74.

Leacock, R. (2005). (Personal communication, March 25, 2005)

Lehnert, W., Dyer, M., Johnson, P., Yang, C.J. & Harley, S. (1983). BORIS – An Experiment in In-Depth Understanding of Narratives. *Artificial Intelligence.* 20(1) p. 15-62.

Lenat, D. (1995). CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11), 33-38.

Henry Lieberman, Hugo Liu, Push Singh, and Barbara Barry (2004). Beating common sense into interactive applications. *AI Magazine*, Winter 2004, 25(4):63-76. AAAI Press

Lin, X., Hmelo, C., Kinzer, C. K., & Secules, T. J (1999). Designing technology to support reflection, *Educational Technology Research & Development*, pp. 43-62.

Lippman, A. (1980). Movie Maps: An application of the Optical Videodisk to Computer Graphics. SIGGRAPH '80.

Lin, X.D. & Lehman, J. (1999). Supporting learning of variable control: On the importance of making students' thinking explicit. *Journal of Research in Science Teaching. 36*(7):837-858.

Liu, H. (2002). Semantic Understanding and Commonsense Reasoning in an Adaptive Photo Agent, Masters Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA.

Hugo Liu, Push Singh. (2002). MAKEBELIEVE: Using Commonsense to Generate Stories. *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, AAAI 2002, July 28 - August 1, 2002, Edmonton, Alberta, Canada. AAAI Press, 2002, pp. 957-958.

Livo, N. & Ritz, S. (1986). *Storytelling: Process and Practice.* Englewood, Colorado: Libraries Unlimited.

Maes, P. (1990) Situated Agents Can Have Goals. *Journal for Robotics and Autonomous Systems*, 6(1):49-70.

Mateas, M., Vanouse, P., & Domike, S. (2000). Generation of Ideologically-Biased Historical Documentaries. In *Proceedings of AAAI 2000.* (pp. 236-242). Austin, TX, 2000.

McCarthy, J. (1959). Programs with Common Sense. In *Proceedings of the Teddington Conference on the Mechanization of Thought Processes.* (pp. 75-91). London, England: His Majesty's Stationary Office.

McCarthy, J. Minsky, M., Sloman, A., Gong, L., Lau, T., Morgenstern, L., Mueller, E., Riecken, D., Singh, M. & Singh, P. (2002). An architecture of diversity of commonsense reasoning. *IBM Systems Journal*, 41(3):530-539

Members of the Human Interface Laboratory (1989). Demonstration of HITS 1.0: The Human Interface Tool Suite, MCC Technical Report ACT-HI-116-89-P. March, 1989.

Menand, L. (1999). *The Metaphysical Club.* New York: Farrar, Straus and Giroux.

Minsky, M. (2005). *The Emotion Machine.* New York: Simon and Schuster.

Minsky, M. (2000). Commonsense-based Interfaces. *Communications of the ACM,* ACM 43(8), pp. 67-73.

Minsky, M. (1974). *A framework for representing knowledge* (AI Laboratory Memo 306). Artificial Intelligence Laboratory, Massachusetts Institute of Technology

Morgan, Bo (2004). LifeNet belief propogation. Unpublished article. Retrieved from http://web.media.mit.edu/~neptune/lifenet_bp_draft.pdf

Mueller, E. T. (1998). *Natural language processing with ThoughtTreasure.* New York: Signiform.

Mueller, E. T. (1999). A database and lexicon of scripts for ThoughtTreasure. CoRR cs.AI/0003004: (2000) http://arxiv.org/abs/cs.AI/0003004.

Mueller, Erik T. (2002). Story understanding. In Lynn Nadel (Ed.), *Encyclopedia of Cognitive Science.* London: Nature Publishing Group.

Murtaugh, M. (1996). The Automatist Storytelling System: Putting the Editor's Knowledge in Software. MIT Master's Thesis.

Nack, F. & Putz, W. (2001) Designing Annotation Before It's Needed. In *Proceedings of the 9th ACM International Conference on Multimedia,* pp. 251 - 260, Ottawa, Canada, Sept. 30 - Oct. 5, 2001.

Negroponte, N. (1979). The Impact of Optical Videodisks on Filmmaking. Unpublished article.

Nichols, B. (2001). *Introduction to Documentary.* Bloomington: Indiana University Press.

Papert, S. (1991). Situating Constructionism. In Edit Harel (Ed.), *Constructionism: research reports and essays*, 1985-1990. Norwood, NJ: Alex Publishing Corp.

Pea, R. (1993). The Collaborative Visualization Project. Communications of the ACM., 36(5)., pp. 66-30.

Pea, R., Mills, M. (2004). The Social and Technological Dimensions of Scaffolding and Related Theoretical Concepts for Learning, Education, and Human Activity, *The Journal of the Learning Sciences*, Lawrence Erlbaum Associates, Inc., 13(3), 423-451

Pinhanez, C. & Bobick, A. (1996). Approximate World Models: Incorporating Qualitative and Linguistic Information into Vision Systems. *Proceedings of the AAAI.* Portland, Oregon, pp. 1116-1123. August 1996.

Pincus, E. (1972). *Guide to filmmaking.* New York: New American Library.

Riloff, E. (1999). Information Extraction as a Stepping Stone toward Story Understanding, In *Computational Models of Reading and Understanding*, Ashwin Ram and Kenneth Moorman, eds., The MIT Press.

Roschelle, J., Pea, R. D., & Trigg, R. (1990). VideoNoter: A tool for exploratory video analysis. Institute for Research on Learning, Technical Report, No. 17.

Scardamailia, M., & Bereiter, C. (1996). Adaptation and understanding: a case for new cultures of schooling. In S. Vosniadou, E. De Corte, R. Glasser & H.. Mandl (Eds.), *International perspectives on the psychological foundations of technology-based learning environments* (pp. 149-165). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Schank, R. (1998). *Inside Multi-Media Case Based Instruction.* Hillsdale, NJ: Lawrence Erlbaum.

Schank, Roger (1995). *Tell Me a Story: Narrative and Intelligence.* Evanston, IL: Northwestern University Press.

Schank, R., & Abelson, R. (1977). *Scripts, plans, goals, and understanding.* Hillsdale, NJ: Lawrence Erlbaum.

Schön, D. (1983). *The Reflective Practitioner: How Professionals Think in Action.* New York: Basic Books.

Schön, D. (1987). *Educating the Reflective Practitioner.* San Francisco: Jossey-Bass.

Schroeder, C. (1987). Computerized Film Directing. Undergraduate Thesis, Massachusetts Institute of Technology.

Singh, P. (2002). The public acquisition of commonsense knowledge. In *Proceedings of AAAI Spring Symposium: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access*. Palo Alto, CA: AAAI.

Singh, P. & Barry, B. (2003) Collecting Common Sense Experiences. In *Proceedings of the Second International Conference on Knowledge Capture (K-CAP 2003)*. Sanibel Island, Florida, USA.

Singh, P., Barry, B. & Liu, H. (2004). Teaching machines about everyday life. *British Telecom Technology Journal*, *22*(4):227-240.

Smith, B. & Reiser, B. (1997). What should a wildebeest say? Interactive nature films for high school classrooms. In *Proceedings of the Fifth ACM International Conference on Multimedia*. Seattle, Washington. pp: 193 – 201.

Webster, N. (1983). *Webster's New Unabridged Dictionary*. New York: Simon and Schuster.

Wexelblat, A. & Maes, P. (1997). Issues for Software Agent UI. Retrieved from http://web.media.mit.edu/~wex/agent-ui-paper/agent-ui.htm. June 2004.

Appendix – Catalogs of Documentary Shoots

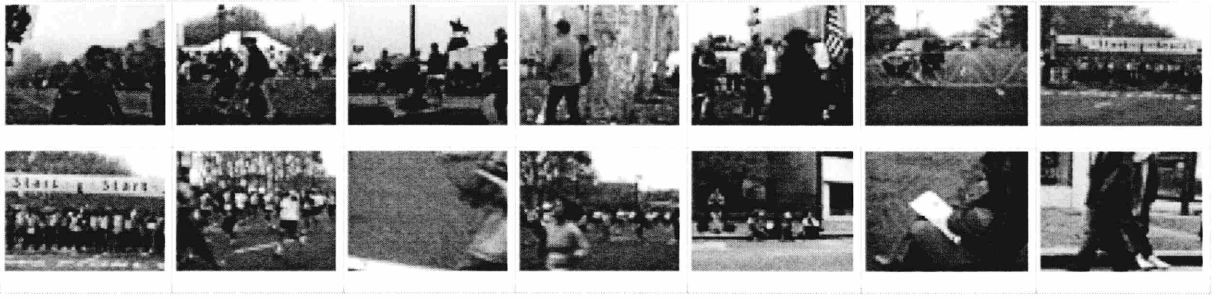
Video clips from each documentary collection created during the evaluations are shown here. To view the actual documentary clips please visit http://www.media.mit.edu/mf/mindfuldoc/shoots

Portland Marathon - Portland, ME, October 3, 2004 – annotation only

Falmouth Marathon - Falmouth, MA, October 31, 2004 – ConceptNet Commonsense suggestions

NYC Marathon – New York, NY, November 7, 2004 – LifeNet commonsense suggestions

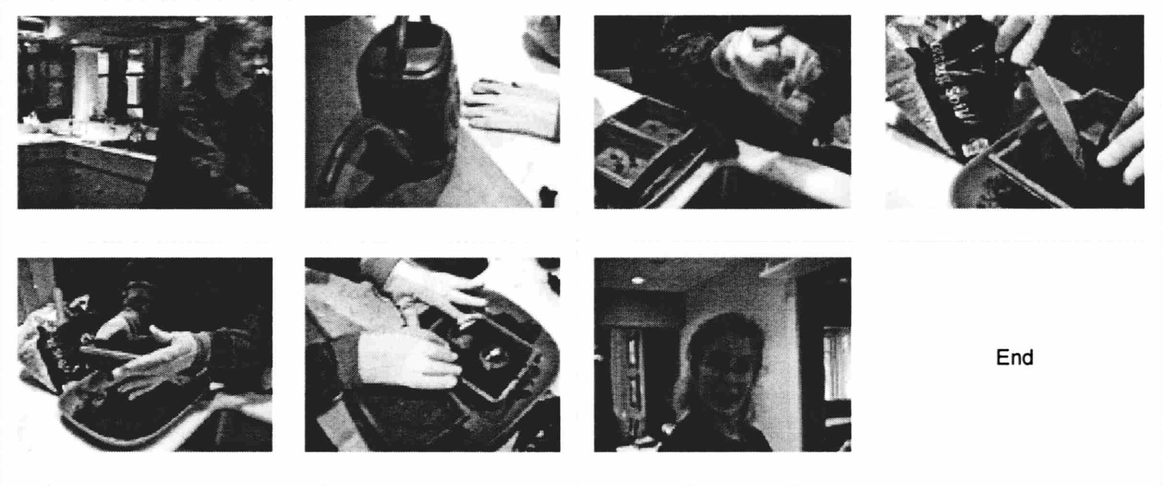
Philadelphia Marathon – Philadelphia, PA - November 21. 2004. LifeNet commonsense suggestions.



Seed shoots

Videographer 1 - Annotation only:

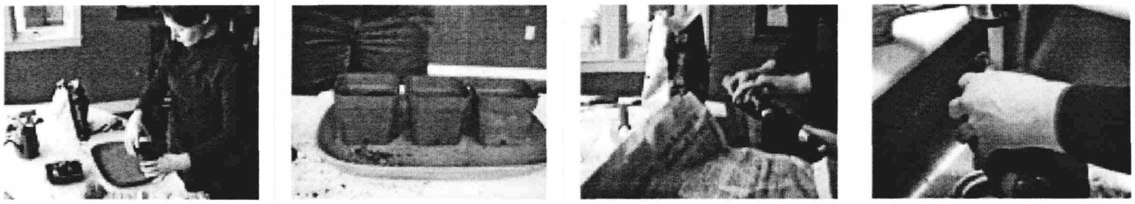
Videographer 2 – Annotation only:



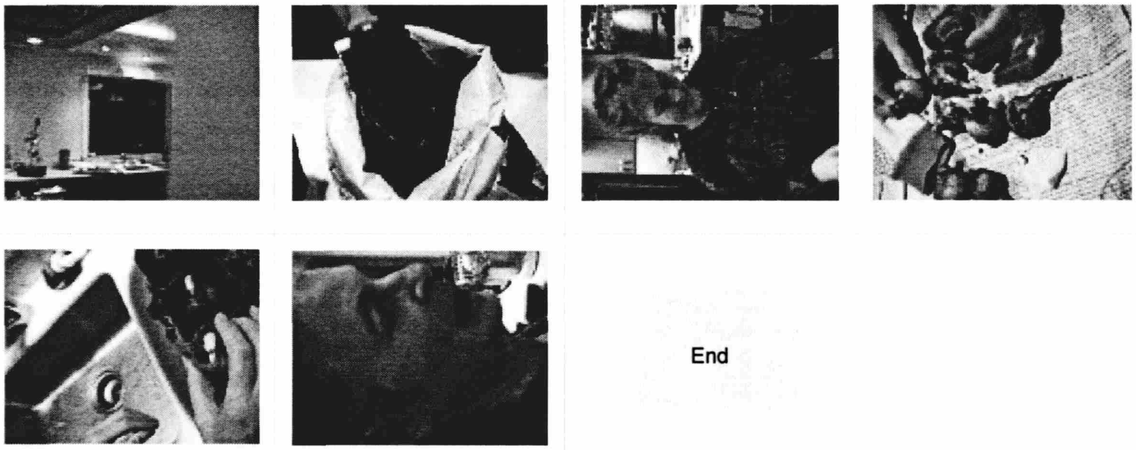
Videographer 3 – Annotation only:

Videographer 4 – Annotation only:

Videographer 1 – ConceptNet suggestions:

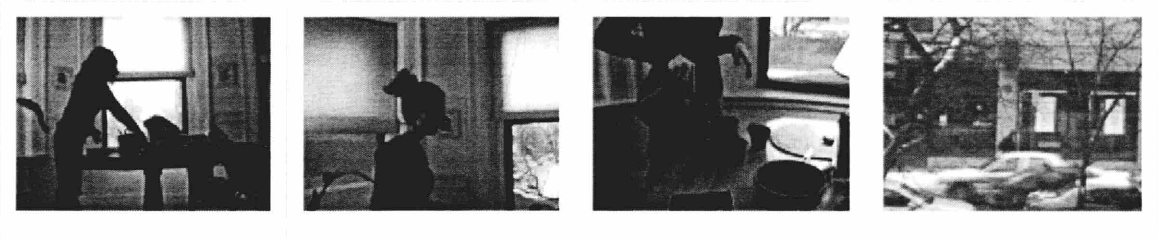


Videographer 2 – ConceptNet suggestions:



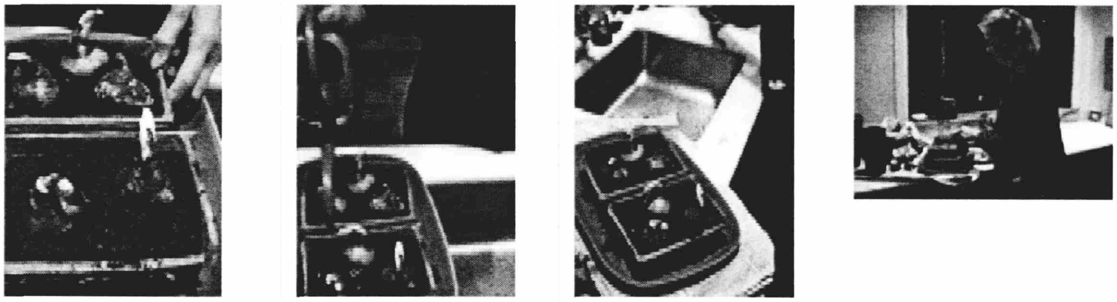
Videographer 3 – ConceptNet suggestions:

Videographer 4 – ConceptNet suggestions:

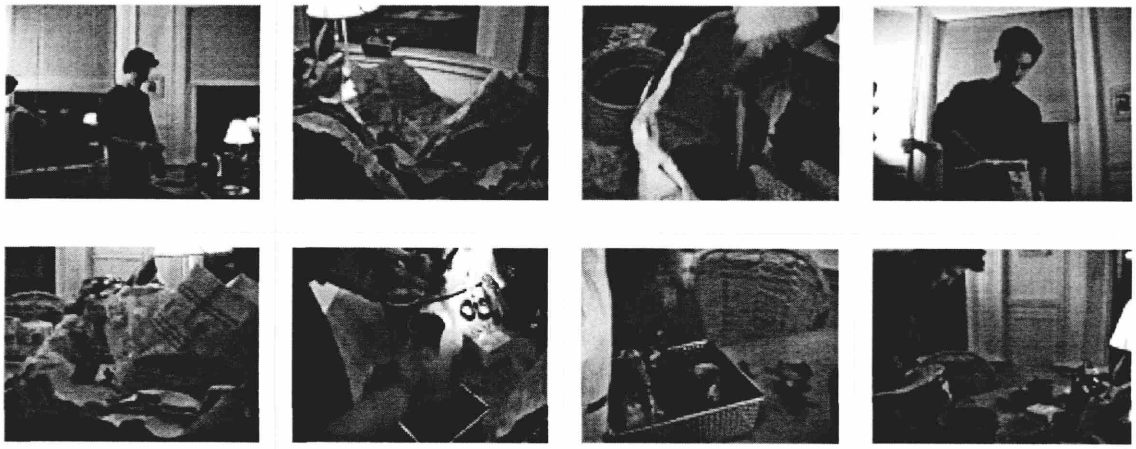


Videographer 1 – StoryNet suggestions:

Videographer 2 – StoryNet suggestions:



Videographer 3 – StoryNet suggestions:



Videographer 4 – StoryNet suggestions: