# An Object-centric Storytelling Framework Using Ubiquitous Sensor Technology

**Norman Lin**
ATR Media Information
Science Laboratories
Seika-cho, Soraku-gun,
Kyoto 619-02 JAPAN
nlin@atr.jp

**Kenji Mase**
Nagoya University
Furu-cho, Chigusa-ku,
Nagoya City 404-8603
JAPAN
mase@itc.nagoya-u.ac.jp

**Yasuyuki Sumi**
Kyoto University
Yoshida-Honmachi,
Sakyo-ku, Kyoto 606-8501
JAPAN
sumi@acm.org

## ABSTRACT

Using ubiquitous and wearable sensors and cameras, it is possible to capture a large amount of video, audio, and interaction data from multiple viewpoints over a period of time. This paper proposes a structure for a storytelling system using such captured data, based on the object-centric idea of visualized object histories. The rationale for using an object-centric approach is discussed, and the possibility of developing an observational algebra is suggested.

**Author Keywords**
ubiquitous sensors, storytelling, co-experience, experience sharing

**ACM Classification Keywords**
H.5.1. Information Interfaces and Presentation (e.g., HCI): Multimedia Information Systems

## INTRODUCTION

Previous work has developed an ubiquitous sensor room and wearable computer technology capable of capturing audio, video, and gazing information of individuals within the ubiquitous sensor environment[4 6]. Ubiquitous machine-readable ID tags, based on infrared light emitting diodes (IR LED's), are mounted throughout the environment on objects of interest, and a wearable headset captures both a first-person video stream as well as continuous data on the ID tags currently in the field of view (Figure 1). The data on the ID tags currently in the field of view represents, at least at a coarse level, which objects the user was gazing at during the course of an experience.

In this paper, we propose using an object-centric approach to organizing and re-experiencing captured experience data. The result should be a storytelling system structure based on visualized object histories. In the following sections we explore what is meant by an object-centric organizational approach, and present a structure for a storytelling system based on this object-centric idea.

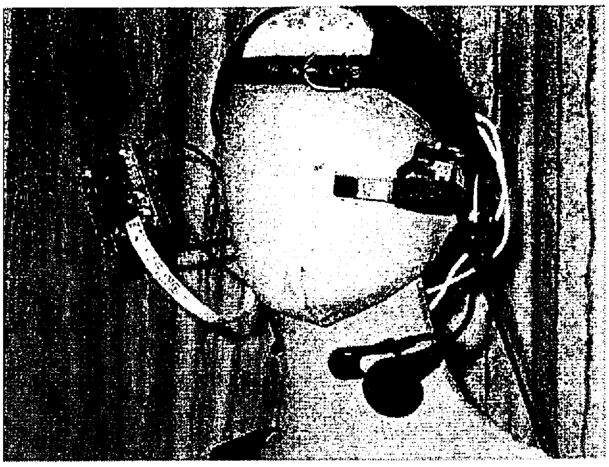

**Figure 1: Head-mounted sensors for capturing video and gazing information.**

## GOALS

The long-term goal of this research is to develop a paradigm and the supporting technology for experience sharing based on data captured by ubiquitous and personal sensors. In a broad sense, the paradigm and technology should assist users in (a) sharing, (b) contextualizing, and (c) re-contextualizing captured experiences. Ubiquitous sensors should automatically capture content and context of an experience. A storytelling system should then allow users to extract and interact with video-clip based representations of the objects or persons involved in the original experience, in a virtual 3D stage space (Figure 2).

## AN OBJECT-CENTRIC APPROACH

The central structuring idea is to focus on objects – physical artifacts in the real world, tagged with IR tags and identifiable via gazing – as the main mechanism or agent of experience generation. Other projects using objects to collect histories include StoryMat[5] and Rosebud[2]; also, [7] discusses the importance of using objects to share experience. By focusing on objects in this way, ubiquitous and wearable sensors and cameras allow the capturing and playback of personalized object histories from different participants in the experience. An object "accumulates" a history based on persons' interactions with it, and the ubiquitous sensor and capture system records this history. A storytelling system should allow playback and sharing of these personalized object histories. By communicat-

**Figure 2: Virtual 3D stage space for storytelling using visualized object histories.**

ing personalized object histories to others, personal experience can be shared.

### Storytelling: Visualizing and Interacting with Object Histories

Having captured video of personalized object histories, we would like to allow the interaction with those objects and their personalized histories in a multi-user environment to facilitate storytelling and experience sharing. Currently, a 3D stage space based on 3D game engine technology is being implemented (Figure 2). Within this 3D stage space, users can navigate an avatar through a virtual "stage set" and interact with video-billboard representations of objects in the captured experiences.

### Contextualization and Re-contextualization

Objects are typically not observed or interacted with in isolation; instead, objects are typically dealt with in groups. In terms of the physical sensor technology, this means that when one (tagged) object A of interest is being gazed at, another object B which is also in the current field of view becomes associated with object A. This is one form of context: object A was involved with object B during the course of the experience. The current working thesis is that an object-centric storytelling system should remember the original context, but should also separate or loosen an object from its original context. The reasoning is that by remembering but loosening the original context, we can both remind the user of the original context (contextualization) as well as allowing the user to reuse the object in another context (re-contextualization).

Concretely, for instance, consider the case where we have two objects, object A and object B, which are recorded on video by a personal head-mounted camera, and which are seen simultaneously in the field of view. Later, if the storyteller plays back a video clip of object A in order to share his experience about object A with someone else, the storyteller should be reminded in some way that object B is also relevant, because it was also observed or gazed at with object A. This is what is meant by system support for contextualization of experience. Essentially, the storytelling system serves as a memory aid to remember the original context of certain objects with respect to a personalized experience.

Re-contextualization, on the other hand, would involve using object A in a new contextstorytelluppose that a second user never saw object A and object B together, but instead saw object A and object C together. From this second user's personal perspective, A and C are related, but from the first person's personal perspective, A and B are related. By allowing video clips of A, B, and C to be freely combined in a storytelling environment, and by comparing the current context with pre-recorded and differing personal contexts, we allow the storytellers and audience to illustrate and discover new perspectives, or new contexts, on objects of interest. New stories and new contexts about the objects can be created by combining their captured video histories in new ways.

### Towards an Algebra of Observations

The object-centric idea presented above is that an object accumulates history, and that this object history is an agent for generating experience and an agent for transmitting experience to others. Part of this idea is that not only do individual objects have experiential significance, but also *groups* of objects carry some semantic meaning. A group of objects can be considered to be a "configuration" or a "situation" - in other words, a higher-level semantic unit of the experience. An object's history should be associated with the situations in which that object was seen. For example, the fact that objects A and B are observed together by a user means that object A is related, through situation AB, with object B. The situation AB is the higher-level grouping mechanism which relates objects to one another through their common observational history.

If we accept this, then, just as we can speak of the history of an object, we can also speak of the history of a situation. Just as we can relate objects with one another, we can relate situations with one another, or objects with situations. These situations can furthermore be grouped into even larger situations. This leads to a sort of hierarchy or continuous incremental experience structure, and suggests the possibility of developing an algebra for describing and reasoning about observations, situations, and higher-level groups. As an example of the kind of questions which such an observational algebra might answer, consider the case of three objects A, B, and C. User 1 observes objects A and B together, forming situation AB. User 2 observes objects B and C together, forming situation BC. Then, in the storytelling environment, user 1 and user 2 collaboratively talk about objects A and C together, forming new situation AC. What then is the relationship among situations A, AB, B, BC, C, and AC? Future work will explore this idea further.

### The Value of Context

By capturing the context of objects as observed, we provide for the later possibility to understand the original context of objects or object groups. We aim to answer questions of the following forms: In what situations was a particular object involved? In what situations was a particular group of objects involved? To what degree are other situations related to the currently chosen object or object group? By capturing context and defining a comparison metric, these types of questions can be answered.

The value of answering these questions is that it provides an intuitive, object-centric way of understanding, organizing, and telling stories about experience. It also provides a method of showing both strong and weak relations to other parts of the experience. The MyLifeBits project [1] also emphasizes the importance of "linking" to allow the user to understand context of captured bits of experience.

As an example, one can imagine a person who works as a home decorator, who uses a variety of furnishings meant to decorate the interior of a home. The decorator has built up an experience of creating several different configurations of objects in different situations. When trying to create a new decoration, it can be useful to try to group objects together (e.g. potted plant, shelf, and lamp), then see what previous situations, from the personalized experience corpus, have used this object group before. When illustrating to a client the decoration possibilities, the decorator, as a storyteller, could select candidate object groups and tell a story (by playing back related, captured video clips in a 3D stage space) about how those object groups have been used in previous designs. This also points out the value of using other persons' experience corpora, as it can provide new and different perspectives on how those objects might be combined.

The preceding example raises a subtle point not yet addressed, namely, that object *types*, and not just objects themselves, can also be important in classifying experience. In the above example, the decorator may be less interested in the history of one *particular* furnishing (e.g. one particular potted plant), but rather may be more interested in past related experiences using some *types* of furnishings (e.g. past decoration designs using potted plants in general). On the other hand, there are also situations where we are indeed interested in the history of one particular instance of an object. For instance, if a home decorator is involved with regularly re-decorating several homes, then within one specific home it can be useful to understand the specific object history of a specific furnishing.

## A STORYTELLING SYSTEM

Based on the previous object-centric paradigm for experience, this section presents a proposed structure for a storytelling system using visualized object histories. Core technology for this system is currently being implemented. The structure consists of five phases, each of which will be described separately. The five phases are capture, segmentation, clustering, primitive extraction, and storytelling.

### Capture

In the capture phase, a user captures an experience by wearing a head-mounted camera which records a video stream as well as continuous gazing data, in other words, which tagged objects were seen at any particular point in time during the experience. A tagged object can be either an inanimate artifact or another person; the only importance is that the object has a tag on it so that it can be recognized and recorded by the capture system.

### Segmentation

The goal of the segmentation phase is to break up the captured video data into chunks or segments which can then be compared (the comparison measure is discussed in the next section) and clustered. Two main approaches have been developed for the segmentation. The first approach is simply to divide the captured video into equally-sized segments. The second approach is to define an observation vector for each instant[1] of the captured video, and to cause formation of a new video segment whenever the observation vector "significantly" changes. The observation vector for an instant is the set of all objects observed by the user during that instant. Therefore, this second approach reasons that whenever the set of observed objects "significantly" changes, that a new situation has in some sense occurred.

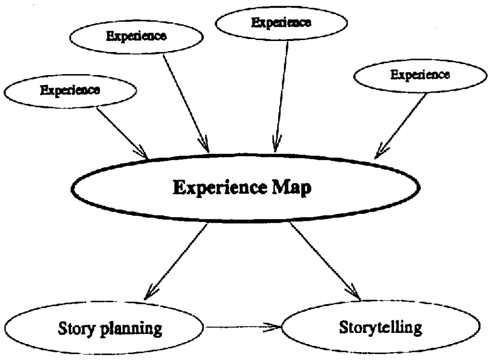

**Figure 3: Role of the experience map in the storytelling process. Objects are grouped into situations based on subjective observation. The experience map shows the situations and allows planning and telling a story about the objects in the situations.**

The reason that two approaches have been considered is that the two approaches tend to yield units (video clips) of different lengths. With the first approach, all resulting video clips are of equal, and short, length. With the second approach, resulting video clips tend to be longer; a new clip starts only when the situation changes. The first approach is more likely to uncover "hidden" patterns in the data because it imposes little structure on the data; the second approach introduces some sort of algorithmic bias, due to the more complicated decision on when a segment ends, but the hope is that this will yield longer, more semantically meaningful segments. The reason that longer video segments may be desirable is that they may serve as a better basis for extracting useful primitives which can be used in storytelling.

### Clustering

Given the segments from the previous segmentation step, the clustering phase compares the segments and clusters them together. The idea is that groups of similar segments form situations. Again, this is based on the object-centric organizing principle discussed earlier. For each segment under consideration, we first generate the observation vector over the entire time interval of that segment. An observation vector for a particular interval of time is a binary-valued vector representing, for all objects under consideration, whether that object was seen or not. Then, we use a clustering algorithm to compare the similarity of segments by comparing their observation vectors. To compare similarity of observation vectors, we have

[1] Technically, due to sampling issues, the observation vector cannot be measured instantaneously, but is instead aggregated over a small time-slice with an epsilon duration.

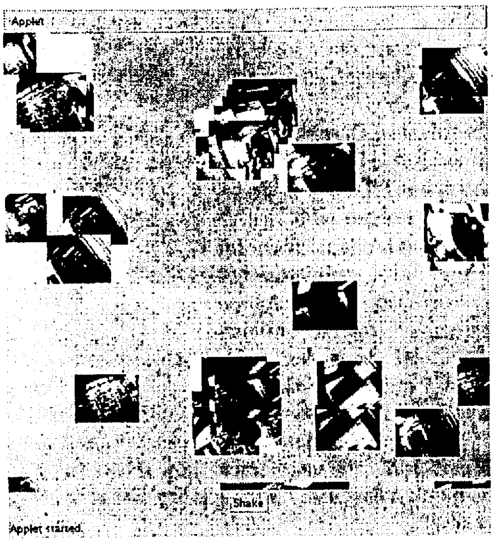

**Figure 4: A sample experience map illustrating clusters forming situations.**

chosen to use the Tanimoto similarity measure[3, p. 16-17], $S_T(a,b) = (a \cdot b)/((a \cdot a) + (b \cdot b) - (a \cdot b))$, with the $\cdot$ operator representing the inner dot product. This essentially is the ratio of common elements to the number of different elements.

Clusters represent situations in the original captured experience; they are groups of video segments, each involving a similar set of objects. By displaying the clusters on a 2D map, and by mapping similar clusters close to each other, we can create a "map" of the experience which can serve as a structural guide and memory aid during storytelling. Figure 3 shows the conceptual role of the experience map, and Figure 4 shows a sample interactive experience map created using magnetic attractive/repulsive forces.

### Primitive Extraction

Given the clusters from the previous clustering step, the primitive extraction phase aims to extract reusable video primitives from the situation clusters. By "reusable" we refer to the "loosening of context" discussed earlier. We aim to extract temporal and spatial subsets of the video which can be used in a variety of contexts to tell many stories relating to the original captured experience. The output of this phase should be a pool of video clips which represent object histories. This phase requires human intervention to decide which video clips from the situation clusters are representative of the experience and which have communicative value.

### Storytelling

In this final phase, a storyteller uses the video primitives extracted from the previous phase to tell a story to others. Within a virtual 3D stage space, video billboards representing the objects are placed in the environment and can be moved and activated by storytellers or participants in the space. Video billboards of objects can be activated in order to play back their object histories which were extracted in the primitive extraction phase. The current object configuration can be measured in the 3D stage space by generating an observation vector, just as is done in physical space during experience capture; this virtual observation vector defines a current "storytelling situation" which can be mapped onto the experience map to illustrate the storyteller's "location" in conceptual "story space."

## CONCLUSION

Given an ubiquitous sensor room capable of capturing video, audio, and gazing data, this paper described the use of an object-centric approach to organizing and communicating experience by visualizing personalized object histories. A storytelling system structure based on this object-centric idea was proposed. Core technology for this storytelling system is being developed, and work continues on gaining insight into a reasoning framework or algebra for observations.

## ACKNOWLEDGEMENTS

This research was supported by the Telecommunications Advancement Organization of Japan. Highly valuable contributions to this work were made by Sadanori Itoh, Atsushi Nakahara, and Masahi Takahashi.

## REFERENCES

1. J. Gemmell, G. Bell, R. Lueder, S. Drucker, and C. Wong. Mylifebits: Fulfilling the memex vision, 2002.
2. J. Glos and J. Cassell. Rosebud: Technological toys for storytelling. In *Proceedings of CHI 1997 Extended Abstracts*, pages 359–360, 1997.
3. Teuvo Kohonen. *Self-Organizing Maps*. Springer-Verlag Berlin Heidelberg, 1995.
4. Tetsuya Matsuguchi, Yasuyuki Sumi, and Kenji Mase. Deciphering interactions from spatio-temporal data. *ISPJ SIGNotes Human Interface*, (102), 2002.
5. Kimiko Ryokai and Justine Cassell. Storymat: A play space with narrative memories. In *Intelligent User Interfaces*, page 201, 1999.
6. Yasuyuki Sumi. Collaborative capturing of interactions by wearable/ubiquitous sensors. The 2nd CREST Workshop on Advanced Computing and Communicating Techniques for Wearable Information Playing, Panel "Killer Applications to Implement Wearable Information Playing Stations Used in Daily Life", Nara, Japan, May 2003.
7. Steve Whittaker. Things to talk about when talking about things. *Human-Computer Interaction*, 18:149–170, 2003.