Distributed Narrative Extraction Using Imaging Sensor Networks

by
Mathew Laibowitz
12 April 2007

Proposal for Thesis Research in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

Massachusetts Institute of Technology
Department of Media Arts and Sciences

THESIS SUPERVISOR: ______________________

DR. JOSEPH A. PARADISO
PROFESSOR OF MEDIA ARTS AND SCIENCES
MIT MEDIA LAB

THESIS READER: ______________________

ALEX "SANDY" PENTLAND
PROFESSOR OF MEDIA ARTS AND SCIENCES
MIT MEDIA LAB

THESIS READER: ______________________

BLAKE SNYDER
SCREENWRITER, PRODUCER, AND AUTHOR OF "SAVE THE CAT!"
SCREENWRITING/STORY STRUCTURE PROFESSOR AT UCLA

THESIS READER: ______________________

DAVID RAKOFF
ESSAYIST, ACTOR, JOURNALIST
BESTSELLING AUTHOR OF "DON'T GET TOO COMFORTABLE" AND "FRAUD"

AUTHOR: ______________________

MATHEW LAIBOWITZ

Abstract

This dissertation will introduce a novel sensor network system designed to detect and capture fragmented events of human behavior that can be collected and sequenced into a cohesive narrative that conveys a larger overall meaning. Research into narratology will be performed in order to develop a parametric model of effective narratives that can be mapped onto sensor-detectable elements of human behavior. The network will be comprised of wearable sensors, environmental sensors, and video sensors that can identify and record events that fit specific narratives. Alternatively, the system can capture all events along with narrative meta-data for cataloging and browsing. This research seeks to provide new perspectives into everyday events by incorporating them into a larger story flow. It further seeks to create a new form of entertainment and journaling in which users are automatically presented with coherent videos created from their naturally-occurring events that follow the dynamics of a specific story of their choice or creation. The particular narrative models that are used to seed the system can be traded along with the created videos to form a new type of online community. An overall goal of this research is to investigate a method for understanding, reducing, and presenting large amounts of data using our innate ability to understand and interpret stories.

Contents

Chapter 1

Introduction

As humans, we are inherently able to identify relationships between events and extract some sort of overall meaning. We transform disconnected occurrences and ideas into cohesive narratives as a method to understand, remember, and communicate meaning. The main goal of this work is to design a ubiquitous sensor network that can identify and capture disjointed real-world events that fit a specific narrative and contribute to its overall story.

The first requirement of such a system is to design a parametric model of narrative built from elements that are detectable with the sensor network. Possible examples of these parameters are the level of conflict and agreement between characters, the level of social respect and disrespect towards each character, and the distance between a character and a particular goal. A story can then be written from which a parameterized model can be created. This particular model is then fed to the sensor network, which can search for events that exhibit similar dynamic shape (such as a character losing a lot of respect) to subsections of the model.

The sensor network will detect elements of human behavior indicating attention level, conflict level, and contextual changes from activity, location, gesture, and social signaling. This behavioral data will be mapped to the higher level narrative parameters. For example, certain patterns of attention between people combined with unsympathetic gestures can indicate a high level of conflict. An event that illustrates a change from conflict to agreement is usually an important event in any narrative.

As real-world events are detected that fit the mapping, they can be captured by imaging and audio sensors. These individual elements can be assembled based on where they best fit the narrative, resulting in a complete story that follows the dynamics of the authored story but will maintain the characterizations of the real people and the specifics of the real relationships.

The amount of media being captured around us is rapidly increasing. What started with photos and video is now evolving towards site-specific, enveloping media coverage as illustrated by such projects as Gordon Bell's MyLifeBits [1] and the Sensecam [2] initiative, both realizing a vision that dates back as far as 1945 to Vannevar Bush's Memex concept [3]. The research discussed in this proposal presents a system that uses sensor networks as a potential solution to organizing this ever-growing pervasive media acquisition and addresses the general research goal of cataloging and browsing massive amounts of data. It might not be possible to watch all of the captured video in a large-scale video network and there might be important characteristics that are difficult to visually observe. By storing the sensed narrative dynamics of an event along with the video, you can limit your search to clips that exhibit certain characteristics, such as that it is an establishing event, as opposed to part of the denouement.

The immediately apparent application of this system is for a new type of entertainment, where participants watch as events from their life are assembled into a cohesive video that conveys a similar meaning and feel to a story of their choice or creation. This exercise can provide new perspective into the everyday by metaphorically linking real activities and previously non-apparent relationships to a fictional narrative. The resultant video can be seen as a new form of diary/blog and multiple users can use the same story model to see how their video differs. Participants can trade story models which can lead to the creation of a community.

Chapter 2

Motivation

This section discusses the motivation behind the infusion of sensor networks with ideas drawn from narratology and some of the higher-level research goals and novel applications that result. Specific examples of related research are discussed later in Chapter 6.

2.1 Narratology

Humans create stories as a method to understand fragmented events and to extract and communicate meaning. The narratives that we develop provide insight into our lives and the world around us, and are essential mechanisms to our education, entertainment, and social well-being [4]. Or as Toni Morrison said "[Narrative] is, I believe, one of the principle ways in which we absorb knowledge." [5]

The study of the conveyance and the content of narratives is called narratology. The Russian Formalist school began to study Narratology in the 1920s. They divided the composition of a story into fabula and syuzhet. The fabula contains the story's themes, characters, and main points, and the syuzhet is the artistic and syntactic structure by which the fabula is conveyed. This school, as characterized by Vladamir Propp's work with Russian folktales [6] first suggested that all traditional story forms mix and match fabula fragments from a finite set of available scripts such as "hero saves damsel."

This was an early attempt at trying to understand the underlying requirements that distinguish a cohesive narrative from just a series of events. In Propp's case, he suggested that narratives require very specific archetypal elements to cohesively convey their morals.

As humans, we have an innate ability to recognize relationships between events and we can even predict future events based on what will best contribute to the overall point of the sequence of actions. If we can build a model that describes these relationships, we can build systems that can use this model to pull cohesive, effective narratives out of a complex stream of behavioral data. The proposed research into a model of this sort is described in section 3.1.

2.2 Sensor Networks

Advances in technology have allowed us to build and deploy systems comprised of enough sensate nodes to cover an environment with sensing capabilities that are always on and in communication. These systems provide us with a detailed and often not visually observable view of the covered area without interfering with the subject. The goal is then to use this information as an input to a variety of monitoring applications that can better understand and predict events.

At the time of writing this, most deployed sensor networks provide only the simplest of states, such as if there is a person in the room, if they have turned on the light, or if they are using the computer. In other words, these systems tend to simply return a sequence of events, but as any writer will say, the "what" is not that important, only the "why" matters. As narratives can be seen as a sequence of events with a "why", this research seeks to use a narrative model to create a sensor network application that detects beyond the simplest of disjointed human behaviors.

The field of distributed camera networks has been exploding lately, as illustrated by a dedicated workshop at SenSys 2006 [7]. But without a decent method for understanding the detectable events, it is pointless to use high bandwidth sensing. For example, capturing events on video with a multi-node distributed imaging sensor network would require large amounts of power to record, store, and transmit the video. It is really a requirement of a system such as this that the video gets filtered at the source and only events of significance should get captured. As vision systems are generally too large and power hungry to be used without affecting the subjects, the filtering has to happen through the use of other sensors or by extracting simple features from the live video stream [8]. These sensors can filter the video based on narrative content resulting in only clips that are involved in the story and convey something of importance.

2.3 New Perspectives

It is a natural goal of any human-centric research project to provide new perspectives into our lives and who we are as an individual and part of a larger social group. This research program seeks to do that by finding, capturing, and presenting the underlying narrative currents in our daily existence. We can use the techniques developed in this research to model specific narratives, detect elements of these narratives in natural human behavior, capture these events using image sensors, reorganize the events to tell the desired story, and present it to the participant. What is presented is essentially a new way to view the real events with them contributing to a cohesive narrative. This could bring to light previously unnoticed stories and relationships throughout the rest of one's life.

2.4 Entertainment and Online Communities

A major goal of this research is to create a new form of entertainment. The internet, more specifically YouTube, has shown a desire for user-created video content. Through the use of a narrative-aware imaging sensor network, one can automatically create a video comprised of real clips that follows a cohesive narrative based on a user-selected story.

2.4.1 Motivational Scenario

Reyna decides to create a new video over the course of the next month. She lives and works in an area covered by a ubiquitous imaging-enabled sensor network. Her favorite movie is The Karate Kid and she downloads a parameterized model of its story. She downloads this model into a small, wearable sensor node embedded in her watch. She will wear this for the duration of the month.

This identifies her to the imaging sensor network in her environments. When she, or anyone else she is with who is wearing a sensor node allowing themselves to be included, behaves in such a way that fits the narrative model of the Karate Kid, the imaging system captures the event. The video clips are collected and sequenced according to the Karate Kid's timeline.

Each day she can check the video as more and more clips are added and the video better fits the ebbs and flows of the original story. The final video follows the parametric model of the original story but has the personality of the events and relationships in Reyna's life.

Instead of selecting the Karate Kid story model, she could have made her own based on her own story or a story that has yet to have a model created. She will have access to a graphical tool for creating new models by creating curves for each parametric in the model. Once she creates a new model, she can post it online and add to the community.

Besides being entertaining user-created content, the resultant videos can be seen as a new form of online journal. Multiple people can use the same story model and she how the particulars of their lives alter the final video.

While a story model is running and a video is being generated, Reyna finds her life more interesting as she wonders what part of the Karate Kid will happen today. She sometimes even changes her behavior to affect the outcome of the *video and in the process adds some much needed variety to her month. Even after the month is over, she continues to enjoy her newfound perspectives on the world around her.*

2.5 Cataloging and Browsing of Data

As mentioned previously, the level of detail and amount of data that can be collected is limited by the system's ability to understand the data. In many situations, by understand we mean organize the data into meaningful categories that allows the correct elements to be appropriately retrieved and used. For example, if someone is making a film with 1000 cameras shooting as many as 1000 takes of a scene, it can be quite difficult to annotate the clips by hand. And even so, there may be qualities of the take that might not be obviously apparent when viewing that many clips, if it is even possible to view that many clips. By collecting meaningful sensor data, one can view only clips that fit a particular narrative transition, such as a rise in conflict level between two characters. If the mood of the story changes, the search can change to find a clip that has a slightly more forgiving interaction.

Chapter 3

Proposed Research

To complete this thesis, a human-centric sensor network will be deployed that is capable of evaluating sensed events with respect to their contribution to a specific narrative. In order to build this, a parametric model describing the components of an effective narrative will have to be developed and mapped onto the sensing capabilities of the network. A suite of software tools will enable the creation of models of specific narratives. A second device network will be deployed consisting of image sensing nodes which will provide distributed video capture for the system to call upon. An architectural diagram of the overall system is shown below in Figure 3.1. The following sections describe each of these aspects in more detail and section 5.1 describes already completed work.

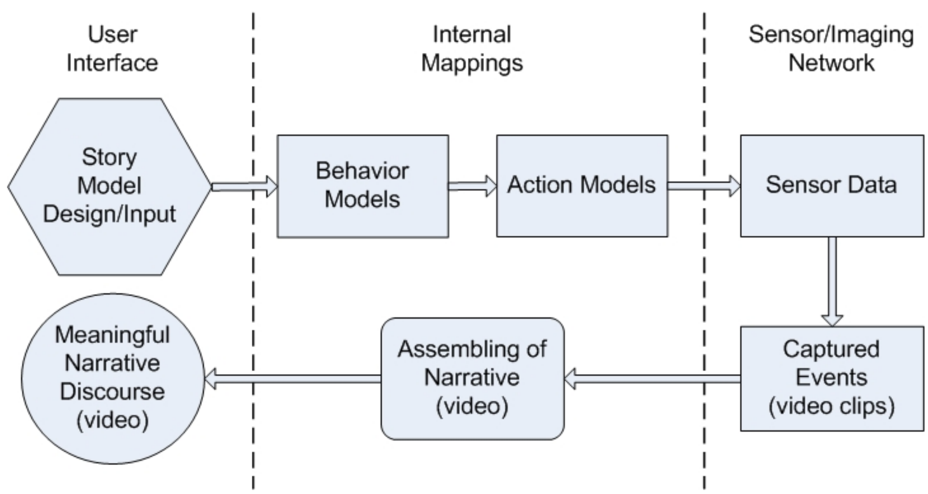


Figure 3.1: High-level architectural diagram of proposed system.

3.1 Parametric Model of Narrative

The first requirement of this research is to be able to represent stories in a structured form built from quantifiable parameters. The sensor network can then search for these parameters and identify events that fit a particular narrative or catalog events according to high-level narrative descriptors. These parameters are the essential components that separate a cohesive narrative from a disjointed sequence of events. Sections 3.1.1 through 3.1.3 present specific potential examples of these parameters and section 3.4 suggests how these rather elusive parameters might be sensed and measured.

The motivational scenario described in section 2.4.1 and the beginning of the process illustrated in figure 3.1 both call for a software interface to select and create specific parametric models. Shown in figure 3.2 is a mocked-up screenshot of an application intended for this use.

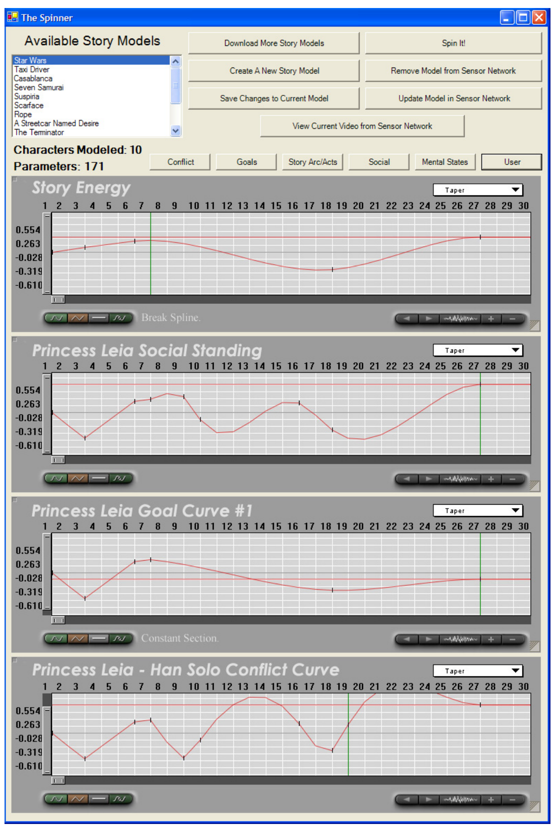


Figure 3.2: Screenshot of proposed story modeling software.

The above screenshot shows the software being used to view or edit a parametric model of the narrative of Star Wars. It reports that this model has 10 characters and 171 parameters. The interface shows the parametric curve of several of these parameters and how their value changes along the timeline of the story. The user can examine and make adjustments to all 171 parameters that make up this particular model. The number of parameters is determined by the number of characters in the model because there are certain parameters that exist for every character or every permutation of relationship between pairs of characters.

The sensor network can then detect real-world events that contribute something to the overall story that makes up the final video in the same way as particular events in Star Wars contribute to its overall narrative by detecting the parameters and looking for correlations with as many of the 170 parameters as possible.

The system can also work without a specific narrative the search for. In this mode, it looks for the raw narrative parameters and can use this information for cataloging applications as well as interactive applications where something happens in response to a change in a story parameter, such as conflict turning into agreement.

This leads to the most important part of this sub-topic of the proposed research: the actual, sensor-detectable parameters that describe a narrative. The following subsections outline some types of parameters that are being investigated for possible inclusion in this model.

3.1.1 Story Energy

We can define story energy as the effect the current narrative situation has on the character or characters and in turn the audience. It can be further described as having polarity (positive, negative, neutral) and an intensity level. Events such as the death of a character's mentor are considered negative with high intensity. Any dialog phrased negatively, such as simply saying "no, thanks" to someone is considered negative story energy, however, it is very mild. In non-abstract narrative, neutral energy is usually avoided. An act is characterized by a change in story energy, such as going from negative to positive or vice-versa. Effective narratives tend to alternate the polarity from act to act, as well as the intensity of the change. In other words, the shape of the story energy curve is often enough to determine if the scene falls in an establishing act, a climatic act, or a resolution act.

3.1.2 Conflict

Stories, and arguably anything of interest to humans, arise from conflict. If the main character is on a quest and nothing stands in their way, it is not much of a story. Despite the conflict being what keeps the story going, conflict is not action. Conflict is the 'why' behind the action and it can be internal, personal, social, and/or elemental. The way a character deals with conflict defines them to the audience. Conflict can be seen as the driving force behind the aforementioned story energy, changes in conflict can be seen as the delimiter for an act, and conflicts can affect mental states, in fact, conflict can be a mental state. With conflict being fundamentally important for storytelling, it may be possible to create an entire story model using conflict. An example of this is the interactive drama Façade [9]. In this game, a user witnesses a conversation between a husband and wife and can type in things to say at will. If the user types in nothing, the discussion continues. Whenever the user types in something, the software determines if it is in conflict or agreement with one of the characters and to what extent. This modifies the dialog and the narrative drastically, and the story unfolds in a way that seems very appropriate according to what the user has entered.

For the example of the constrained interactive drama, conflict was enough to completely steer the narrative. In the case of our model, conflict becomes an additional parameter to add detail to our model. We can add conflict timelines for each character for internal, social, and elemental conflict and we can add timelines for personal conflict with each of the other characters. Each timeline plots the intensity of conflict or agreement over time and most likely will show points of high rates of change located near event nodes from the act structure.

3.1.3 Social and Personal Character Goals

In narratives, characters are often defined by one or more goals. At various times in the story, the goal that is in the forefront of the character's action may change. The goals may be internal, such as a quest for some knowledge, personal, such as trying to stop another character from succeeding or social, such as trying to increase standing, popularity, or wealth.

We can add to our structural model a timeline for each goal of each character that relates how close the character is to achieving this goal to time. Even if the goal is not known to the character at a certain time, it can still have a value because a character can still be close or far away from reaching a goal regardless of if they realize it. If a goal is no longer desired by a character, it can be considered a reached goal.

In storytelling, there is often the concept of the ticking clock. This is some sort of time pressure to reach a particular goal. As the ticking clock winds down, the pressure to complete the goal increases, which will increase the conflict intensity unless the goal is achieved.

Social standing may warrant its own parameter. We have shown in past research [10] [11] [12] that we can detect social standing and interpersonal interests using basic sensing. Most stories involve some sort of social goals, in fact many stories are entirely a process of the character achieving popularity, respect, or influence. We can call this the respect curve.

3.2 Imaging Sensor Network

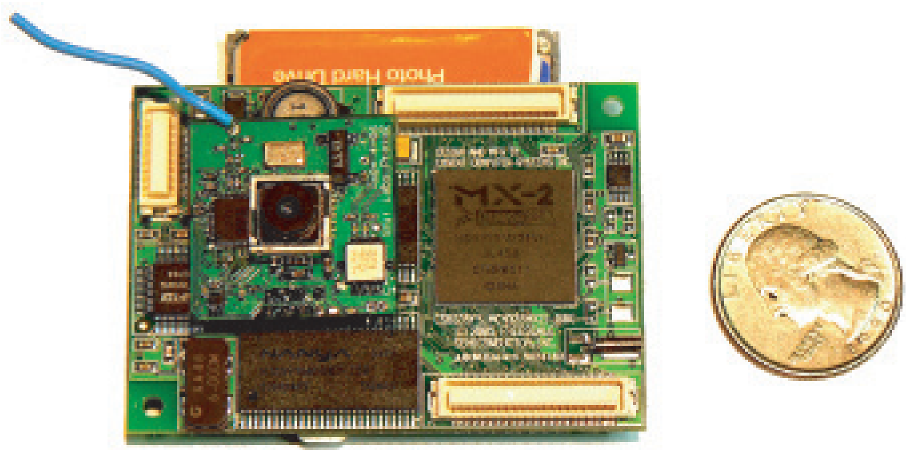


Figure 3.3: Early prototype of an imaging sensor node.

The imaging sensor network is comprised of nodes, such as the one shown in figure 3.3. It is a requirement that the nodes be small enough and cheap enough so that the network can cover the entire environment with video capturing ability. The nodes will have the following features:

- *CMOS Camera Sensor w/ Microphone* - A tiny video camera, identical to the ones used in high-end mobile phones. If available, it will have an autofocus feature. A pan/tilt feature is not needed due to the density of camera nodes.
- *TI DaVinci Processor* - A dual core microcontroller with an ARM-core for running an OS and a video DSP core that has a dedicated camera interface, video processors unit, and storage interfaces.
- *Flash Storage Device* - A 1-8GB flash storage solution for storing captured video and narrative sensor data.

- *RF Wireless for Data* - A wireless transceiver that allows it to communicate with wearable and other devices intended to observe human behavior. It is through this channel that the network asks the cameras to capture an event. It is also through this channel that the image node receives the sensor data to store with the video. The image nodes are fixed in place and provide location data to the mobile sensor nodes using the RF systems location engine.

- *High Bandwidth Wireless* - An 802.11 WiFi transceiver is included for retrieving the video clips. It can also be used for a backend infrastructure for the entire sensor network.

- *Sensors* - Nodes can be equipped with other sensors such as PIR motion detectors to help with power management and to provide stationary sensing channels to the rest of the network.

3.3 Human Behavior Sensor Network

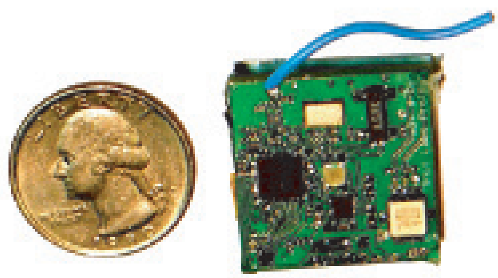


Figure 3.4: Early prototype of an narrative sensor node.

The sensor network charged with detecting the narrative elements of various events consists of nodes with the following features:

- *Wearable* - The nodes are worn to provide the best access to the human subjects. They may also be embedded in objects or loose in the environment.

- *Locateable* - Using the proprietary wireless, these nodes can locate themselves in reference to the image sensor infrastructure. The wearable nodes also are used to identify themselves to the imaging sensor network and to other narrative sensor nodes when they come into proximity of the user.

- *Sensate* - These nodes are equipped with a suite of inertial, gestural, aural, and optical sensors to determine activity, affect, and social conditions.

3.4 Mapping of Narrative Model to Behavior

The next piece of the proposed research plan is to create the mapping between sensor-detectable elements of human behavior and the components that make up a cohesive narrative. We can then use narrative structure to browse and understand the events that make up our lives. We can also use this mapping to detect and capture real events that fit a particular narrative which would allow the creation of an interpretation of a story using real-world events.

The following sub-sections contain thoughts on elements of this mapping. These ideas are based on research from the current field of sensing human behavior as exemplified by projects such as Singh's LifeNet [13], Weld's work in personalized user interfaces [14], Nathan Eagle's Reality Mining project [10], the previously mentioned Uber-Badge project [12], and Wolf's work on understanding the purpose of travel from GPS data [15]. There is also a rich body of literature relevant to designing a mapping of this sort including Ro et al's Pattern Classification text [16] and Wertheim's discussion of organizational behavior [17]. The mapping will develop with experimentation throughout the course of this research.

3.4.1 Conflict

While we can't sense all types of conflict, we can detect the level of conflict and agreement between two people when they are interacting. This can be achieved using activity level, body movement, gesture, and engagement. The first thing that can be detected is matching gestures, such as both people nodding and so on. In general, sympathetic motions can denote agreement and dissentious motions can denote conflict. If both parties are talking at a similar volume (below a certain threshold that indicates yelling) it can indicate agreement. This goes for activity level as well. Discrepancies in speech volume can indicate a conflict. The space between each person's turn to talk can be mapped to conflict level as well, too short can indicate an interruption, and too long can indicate lack of interest, both can indicate conflict. In general, people in agreement find a rhythm of communication both verbally and with their activity level and body language.

Certain specific gestures can show a certain conflict level, a handshake can indicate agreement whereas a clenched fist can indicate conflict. Relative body position can further present a conflict value, a hug gives one value and getting into someone's face gives a different value.

If two people are highly engaged steadily for a decent length of time (interested and attentive) it probably means they have a high level of agreement. If it is spiky, as in they engage and then disengage rapidly, it can indicate conflict because people generally do not stay engaged with someone they have a conflict with. The exception to this being both people staying engaged in a lengthy physical or verbal battle, but this can be detected by the motion and audio sensors. A difference in engagement level between the two parties indicates an unrequited attempt to get someone's attention or a flat-out dismissal of one of the parties, both can create/indicate tension. If both parties are disengaged, there may not be enough information to use engagement to determine conflict/agreement, or it may be that they are apathetic towards each other which is not a conflict and can be considered as just barely an agreement.

3.4.2 Social Respect

The narrative model contains a parametric curve for the social standing (respect) of each character. This can mapped to the engagement level that is detected from face-to-face time, body language (distraction), and conversational dynamics. Unlike conflict, social respect is an aggregate of all the characters engagement towards the character in question.

More specifically, social respect can be mapped to the speed at which the engagement level passes a certain threshold when the character enters a new situation. This separates engagement due to social respect from engagement due to some sort of situational content. Characters with a high level of social respect are engaging the moment they enter the room. Increases in social respect level are usually follows one or more situations where the character creates some sort of engaging content. In other words, it will be common to see several instances of people slowly becoming engaged with the main character followed by an instance of people instantly becoming engaged.

3.4.3 Goals

The next parameter of interest is the goal curve. Each character has a goal curve that indicates how far away they are from reaching their current objective. If the narrative contains a series of goals such that when the character reaches one goal they are given another one, the curve would most likely have a sawtooth shape.

The goal curve can be mapped to sensed parameters indicating a search. A search can be detected by rapid changes in location and physical state. Constantly changing attention from one person or activity to the next can further indicate attempting to reach a goal or find a solution. If a person increases their use of certain tools for searching such as the internet, the phone, and writing implements, it can indicate a large distance from reaching the goal. One-way engagement, such as eavesdropping or barking orders, is an important factor. As these things start to slow up and the person becomes more engaged with a single element, it can be related to reaching some sort of objective.

3.4.4 Story Energy

Story energy is considered the effect that the current narrative situation has on the audience. It is described as either negative or positive and has amplitude. In the simplest terms this is how happy or sad the narrative makes the audience feel. The author of the narrative manipulates the audience through the main character. By creating a sad or happy situation for the character, the audience is sympathetic and follows. In general, story energy is an aggregate of all the other parameters and might not need to be sensed individually. However, it may be possible to properly detect all the other parameters and yet capture a happy scene that is supposed to be highly negative on the story energy curve. Solving this will require additional analysis on the sensor data from the main character whenever he/she is present. It might be possible to determine happiness through very specific body motion curves and speech qualities. Speech recognition and facial expression recognition are sensing modalities not necessarily available for systems of this sort and require sophisticated analysis to gain high-level information.

If the main character from the narrative is manually linked to a particular person in the real world, it may be possible to train the system to determine story energy. In general people have certain indicators of their happiness such as posture, breathing, gait, and hand position. We can generalize assumptions about the formula that indicates sadness and happiness and wherever possible train the system according to a particular person to serve as the indication of story energy.

3.5 Story Structure and Cataloging

Generalized story structure maps such as the classic three act paradigm [18] and Blake Snyder's more detailed beat sheet [19] are used by writers to organize story events into effective story arcs. For this project, we can use the elements of these generalized story models to classify events that the sensors witness. For example, we can identify an event that exhibits the characteristics of the "All is Lost" beat in the story structure and identify an event that is a better fit with the "Catalyst" story beat. The best way to do this is by training the system by feeding it many known examples. These examples can be recorded using actors performing known scenes with the sensors. There exist many sophisticated algorithms and commercial software packages for creating mappings from training data. One such commercially available software package is NeuroSolutions [20] which provides data modeling algorithms based on neural networks.

3.6 Evaluation

As with any technological system, how well it performs fundamental tasks needs to be examined. For this system, this is how well the sensors track the desired story parameters and how well the video network captures individual comprehensible video clips.

The base hypothesis for this research is that we can use narrative modeling to filter and present data in such a way that it is understandable to us. To support our desired applications, the data being used is video captured human behavior. A major point of evaluation is whether or not the captured, collected, and sequenced video makes sense. In other words, does it exhibit a cohesive story that presents meaningful information about the characters and activities observed despite being built from fragments of events? If it uses a specific narrative as a base, does the captured video follow the dynamics of the original story? Does it take the audience on a similar journey despite coming from an entirely different setting with entirely unrelated dialog? We can also evaluate the system by cataloging all the video data with the narrative sensor data and trying to create video stories by using only the narrative data. This could identify which elements of the model work and which need more work.

This research contains many hard objectives, especially sensing elements such as story energy and goal achievement, but a high success rate might not be necessary. The story model will have some built-in redundancy as many of the parameters parallel each other. Misclassified events can be overwritten by better fit events as they happen and an occasional poorly fit clip might add to the enjoyment of the output. Furthermore, the system can be trained to the specific environment and participants, and a user-intervention step can be included to select from potential clips resulting in a higher success rate for future events.

If the output of the system is comprehensible, it indicates that our story model parameters and our mapping to the sensors are valid.

The final mark of evaluation is if our system has achieved its motivational goals and created new experiences and new perspectives to make our lives more fruitful and enjoyable.

Chapter 4

Contributions

If all goes according to plan, this research will make the following contributions to the world at large:

- A tested model of narrative that can be used across many fields including story generation and understanding, screenwriting, and interactive entertainment.
- A tested mapping of narrative to detectable occurrences that can support applications in social networking, behavior capture and predication, contextual computing applications and human-computer interaction, interactive entertainment, and filmmaking.
- A methodology for humans and machines to filter, organize, and understand large streams of data.
- A new form of entertainment in which participants and their surrounding life become the star their created story. The process by which this story is created effects the participants entire life and provides new ways to see everyday events.
- A new form of online journal and online community.
- A software application for creating narrative models.
- A software application for organizing video clips based on narrative content.
- Many advances in sensor network hardware design and communication protocols.

In the context of the Media Lab, the work proposed here incorporates and builds upon many of the Media Labs overarching research agendas and takes advantage of its multidisciplinary nature. There are clear connections with the Labs efforts to bridge the physical and digital worlds and extend human capabilities of perception, memory, and communication.

On the physical level, this project will augment the Media Lab itself with a permanently deployed dense imaging network and sensor network infrastructure that can support many future applications.

Chapter 5

Research Plan

This section outlines the the concrete steps I will take to complete the research proposed in section 3.

5.1 Completed Work

I have tested some of the behavioral and social sensing techniques using the UbER-Badge [12].

I have developed the hardware, protocols, and software for the proprietary sensor network wireless communication and it is currently being used in many different projects. I am working on a new version of this with sponsor support from Chipcon that uses 802.15.4 and/or Zigbee to provide device discovery, data communication, and a location engine.

The Communicator Badge is currently under development and may be all that is needed hardware-wise for the wearable narrative sensor node.

I have been working on the video sensor node for some time now and Texas Instruments, Alps, and Seagate are all on board to donate components. This system is described with more detail in Appendix A.

Through the general exam process and through discussions inside and outside of the Media Lab, a beginning version of the narrative model and sensor mapping has been defined.

5.2 Privacy, Security, and the Use of Human Subjects

Deploying a video network in the lab raises obvious privacy and security concerns. See [21] for a good summary of these issues. The video network will only record when it detects someone wearing a sensor node identifying to the system that they have agreed to be part of the test. This research will require approval from the Committee on the Use of Humans as Experimental Subjects (COUHES). Further notes on privacy are given in Appendix A.

5.3 Timeline

- 17 April 2007 : Design imaging sensor node and start fabrication.
- 30 May 2007 : Develop story modeling software.
- 15 June 2007 : Deploy imaging sensor network.
- 30 June 2007 : Evaluate Communicator badge or build demo units of new wearable sensor platform.
- 15 July 2007 : Complete working prototypes of hardware and software for narrative sensor nodes.
- 30 August 2007 : Demonstrate imaging network, sensor nodes, and narrative software individually.
- 1 September 2007 : Begin integration of all components.
- 15 September 2007 : Extensive testing on hardware and software. Begin to evaluate mappings.
- 30 September 2007 : Begin story structure training for cataloging and browsing test applications.
- 15 October 2007 : Develop story/entertainment application.
- 30 October 2007 : Develop many specific test narratives and evaluate system.
- 15 November 2007 : Perform two-day test with multiple subjects.
- 30 November 2007 : Perform one-week test with multiple subjects.
- 15 December 2007 : Finish evaluations.
- 31 December 2007 : Develop online community sites and applications.
- 15 January 2008 : Finish evaluations.
- 31 February 2008 : First Draft of dissertation.
- 7 March 2008 : Second Draft of dissertation.
- 15 April 2008 : Public Defense of dissertation.
- 30 April 2008 : Submit final draft.

5.4 Required Resources

In addition to the tools and software we already use everyday, the following resources are needed:

- Parts, tools, and manufacturing facilities to complete the imaging sensor network. Several parts are being donated by sponsors.
- As many as 20 Communicator Badges with special modifications or the facilities to design new wearable sensor nodes.
- Powerful server to archive video data.

Chapter 6

Related Work

Although the use of narrative modeling to understand and use imaging sensor network data has no direct precedent, it builds upon current work in the encompassed fields of distributed sensor networks, human and social sensing, knowledge extraction, and narratology.

Wireless sensor networks have become a large area of research, with many universities and institutes contributing. Strategic seed programs begun in the 1990s such as DARPAs SENSIT initiative [22], have grown into an international research movement.

Advances in this field have lead to smaller, cheaper sensor nodes [23]. The SenseCam [2] being developed at Microsoft research brings video sensing and image gathering to a small, low-power node. Devices such as this have started to support a number of distributed camera systems such as the ones developed by Wayne Wolf at Princeton as a test platform for distributed vision algorithms [8]. The Panoptes system [24] developed at OHSU shows how redundancy in video camera systems can keep the information detail even in the event of a network outage as well as demostrated a reprogrammable video platform.

Advances in ubiquitous and wearable sensing have allowed us to observe human subjects without interference. This supports investigations into organizational behavior, social networking, and information diffusion. The Human Dynamics Group at the MIT Media Lab has developed several projects that look at human social behavior, such as the Sociometer [11] and the Reality Mining project [10]. Recent collaborations with the Responsive Environments group have utilized the Uber-Badge platform to look at group social signaling and interest [12]. This has led to a new platform called the Wireless Communicator [25] which is being considered as a platform to initially develop the ideas described in this proposal.

In section 3.4, related work from outside the MIT Media Lab addressing human behavior sensing and understanding is listed through references to Weld's work in personalized user interfaces [14], Wolf's work on understanding the purpose of travel from GPS data [15], Ro et al's Pattern Classification text [16], and Wertheim's discussion of organizational behavior [17].

In Bove and Mallett's paper "Collaborative knowledge building by smart sensors" they explore "decentralized approaches for gathering knowledge from sensing devices" [26]. This body of research is perhaps the closest relative of the proposed research which attempts to identify narrative-based knowledge in the data collected from reality. It discusses several examples, including the apropos Two Eyes camera project [27]. In a similar vein, Bo Morgan's research is concerned with the use of commonsense algorithms for understanding sensor network data [28].

Narratology has been discussed in previous sections via references to Façade [9], Kearny [4], Morrison [5], and Propp [6]. Story structure is an important topic for screenwriters [29] to help develop effective narratives. The field of story generation looks into developing parametric models of narrative to create systems for interactive entertainment. An example of this is Meehan's Soap Opera plot generator [30]. Prior to this, in 1973, Klein's essay discussed methods for automatically writing a novel [31]. Glorianna Davenport has continually performed research into our use of stories for interactivity and understanding [32] and her Media Fabrics group at the MIT Media Lab has been using narrative ideas as a means for communication [33] and for looking at the design of and our relationships with everyday objects [34].

Using story concepts to understand, catalog, and browse information has been discussed as far back as 1977 with Schank et al's book "Scripts, Plans, Goals, and Understanding" [35] followed in 1982 by Lehnert's article [36] describing the use of plot unita and narrative to summarize text. Andrew Gordon's PhD thesis [37] discussed using additional knowledge such as narrative structures, common-sense, and sensor data to browse information. More recently, the StoryNet project [38] has built a database of scripts that are used to search through and predict subsequent elements of large collections of data.

Chapter 7

Bios

The four readers involved with this project have been selected to cover the breadth of topics spanning from the creation of story content from everyday events, story structure and modeling, using sensors to detect high-level attributes of human social behavior, and the creation of ubiquitous sensor and imaging systems.

Dr. Joseph A. Paradiso

Joseph Paradiso is Associate Professor and Sony Career Development Professor of Media Arts and Sciences and head of the Lab's Responsive Environments group.

Paradiso has a remarkably diverse background, ranging from high-energy physics detectors and spacecraft control systems to electronic musical instruments. He now explores the development and application of new sensor architectures and extremely dense sensor/processor networks for human-computer interfaces and intelligent spaces. An expert on sensing technology, Paradiso has developed a wide variety of systems that track human activity using electric field sensing, microwaves, laser ranging, passive and active sonar, piezoelectrics, and resonant electromagnetic tags.

A summa cum laude graduate of Tufts University, Paradiso received his PhD in physics from MIT as a Compton Fellow. Before joining the Media Lab, he worked at ETH in Zurich, and the Draper Laboratory in Cambridge, Massachusetts. He is the winner of the 2000 Discover Magazine Award for Technological Innovation for his expressive footwear system.

Alex "Sandy" Pentland

Prof. Alex (Sandy) Pentland is a pioneer in wearable computers, health systems, smart environments, and technology for developing countries. He is one of the most-cited computer scientists in the world.

He is a co-founder of the Wearable Computing research community, the Autonomous Mental Development research community, the Center for Future Health, and was the founding director of the Media Lab Asia. He was formerly the Academic Head of the MIT Media Laboratory, and is MITs Toshiba Professor of Media Arts and Sciences, and Director of Human Dynamics Research.

He has won numerous international awards in the Arts, Sciences and Engineering. He was chosen by Newsweek as one of the 100 Americans most likely to shape the next century.

Blake Snyder

In his 20-year career as a screenwriter and producer, Blake Snyder has sold dozens of scripts, including co-writing Blank Check, which became a hit for Disney, and Nuclear Family for Steven Spielberg – both million-dollar sales. Named "one of Hollywood's most successful spec screenwriters," Blake continues to write and sell screenplays, most recently a 2006 sale of a horror-comedy that will be in theatres in 2007.

His book, Save the Cat! The Last Book on Screenwriting You'll Ever Need, was published in May, 2005, and is now in its sixth printing. It has prompted "standing room only" appearances by Blake in New York, London, Chicago, and at the 2006 Screenwriting Expo in Los Angeles.

Blake's method has become the "secret weapon" of many development executives, managers, and producers for its precise, easy, and honest appraisal of what it takes to write and develop stories that resonate. And the release of Save the Cat! The Last Story Structure Software You'll Ever Need has codified this method into an easy to use format that has been selling out at the Writers Store in Los Angeles since its debut in November 2006 and is also available online.

Blake has taught at Chapman University, UCLA, Vanderbilt University, and Rockport College. His book is the basis of screenwriting classes at Cornell University, The New School, and many major universities in the U.S. and Canada. Apparently it is not quite the last book on screenwriting youll ever need, as Blake is currently writing the eagerly awaited sequel, Save the Cat! Goes to the Movies: The Screenwriter's Guide to Every Story Ever Told, which will be out in fall 2007.

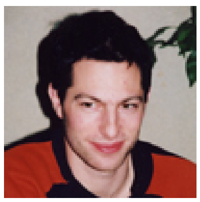
Blake has a B.A. in English from Georgetown University and lives in Beverly Hills, California. He is a member of the Writers Guild of America, west.

David Rakoff

David Rakoff (born 1964) is an essayist, journalist, and actor. Originally from Canada, he obtained dual Canadian-American citizenship in 2003, and currently resides in New York City.

Rakoff has written for the New York Times Magazine, Outside, GQ, Vogue and Salon. He has also been a frequent contributor to the radio program This American Life on Public Radio International. Rakoff's essays have been collected in the books Fraud (2001, ISBN 0-385-50084-X) and Dont Get Too Comfortable (2005, ISBN 0-385-51036-5).

He was featured on This American Life's holiday show (Episode 305) on December 23, 2005. Rakoff also has the distinction of being the only other person aside from Ira Glass to host an episode of This American Life (Episode 248 - Like It Or Not). He appeared on The Daily Show on October 5, 2006 and on the Late Show with David Letterman on October 31. 2006.



Mathew Laibowitz

Mat received a BS in Electrical Engineering and Computer Science from Columbia University in 1997. He received a second degree from NYU in Film in 2000 while working at IBM's TJ Watson Research Center. He joined the Media Lab's Responsive Environments Group under Dr. Joseph Paradiso as a master's student in 2002 and received his MS in 2004. His research focuses on novel applications of human-centric sensor networks and the creation of technologies and methodologies for new types of experience.

Appendix A

Proposal for a Sensate Media Lab

This proposal presents a new system intended to be permanently installed throughout the new Media Lab building during construction to provide a platform for the research and deployment of sophisticated distributed applications starting immediate upon the building's occupation.

This is a key step towards:

- Enhancing the Media Lab experience for visitors and the extended community
- Fostering collaboration between research projects by providing a common framework
- Jumpstarting research into new application domains by providing a platform of services including communication, sensing, and storage

A.1 System Architecture

The system will be comprised of deck-of-cards sized devices mounted on the ceiling or walls every 15-20 feet covering the entire building. A thin power wire will be connected from device to device.

The standard features of each node are:

- Dual-core processor including an ARM processor and video DSP which can run video codecs and vision algorithms
- 802.11g WiFi connectivity for backend data communication and deployment of new applications

- Dedicated device wireless communication channel, backwards compatible with the Plug, Tricorder, and Wireless Communicator badge, provides an interface to sensor network devices including wearable and mobile devices. Through this channel, the sensor devices can receive location information, communicate to the backend WiFi network, and utilize the additional sensors and storage in the infrastructure. Will provide use 802.15.14 and Zigbee support.
- Up to 8GB of storage in every node providing location-specific distributed storage, further accessible through the backend network
- Image Sensor provides minimum 1.3MP video and stills to the video dsp subsystem
- Motion Sensor
- Small cellphone style LCD for status information
- Expansion Bus for addition of custom hardware
- Optional sensors on select nodes, such as the Canesta ranging imager
- Audio input and output
- Optional Bluetooth
- Connection to handheld cameras and cellphone cameras to support digital graffiti and other content creation applications

A.2 Application Framework

Each node will have a custom Real-Time OS built from the Windows CE kernel. This is a multitasking OS allowing multiple research projects to be active simultaneously. It will provide easy remote development, debugging, and deployment of applications using standard Microsoft Visual Studio development tools. In addition, the OS will provide the following features:

- The nodes will be accessible via SFTP, Samba, HTTP, and SSH.
- A complete set of device drivers for all hardware features will speed application development
- Standard file systems and file formats allowing easy access to data
- Easy remote administration and remote debugging/deployment of applications

A.3 Privacy

As with any publicly installed system, privacy is a major issue. In this case, the platform does nothing without applications. All applications that run on this platform will require approval from the Committee on the Use of Humans as Experimental Subjects. In general, the applications can be respectful of privacy by providing methods and requirements for people to Opt-in, Opt-out, or both. An opt-in application would require people to carry a token that identifies them to the system and turns it on. An opt-out system would require anyone people who do not wish to be involved to carry a token that turns off the system when they are in range. A combination system would allow people without any tokens to only be sensed when someone with an opt-in token was in range and there are no opt-out tokens in range.

A.4 Sponsor Collaborations

There has already be substantial support from Texas Instruments who have provided the processor development platform, Seagate who have provided sample 8GB storage devices, and ALPS who intend to provide us with image sensors. As this system will support many applications and be highly visible, support from many more sponsors is expected.

References

[1] Jim Gemmell, Roger Lueder, and Gordon Bell. The mylifebits lifetime store. In *ETP '03: Proceedings of the 2003 ACM SIGMM workshop on Experiential telepresence*, pages 80–83, New York, NY, USA, 2003. ACM Press. 5

[2] Jim Gemmell, Lyndsay Williams, Ken Wood, Roger Lueder, and Gordon Bell. Passive capture and ensuing issues for a personal lifetime store. In *CARPE'04: Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences*, pages 48–55, New York, NY, USA, 2004. ACM Press. 5, 25

[3] Vannevar Bush. As we may think. *The Atlantic Monthly*, 176(1):101–108, 1945. 5

[4] R Kearney. *On Stories*. Routledge, 2002. 6, 26

[5] Philip Page. Furrowing all the brows: Interpretation and the transcendent in toni morrison's paradise. *African American Review*, 35(4):637–649, 2001. 6, 26

[6] Vladimir A. Propp. *Morphology of the Folktale*. Publications of the American Folklore Society. University of Texas Press, 1928. 6, 26

[7] *SenSys '06: Proceedings of the 4th international conference on Embedded networked sensor systems, Workshop on Distributed Smart Cameras*, New York, NY, USA, 2006. ACM Press. Chair-B. Rinner and Chair-W. Wolf. 7

[8] Burak Ozer Wayne Wolf and Tiehan Lv. Smart cameras for embedded systems. In *IEEE Computer*, volume 35, pages 48–53, September 2002. 7, 25

[9] Procedural Arts. Facade. `http://www.interactivestory.net`. 13, 26

[10] Nathan Eagle and Alex (Sandy) Pentland. Reality mining: sensing complex social systems. *Personal Ubiquitous Comput.*, 10(4):255–268, 2006. 13, 16, 25

[11] T. Choudhury and A. Pentland. Sensing and modeling human networks using the sociometer. In *Wearable Computers, 2003. Proceedings. Seventh IEEE International Symposium on*, pages 216–222, 18-21 Oct. 2005. 13, 25

[12] M. Laibowitz, J. Gips, R. AyIward, A. Pentland, and J.A. Paradiso. A sensor network for social dynamics. In *Information Processing in Sensor Networks, 2006. IPSN 2006. The Fifth International Conference on*, pages 483–491, 19-21 April 2006. 13, 16, 22, 25

[13] Push Singh and William Williams. Lifenet: a propositional model of ordinary human activity. In *Workshop on Distributed and Collaborative Knowledge Capture (DC-KCAP) at K-CAP 2003*, 2003. 16

[14] Daniel S. Weld, Corin Anderson, Pedro Domingos, Oren Etzioni, Krzysztof Gajos, Tessa Lau, and Steve Wolfman. Automatically personalizing user interfaces. In *IJCAI03*, Acapulco, Mexico, August 2003. 16, 26

[15] Guensler R. Wolf J and Bachman W. Elimination of the travel diary: an experiment to derive trip purpose from gps travel data. In *Transportation Research Board 80th annual meeting*, 2001. 16, 26

[16] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, November 2000. 16, 26

[17] E. Wertheim. Historical background of organizational behavior. 16, 26

[18] Syd Field. *Screenplay: The Foundations of Screenwriting*. Dell Publishing, New York, 1982. 18

[19] Blake Snyder. *Save the Cat! The Last Book on Screenwriting You'll Ever Need*. Michael Wiese Productions, 2005. 18

[20] NeuroDimension. Neurosolutions. 18

[21] A. Soppera and T. Burbridge. Wireless identification privacy and security. *BT Technology Journal*, 23(4):54–64, October 2005. 22

[22] Chee-Yee Chong and S.P. Kumar. Sensor networks: evolution, opportunities, and challenges. *Proceedings of the IEEE*, 91(8):1247–1256, Aug. 2003. 25

[23] B. Warneke, M. Last, B. Liebowitz, and K.S.J. Pister. Smart dust: communicating with a cubic-millimeter computer. *Computer*, 34(1):44–51, Jan. 2001. 25

[24] Wu chi Feng, Brian Code, Ed Kaiser, Mike Shea, and Wu chang Feng. Panoptes: scalable low-power video sensor networking technologies. In *MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia*, pages 90–91, New York, NY, USA, 2003. ACM Press. 25

[25] Joseph A. Paradiso Daniel Olgun Olgun and Alex (Sandy) Pentland. Wearable communicator badge: Designing a new platform for revealing organizational dynamics. In *10th International Symposium on Wearable Computing (Student Colloquium)*, 2006. 25

[26] V. M. Bove and J. Mallett. Collaborative knowledge building by smart sensors. *BT Technology Journal*, 22(4):45–51, 2004. 26

[27] J. Mallett and Jr. Bove, M.V. Eye society. In *Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on*, volume 2, pages II–17–20vol.2, 6-9 July 2003. 26

[28] Bo Morgan and Push Singh. Elaborating sensor data using temporal and spatial commonsense reasoning. In *BSN '06: Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks (BSN'06)*, pages 187–190, Washington, DC, USA, 2006. IEEE Computer Society. 26

[29] June Noble. *Steal This Plot: A Writer's Guide to Story Structure and Plagiarism*. Paul S. Eriksson, 1985. 26

[30] J. Meehan. Tale-spin. In R. Schank and C. Reisbeck, editors, *Inside Computer Understanding: Five Programs Plus Miniatures*, pages 197–226. Lawrence Erlbaum, Englewood Cliffs, NJ, 1981. 26

[31] Sheldon Klein, J. F. Aeschlimann, D. F. Balsiger, S. L. Converse, C. Court, M. Foster, R. Lao, J. D. Oakely, and J. D. Smith. Automatic novel writing. In W. Burghardt and K. Hölker, editors, *Text Processing/Textverarbeitung*, pages 338–412. Walter de Gruyter, Berlin and New York, 1979. (Published version of UWCS Technical Report No. 186, 1973). 26

[32] Ali Mazalek and Glorianna Davenport. A tangible platform for documenting experiences and sharing multimedia stories. In *ETP '03: Proceedings of the 2003 ACM SIGMM workshop on Experiential telepresence*, pages 105–109, New York, NY, USA, 2003. ACM Press. 26

[33] Aisling Kelliher. Everyday cinema. In *SRMC '04: Proceedings of the 1st ACM workshop on Story representation, mechanism and context*, pages 59–62, New York, NY, USA, 2004. ACM Press. 26

[34] Hyun-Yeul Lee. Provenance as narrative history. MIT Media Lab General Exams. 26

[35] Roger C. Schank and Robert P. Abelson. *Scripts, plans, goals, and understanding*. Lawrence Erlbaum, 1977. 26

[36] W Lehnert. Plot units: A narrative summarization strategy. *Strategies for natural language processing*, pages 375–414, 1982. 26

[37] Andrew Steven Gordon. *The design of knowledge-rich browsing interfaces for retrieval in digital libraries*. PhD thesis, 1998. Adviser-Roger C. Schank. 26

[38] Ryan Williams, Barbara Barry, and Push Singh. Comickit: acquiring story scripts using common sense feedback. In *IUI '05: Proceedings of the 10th international conference on Intelligent user interfaces*, pages 302–304, New York, NY, USA, 2005. ACM Press. 26