Interactive Documentary: A Production Model for Nonfiction Multimedia Narratives

Insook Choi

Emerging Media Technologies, Department of Entertainment Technology
New York City College of Technology of the City University of New York
300 Jay Street V411, Brooklyn, NY 11201, USA
ichoi@CityTech.CUNY.edu

Abstract. This paper presents an interactive production model for nonfiction multimedia, referred to as interactive documentary. We discuss the design of ontologies for authoring interactive documentary. A working prototype supports the use of reasoning for retrieving, composing, and displaying media resources in real-time. A GUI is designed to facilitate concept-based navigation which enables queries across media resources of diverse types. A dual-root-node data design links ontological reasoning with metadata, which provides a method for defining hybrid semantic-quantitative relationships. Our application focuses on archiving and retrieving non-text based media resources. The system architecture supports sensory-rich display feedback with real time interactivity for navigating documents' space. We argue an experience of narratives evolves through the performitivity in the interactive narrative structure when the constituents are mediated by common ontology. The consequential experience identifies a renewed practice of oral tradition where the accumulative sensorial propositions inform narratives, such as in performance practice.

Keywords: Documentary, ontology, authoring, media production, cognitive architecture, oral tradition, GUI, interaction design.

1 Introduction

Documentary in film practice is an established genre. Yet its definition often undergoes a discursive path. Two factors play consistently in various definitions: 1) reality is captured in some forms of documents and 2) the documents are subjected to assemblage to serve a larger context. This paper proposes a production model of documentary with respect to the emerging practices with interactive technologies. As a starting position the paper adopts the simplest task definition of traditional documentary, that of Vertov: "to capture fragments of reality and combine them meaningfully" [1]. We attribute Grierson's "the creative treatment of actuality" [1] to be at the heart of documentary practice for both spectators and authors. In this spirit we facilitate interactivity to reduce the gap between the two poles in documentary culture, between users and producers, between production and reproduction, and between authoring and an act of inquiry.

A. Nijholt, D. Reidsma, and H. Hondorp (Eds.): INTETAIN 2009, LNICST 9, pp. 44 – 55, 2009.

The construction of the documentary utilizes both physical and literary narrative devices, found in many forms of storytelling. The devices are engaged by information sources such as first-person accounts to anchor them in factual circumstances. The result is a narrative about factual objects and events which include original and reconstructed documents and representations of them. The work of narrative relies significantly upon human memories for reconstructing observations and experiences. In terms of narrative structure, these basic cognitive functions create a unique role of the narrator as a direct or indirect subject in the narrative.

An act of authoring involves grounding perspectives and rationales as in storytelling. Storytelling in turn is an act of authoring of performance instances. In interactive documentary the performance is an act of authoring parallel to illustrated storytelling or annotating a flow of media resources. In our architecture, the use of ontologies is brought to organize the references and media resources. Currently the working prototype includes 2D images and videos, 3D graphic models and scenes, simulations, and data-driven and procedural auditory and visual processing of resources.

The remaining part of Section 1 covers related works through a comparative analysis with respect to the proposed system. Section 2 describes the production model in our system and we work through the definition of an interactive documentary. Section 3 describes the working prototype focusing on 1) the role of path-making for generating narrative recall, 2) the use of interactive reasoning to retrieve media resources, and 3) a graphical user interface design for queries across media resources of diverse types. Section 4 describes the design of ontological data, specifically a dual-root-node data design linking ontological reasoning with metadata, which also serves as a method for defining hybrid relationships of semantic and quantitative data.

1.1 Related Work

Interactive Documentary combines elements of generative storytelling systems with an interaction framework related to live media, virtual reality performances, and installations. The present application of semantic networks came about while developing generative models for interactive media performance [2]. Bocconi's "Vox Populi" [3] has similarities to the interactive documentary concept in terms of functional goals for preparing semantics. One common goal is to provide production-level access to documentary material for community exploration and reconfiguration beyond the constraints of pre-determined linear presentations. Another similarity is the role of an initial design phase to specify a semantic context that will result in a graph structure of semantic units. This phase is parallel to "pre-authoring," discussed in Section 2. Bocconi describes semantic annotation of individual media resources; we use a process of grouping resources to concepts, rather than annotation of each resource. Vox Populi uses a thesaurus of rhetorical constructs to determine the point of view expressed in spoken text. In distinction our structured vocabulary of concepts describes contents of media resources ("traffic downtown") and real-world objects depicted in media resources ("my car"; "6th avenue and 42nd street"). Both projects generate data-driven media sequences by navigating a graph structure; our prototype operates in real-time with navigation controlled by an observer. Vox Populi is highly specific for rhetorical relationships in video sequences, whereas we utilize a *display grammar* that generates "display prosody" in real-time for multiple types of media,

dynamically determining durations and inter-media composition and tempos. Prosody is determined through ontology of qualitative aspects of media resources such as tone, duration, and other sensorial properties, accounted in the display grammar.

"Evolving Documentary" [4] and "Multi-Threaded Stories" [5] propose models for documentaries that can be modified in successive presentations as situations evolve and new media resources are available. This is similar to our model of an observer with an authored work creating alternate semantic "paths" to customize and extend the work. "Very Distributed Storytelling" [6] sketches an idea of widely embedded interactive media governed by a central narrative production system. Installation-based narratives are similar to our previous work applying motion-sensing and pattern-recognition to observers' movements in a gallery to generate paths through a semantic network [7].

Interactive Documentary most notably differs from previous work in its focus on the integration of media resources of multiple types with semantic composition and real-time interaction. This follows from a body of applications and research in live performance for virtual reality and interactive media installation [8]. Mutual interests shared by these research histories indicate a likelihood of productive exchanges in the areas of semantic computing, media resource analysis and signal processing across multiple display devices and modalities.

1.2 Production in a Heterogeneous Media Ecosystem

Today media content is often created as a convergence of diverse media types deployed on diverse devices. Display systems such as personal communications devices can differ considerably from the production systems where media resources originate. And their roles can be reversed: personal devices can generate media in consumer formats that are repackaged and distributed by media syndication. It is an open, multi-platform ecosystem. Media can be authored to combine resources that are pre-selected with resources that are automatically searched and pulled from diverse providers. Authoring combined with semi-automation can re-use media resources to create new content. Social networks represent heterogeneous communities of media end-users as producers where oral histories are widely exchanged in multiple media formats. These practices may be formalized through a model for interactive documentary production. The model presented here is designed to support a heterogeneous media capacity by systematizing semantic authoring functionality:

1. Common representations for reasoning over media of unlike types;
2. Semantic representations of authored media shared by multiple observers;
3. Ability to author multiple versions from shared semantic structure;
4. Ability to modify the semantic structure without changing the media resources;
5. Ability to modify the media resources without changing the semantic structure.

2 Interactive Documentary Production Model

As a working definition, Interactive Documentary is a production model for telling and retelling nonfiction narratives by means of interactive recall for retrieving media resources in an open data space. Nonfiction refers to representations of real-world

artifacts. Narrative is synthesized through composite interactions of components and constituent media resources. Authoring in the context of interactive media refers to the design of instruction sets integrating contents of multiple types, media devices in an anticipation of observers' actions to realize terminal presentations. The instruction sets are embedded with media content or otherwise transmitted to display systems and devices. An authoring process configures initial conditions, combinations of media resources and conditional procedures for generating program contents. Authoring anticipates actions of observers who are sensed by the system.

Our prototype system combines a capacity for (1) *authoring interactive media* with (2) an *interactive authoring process*. The first refers to production of procedures for processing media resources while responding to observers' actions. The second refers to an authoring environment supported by real-time media processing to display results of procedures. The objective is to support both capacities in close proximity with proper system architecture and performance.

When a system enables interactive authoring with real-time feedback, the authoring process is similar in many ways to a presentation process, placing the author in a role akin to a presenter or a performer who is rehearsing and modifying their material. This quality of performativity emphasizes similarities to oral tradition. For system configuration some aspects of planning, source material acquisition and concept design are not feasible for real-time semantic interaction. These are prepared in advance in a process we refer to as *pre-authoring*. Figure 1 illustrates an interactive documentary production model comprised of a pre-authoring phase and an interactive authoring phase.

Pre-authoring tasks consist of

- Review and ingest media resources to an ontological repository
- Define concept vocabulary
- Assert membership of resources to concepts (i.e. identify concepts that describe the media resources)
- Design *display grammar* for the automated display of resources.

The interactive authoring tasks consist of

- Access the concept graph structure through an interface
- Explore structured vocabulary of concepts
- Generate queries to display media resources associated with a concept
- Explore paths of queries previously authored
- Modify existing paths or create new paths by arranging concept nodes
- Define new concepts by applying logical expressions to existing concepts.

3 Interactive Documentary Prototype

In Figure 1 three layers of function modules in the architecture relate to control flow and data flow with temporal differentiation. Active Observation modules support fine-grained scheduling for immediate action-feedback response in the GUI and media displays. Working memory modules weigh relationships among concepts and media contents and provide heuristics for displaying these relationships. Long-term memory

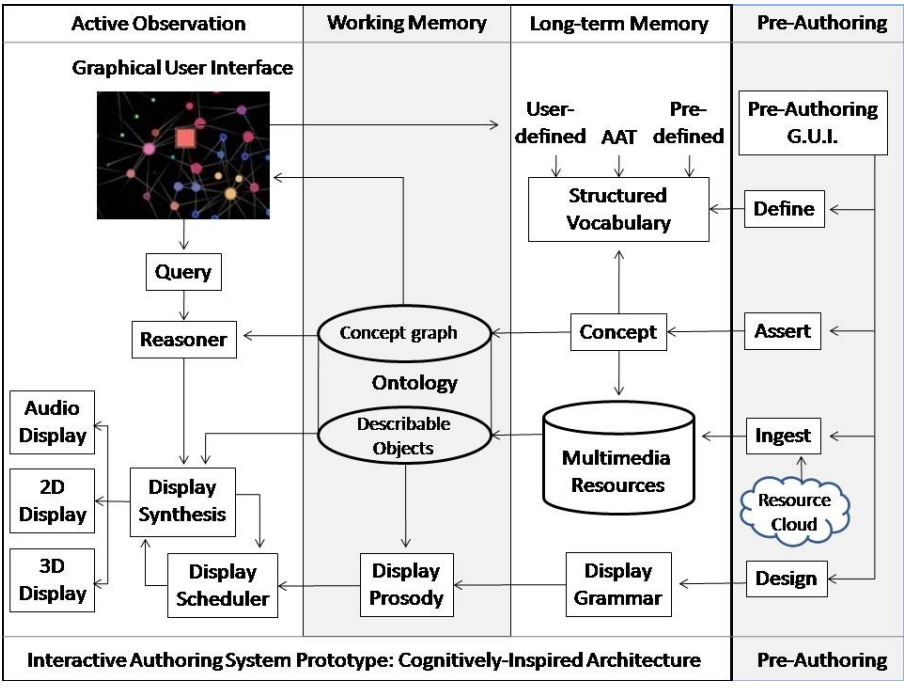

Fig. 1. Interactive Documentary Production Model

modules represent deep structure and media assets supplying production capacity. Semantic and media assets are brought into relationships in the computational workflow.

We describe the architecture as cognitively-inspired in reference to our system design criteria to facilitate concept formation, computational inference, and the observer's cycle of discovery, prediction, and path refinement. Von Foerster draws a distinction between the "invariance of quality" in storage and retrieval, and the synthesis of information in recognition and recall [9]. He regards perception, memory, and inference (prediction) as the requisite components of cognitive processes. The pre-authoring phase supporting the composition of a concept graph indicates a strong role for prediction in the process of interactive exploration and path-making. To design a concept graph is to predict as well as to perceive. To explore a concept graph is to predict as well as to discover. To traverse a path is to recall a sequence of predictions in the form of queries. While computational inference supports storage and retrieval, an author's concept designs and path-making constitute recognition and recall. Paths are coupled to the production architecture for observers' dispositions for learning through ontological inference.

For semantic organization and reasoning we have adopted an ontological data design. The authoring process is formalized as path planning through a semantic data structure—an ontology describing media resources of diverse types. Ontological structure is defined as a set of logical expressions that may be interpreted as a directed graph of concept nodes, where edges represent relatedness of concepts. Structure is encoded in OWL file format using an open source editing tool [10] [11].

3.1 Graphical User Interface: Path-Making for Interactive Authoring

The path-making process is supported by a graphical user interface. A media scheduling engine displays the results of each query as a real-time mixture of media resources. The ontological graph is visualized in the GUI as a 2D network of nodes and edges. Figure 2 shows the GUI as a collection of dynamic nodes; visible nodes represent a limited region of a much larger ontology. Several levels of interaction are defined. Mouse-over a node displays its concept name. Double-click on a node selects that node to generate a resource query and modifies the display to reveal all nodes that are nearest neighbors. Nodes remain hidden until a nearest neighbor is selected. Nodes are displayed with animated ball-and-spring dynamics, aiding visual identification of relationships. The "current location of the user" is defined as the most recently selected node. The square node is anchored; it is a member of an authored path.

The idea of making paths through a digital document space can be traced to multiple sources, including the Memex technology proposed by Bush in the late 1940's [12]. These proposals focus on "trails of documents" using text processing for cross referencing and indexing to achieve more efficient storage and retrieval. Our prototype differs from "trails of documents" proposals by implementing *paths of queries;* paths through concept space generating queries as acts of creative inquiry, recalling in real-time sequences of composite displays of resources, functioning both as dynamic media content and as semantic navigation feedback to an observer.

Figure 2 represents the current prototype interface for query paths. A vertical array on the left of the GUI is a sequence of concept nodes arranged as a path. The path can be traversed in order or explored out of order. In the main GUI a selected path member displays a square anchor node and a concept neighborhood. Any node in the concept neighborhood can be expanded for exploration. Clicking on anchor nodes

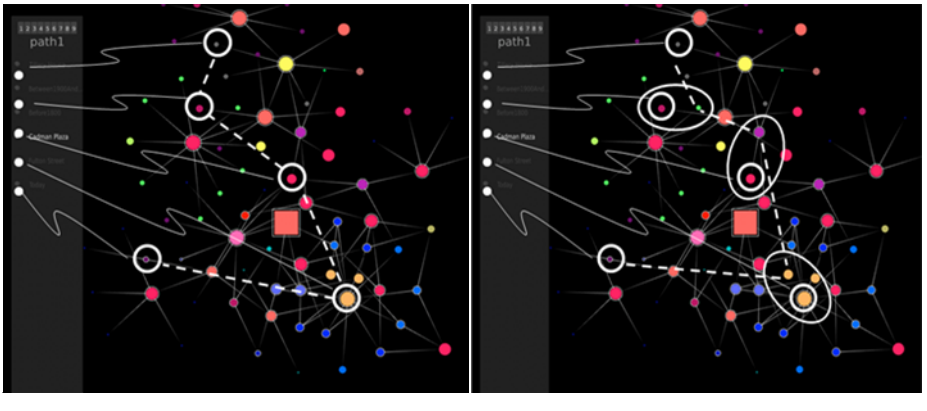

Fig. 2. Graphical User Interface representing concepts as interactive nodes. Size of node indicates number of links to a node. Color indicates subclass-superclass relationship. A Path of nodes is presented in separate vertical array on the left. **2A (left)**: White circles show graph positions of each path node; dotted lines indicate non-adjacent links across semantic regions. **2B (right):** A second narrative instance of the path utilizing alternate nodes in semantic regions of the original path nodes.

generates the queries selected by the path author. Clicking on a non-anchor node returns a related set of media resources and expands the graph visualization to reveal the concept neighborhood of the selected node. Exploration of neighboring nodes produces further queries returning related resources.

We advance the principle of paths of queries not only for initial creative acts but as the regular and normative use model for experience of Interactive Documentary. This is to say that an Interactive Documentary is generated and re-generated through path recall, and the distinction between authors and observers is fluid. An observer may follow a path verbatim or introduce variations, which can be fleeting or result in multiple path versions we refer to as *narrative instances*.

3.2 Interactive Media Use Case

To develop examples of nonfiction narratives, a set of media resources was gathered in reference to present-day and historical Brooklyn [13]. A tiled large-screen format is used to display 2D images and 3D scenes. The GUI is accessed at a small kiosk. To render a path, the system responds to queries by scheduling the concurrent display of sounds, 2D images, and 3D virtual camera movements.

In an example query path the first node is the concept "FultonStreet2000toPresent"; it returns photographs of storefronts, sounds of bus traffic, pedestrians and street vendors captured 2006-08 on Fulton Street, and a 3D camera movement slowly "flying" (tracking) along virtual Fulton street. Selecting a second path node while these resources are displayed, "BoroHall2000toPresent" introduces new photos and sounds, with smooth visual and audio cross-fades effecting the transition. The 3D camera movement interpolates from Fulton Street to a new position hovering above the model of Borough Hall. A subset of images and sounds may be common to both queries; these are brought forward in the display to effect smooth transition. The third path node "FultonStreet1880to1920," sends the 3D camera to resume a flyover of Fulton Street; the 3D scene now includes a model of the Brooklyn elevated train from the early 20th century, and postwar buildings are removed. Photographs of hip-hop shops and cell phone vendors are replaced by historical drawings, lithographs, and photos including images of the elevated train that were used as references for modeling its 3D graphics counterpart. Sounds captured on Fulton Street are replaced by sounds from a SFX library: horses, carriages, a steam engine, and pedestrians on a boardwalk.

Durations and transitions in each media display are qualitative and critical to narrative interpretation. Transitions are computed dynamically in reference to qualitative data of media resources, using metadata stored as properties of individuals. We introduce the term *Display Grammar* for heuristics of display signal processing and resource scheduling. We refer to *Display Prosody* as the application of heuristics with respect to qualitative properties of resources. These subsystems are well-developed for signal processing in computational performance media though not semantically well-formed for reasoning. We advance them as architecturally essential for interactive documentary.

3.3 Narrative Instances

A media archive is like memory without recall; media resources cannot encode their meaning as metadata. Archaeology (cognition) is required to determine or assert what

resources may mean. An author performs an archaeology reflecting upon what is recognized and drawing associations to account for what is missed; these constitute the author's perspective in associative aspects. Interactive recall is a retelling accompanied by retrieval of media resources. To traverse a path of queries is to recall the narrative that binds them and entails a performative quality shared by path authors and observers retracing paths.

Our system architecture and graphical interface are designed to facilitate performance aspects for observers to extend and vary the experience. Paths make leaps of narrative association to nonadjacent graph regions. A path member centers a semantic region; within a region there may be multiple nodes that support associative coherence. Multiple narrative instances are generated from a pool of neighboring concepts creating variations in a query path. Figure 2A shows the nonadjacent graph members' narrative associations in a query path. Figure 2B represents a narrative instance of the query path in Figure 2A.

4 Ontological Data Design for Narrative Association among Media of Multiple Types

Authoring applied to media of multiple types requires structured access to diverse media resources; it is desirable to develop uniform and extensible authoring procedures rather than tailoring a process for each media type. Ontological data design provides a means for designing uniform criteria for organizing media resources of multiple types.

Ontologies are relationships of concepts that describe media resources. "Describe" is a principle of set membership: concepts describe sets; resources are unordered members of one or more sets. "Relatedness" may be hierarchical in the form of set membership or non-hierarchical in the form of semantic associations. Navigation of an ontological data structure involves traversing subclass-superclass relationships and non-hierarchical associations. This flexibility is desirable for authoring both level-of-detail and narrative.

Concepts represent queries that retrieve media resources by assertion and by inference. Assertion is a direct assignment of set membership when a resource is entered into the data set. Inference is a computational evaluation that mines relationships to be discovered by an observer. Ontological reasoning combines both types of discovery.

4.1 Dual Root Data Design Relating Concepts to Individual Media Resources

Figure 3 summarizes the main components in the ontology: a dual-root node structure of Concepts and Describable Objects. *Concepts* describe media resources of multiple types. *Describable Objects* are classed as individual media *Resources*, or as *Content Objects*—entities depicted by resources. *Properties* designate metadata of individual resources or relationships between individuals, rather than between individuals and concepts. Media metadata is stored as Properties of individual resources, and Properties may store other types of keywords, labels and numerical data.

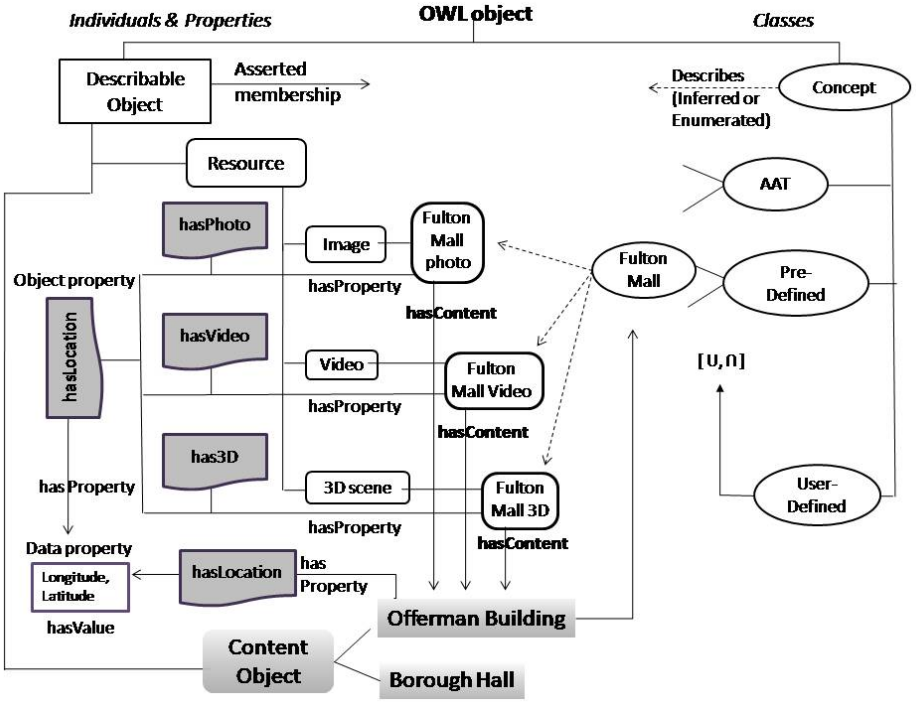

Fig. 3. A photograph, a video and a 3D scene are Resources inferred as members of "Fulton Street." The inference evaluates the GPS metadata encoded in the "hasLocation" Data property, and matches it with the location data of the "Offerman Building" Content Object that is independently asserted as a member of "Fulton Street." The Offerman Building is depicted in the three Resources.

Object Properties are extensible and inferable: they may refer to other properties; or to metadata common to multiple types of media resources. *Data Properties* are terminal, not extensible, and may be asserted only, not inferred. Figure 3 illustrates this relationship: several resources of different types have an object property "hasLocation," which links to the Data property "hasValue" where an asserted set of GPS metadata are stored. This data matches the location data of the Content Object "Offerman Building," an entity depicted in each of the resources. "Offerman Building" is asserted as a member of the Fulton Mall concept, and this enables the inference that the other resources are also members of "Fulton Street" even though their membership was not previously asserted.

Concepts may have relationships asserted to specific media resources. However many concepts have relationships only to other concepts. The dual-root design enables modifiable concept relationships without affecting the representations of individual media resources. By minimizing direct dependencies between concepts and resources, the concept graph can be changed without modifying resources. Reciprocally, resources can be changed without modifying the concept graph.

4.2 Structured Concept Vocabulary

A controlled vocabulary provides semantic constraints for resources depicting downtown Brooklyn. The Getty Art and Architectural Thesaurus (AAT) [14] provides both hierarchical and associative relationships among 35,000 concepts. We presently use about 3000 concepts in the ontology. The AAT serves as a semantic anchor for describing the built environment. However the AAT does not provide all concepts needed for structuring queries to support media authoring. For example dates independent of cultural and historical periods are not part of the AAT. So we complement the AAT with Predefined concepts.

Predefined_concepts are an application-specific vocabulary specified by the system designers to meet the narrative needs of projects. Predefined_concepts group media resources under complex logical relationships that are not easily represented as a graph nor easily manipulated using a GUI.

User-defined concepts are created on-the-fly by selecting available concepts and applying operations such as unions and intersections, or filters on metadata values. Combinations of predefined concepts and AAT concepts may be grouped in real-time while using the GUI to explore media resources. Examples of metadata filters in user-defined concepts include dates, GPS locations, polygon counts, and focal length settings.

5 Concluding Remarks and Future Direction

Documentary practice is growing beyond its roots in cinematic apparatus. The impulse may be recognized in blog journalism and social networks linked to web media archives. Distributed communities' discourses are decentralized, bringing oral traditions into a leading role. Models of media authoring can facilitate continuity in a distributed participatory documentary production. Continuity is a sense of experience through the assemblage of fragments and redundant structure when entered into an observer's sensory faculty. A garden variety of sensors, processing engines, and media displays may be brought into an interactive structure to support multiple modalities of an observer's senses and actions. In return an observer may actively seek to synthesize media processing layers. To produce coherent works, an authoring model can anticipate distributed collections of devices and processes for generating alternatives.

The described Interactive Documentary system facilitates observation as an act of narrative synthesis. Building upon the multimodal interaction paradigms illuminated in previous performance system research, the present creative pursuit often orients itself towards the narrative objective of intensification, the techniques to intensify the perception of events. The following design requirements facilitate such an objective:

1. Respect tempo: Temporality of media artifacts is surrogated in the tempo of the perception of the events while the events are assimilated by an observer's traces.
2. Empower memory: Respect for practice of oral tradition provides mechanisms for appropriating an observer's memory for generating new media experiences. It is necessary to place the function of memory as an active agent for generating new materials and experiences for both a human agent and system agents.

3. Automate subsystems: Access and analysis mechanisms for acquiring media resources and for streaming from archives into a display subsystem can be managed computationally. Automated subsystems can be optimized for cognitive priority and for system performance to deliver the responsiveness to observers' actions.

Future development depends upon assessment of interface design and computational performance. Feasibility of interactive recall as a vehicle of community narrative requires field study. Open contribution of media resources is desirable and can be facilitated by minimizing the complexity of data entry. The capacity to combine multiple structured vocabularies to define domains of ontology, suggests the value of investigating self-organizing methods. Finally, we look toward an extended repertoire of distributed interactive display devices that can host the documentary process, to enable interactive documentary work cycles across a wider community.

Acknowledgments. Thanks to Arthur Peters for API and software development, Robin Bargar for software design and project supervision, and Brian Macmillan and Marius Schebella for module implementations. Tatiana Malyuta contributed insights toward the potential of ontological reasoning. Robert Zagaroli and students provided archival documents of Brooklyn; Robin Michals and students provided photography; Mark Skwarek and students provided 3D models.

References

1. Barnouw, E.: Documentary: a History of the Non-fiction Film. Oxford University Press, Oxford (1993)
2. Choi, I.: Gestural Primitives and the context for computational processing in an interactive performance system. In: Battier, Wanderly (eds.) Trends in Gestural Control of Music. IRCAM, Paris (2000)
3. Bocconi, S.: Vox Populi: generating video documentaries from semantically annotated media repositories. PhD Thesis, Technical University of Eindhoven (2006)
4. Davenport, G., Murtaugh, M.: ConText towards the evolving documentary. In: Proceedings of the Third ACM International Conference on Multimedia, San Francisco, pp. 381–389 (1995)
5. Halliday, M.: Digital Cinema: an Environment for Multi-Threaded Stories. MSMAS Thesis, Massachusetts Institute of Technology (1993)
6. Davenport, G., Agamanolis, S., Barry, B., Bradley, B., Brooks, K.: Synergistic storyscapes and constructionist cinematic sharing. IBM Systems Journal 39(3&4), 456–469 (2000)
7. Choi, I., Zheng, G., Chen, K.: Embedding a sensory data retrieval system in a movement-sensitive space and a surround sound system. In: Proceedings of the 2000 International Computer Music Conference, pp. 141–144. International Computer Music Assoc., Berlin, Germany (2000)
8. Choi, I., Bargar, R.: Human-Machine Performance Configuration for Multidimensional and Multi-modal Interaction in Virtual Environments. In: Proceedings of the 4th Annual Symposium on Human Interaction with Complex Systems. IEEE Computer Society, Dayton (1998)

9. von Foerster, H.: What is Memory that It May Have Hindsight and Foresight as well? In: Understanding Understanding, pp. 101–131. Springer, New York (2002)
10. W3C: OWL Web Ontology Language Overview (2004), http://www.w3.org/TR/owl-features/
11. Protégé Ontology Editor and Knowledge Acquisition System, http://protege.stanford.edu/
12. Bush, V.: As We May Think. Atlantic Monthly (July 1945)
13. Choi, I.: idBrooklyn, an Interactive Virtual Media Presentation. In: Music and the Moving Image Conference. New York University (2008)
14. Paul, J.: Getty Trust: Art and Architecture Thesaurus (2005)