Multimedia Systems for the Next Generation Museum

David Birchfield Gang Qian Hari Sundaram Frances McMahon-Ward

Arts Media and Engineering Program, AME-TR-2004-04

Arizona State University

e-mail: {david.birchfield, gang.qian, hari.sundaram, frances.ward}@asu.edu

ABSTRACT

Museums play an important role in our society by providing a forum for intellectual discourse and interaction. Despite this critical social role, museums today present artwork in a manner (accompanied by a brief text description) that has not evolved at the same pace as our digitally enabled society. In this paper we describe our work in developing a framework for the meaningful integration of digital technologies that will reshape the way in which museum visitors experience traditional artworks. We are developing an unencumbered video-based movement analysis system that allows a user to interact with the museum exhibit using natural movements and gestures. The multimedia system will respond by creating a context-aware, user centric, multimodal presentation. This project involves three facets: the development of appropriate multimedia context models, vision based interaction and analysis, and generative mechanisms for multimodal feedback in the museum context. We believe that this enriched environment will lead to a better understanding of exhibited artworks and will address the individual needs and desires of all museum visitors.

Author Keywords

Multimedia, generative art, generative music, context model, computer vision, HCI, digital museum.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Museums play an important role in our society by providing a forum for intellectual discourse and interaction. A museum allows us to reflect on the past and contemplate the world we live in as well as helping us to appreciate aesthetic value.

Museums today present artwork in a manner (accompanied by a brief text description) that has not evolved at the same pace as our digitally enabled society. In [11], the author makes several key observations about traditional museums: (1) The ideas behind an exhibit can change over time but most museums are not equipped to easily deal with this situation. (2) Most of the museum's permanent collection is not seen by the visitor (3) The museum does not offer ways to personalize the learning – e.g. a child as well as an adult see the same presentation although they clearly have different educational, presentational and interaction needs. This paper describes our work in addressing these issues in service of designing a more compelling experience for museum curators and visitors.

There has been emerging interest in developing environments (either virtual, or augmented) for the multimedia museum. Augmented reality systems (Tokyo University Digital museum [7] and the Meta-museum [8] use see-through head mounted displays (HMD's), as well as hand-held devices – in both cases, the user's head must be tracked to enable the system to render accurate graphics that register properly with the real world. While we can find examples of specific interactive exhibits there are no information design tools available to the curators, who can use the *same* hardware across multiple exhibits. There is also no formal strategy for content creation – i.e. how and what multimedia data needs to be show to the visitor to generate a compelling experience? How are these to adapt to the user context?

We are working on the meaningful integration of digital technologies that will reshape the way in which museum visitors will experience traditional artworks. We are proposing an unencumbered multi-modal interaction environment that allows a user to interact with the museum exhibit, using natural movements and gestures. The system will respond by creating a context-aware, user centric, multimodal presentation. We believe that this will lead to a better understanding of the artwork.

THE INTERACTION FRAMEWORK

We now describe the typical user experience. Before viewing the exhibition, the visitor enters a user profile that indicates personal background (eg. cultural and educational profiles), goals (eg. the user may want to learn about particular topics such as Impressionism,

Cubism, etc.). After entering the profile, she is given an active badge that enables the system to track her. We are sensitive to privacy issues regarding this tracking system, and note that participation is voluntary and collected information does not include any identifying personal details about visitors.

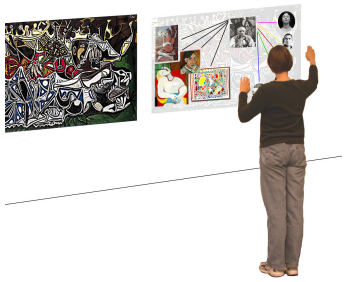

Figure 1: example user scenario in our proposed framework

When the visitor approaches a particular painting, the server updates the node that is driving the multimodal presentation associated with the appropriate user profile. The user can interact in real time with the augmented environment using natural gestures, movements, and poses to explore/understand/re-imagine the painter's life and work. The system presentation is multimodal (images, videos, text, and sound), context aware, and user-centric. The presentation (media and synthesis) is influenced by the current user context that includes previously viewed paintings and media presented to the user. The user context also includes the user behavior and the relationship between behavior and action. For example, a visitor may use a specific pose each time she is in long contemplation of a painting. Finally, once the user leaves the area of a painting, the user context is updated with media information – annotations, feature vectors, time spent, as well as behavior.

This multimodal interaction and feedback is critical to understanding a given artwork since many aspects of art, including style and context, are difficult to convey to a general audience that may not have the language or expertise to articulate such concepts.

USER CONTEXT MODELS

In order to effectively communicate appropriate experiences for individual museum visitors, we are developing novel multimodal user context models. The notion of context-awareness is key to both our vision and multimodal presentation systems. Current models of context [5], do not address the complexity of the information needs associated with the museum user context. There has been much work using the idea of context primarily in the area of context-aware / ubiquitous computing. The work there focuses on *application use* and hence it is primarily concerned with contextual information such as location, identity, activity and time. While this context-model works well in the context of applications in ubiquitous computing, it doesn't deal with *semantic* inter-relationships between concepts.

We have developed a graph based context model [12]. We use the dictionary definition of context as a starting point *– the interrelated conditions in which something exists or occurs* [1]. The formal model is defined using a concept-net – a graph where the nodes represent the concepts and where the edges represent the type of relationship (semantic, spatio-temporal, feature-level) between the nodes, we can also specify the strength of the relationship between the two nodes. A concept node is associated with a specific instance – an image / video / audio / text etc. We define the context be the union of concept-nets.

We plan on significantly expanding the scope the context model from prior work to include: (a) *Profile:* This is the user input data, and includes age, ethnicity, gender, occupation, education and interests. (b) *Media history:* As the user walks through the exhibit, the system keeps track of the media watched / heard by the user. (c) *Behavior:* The system will keep track of the user behavior – movements, poses, as well as control gestures (pointing, zooming, etc.). It will also store low-level associated with each behavior. (d) *Goals:* These are the specific concepts that the user wishes to acquire – e.g. the user may be interested in the life of Pablo Picasso.

As the user wanders though the exhibition, the numbers of concepts that the user acquires will increase. The effect of the concepts on the user in the current artwork depend upon two things – (a) the time spent on watching the artwork, as well as (b) the relationship (semantic, feature-level) of the artwork concepts to the concepts in the user context.

VISION-BASED UNENCUMBERED INTERACTION

The museum environment requires unencumbered human interaction, and we are developing a video based system for extraction and analysis of visitor movements. We are investigating the use of two modes of interacting with the multimedia exhibits, (a) active control and (b) passive interaction. This vision system will allow the general museum population to interact with multimodal displays through intuitive, natural actions. Significantly, this framework will allow both children and adults to navigate the information space.

Context information will be used explicitly in the design of vision sub-system to accomplish real-time, accurate, and robust movement and posture analysis. In the computer vision community, the posture and gesture recognition are treated separately. In our approach, we are investigating the two problems in a unified framework. We are modeling and tracking the movements of the whole body and body parts. Postures can be modeled by body parts joint angles.

The vision sub-system will provide us with real-time feedback that will be utilized for passive interaction and active control of the multimodal presentation. We propose to create functional mappings between presentation-update and results from vision-based sub-system that will effectively control the presentation interface in real-time.

MULTIMODAL INTERACTION

One important goal of this work is to provide museum visitors with an enriched experience that expands beyond the current norm of art exhibition where paintings are coupled with a brief text description. To this end, we are developing three different context-aware multimedia presentation modes for the museum visitor – (a) context exploration and presentation, (b) stylistic understanding, and (c) imaginative soundscapes. The first two modes are *active* – i.e. the movements / poses / gestures actively alter the display in real-time, while the last one is *passive* – the user can just stand and listen to the soundscape while viewing a painting.

Content Exploration and Presentation

We are conducting research on novel, context aware multimodal presentation schemes that adapt in real-time to user actions. The goal behind the context explorer is to allow the user to examine different facets of a painter's life and work through multiple presentation mechanisms, each optimally suited for a specific exploration. Current presentation techniques using storyboards, or video skims are nor adaptive to the content, or the needs of the user. The media presentation problem can be broken down into three sub-problems: (a) media selection (b) media presentation and (c) synthesis.

Media Selection: In the media selection problem, we are interested in selecting media that maximizes the information to be gained by the user.

Media presentation method: Given a media collection, and set of most relevant media elements, what is the optimal method to represent this data? This involves determining first a set of presentations to choose from as well as formulating an objective function that picks the optimal media presentation technique *from an a priori fixed presentation set*.

For example, let us assume that a user is interested in the work of Pablo Picasso. The context explorer will allow the user to examine the context of Picasso's life and work from a number of perspectives. A timeline presentation would enable us to examine the chronological placement of Picasso's life in relation to other artists whose lives overlapped with his life. Alternatively, if the user is interested in the geographical influences (e.g. influence African art on Picasso's development of Cubism), then a map may be optimal. We are developing media representation *method* $\mathbf{V}$ that characterizes presentations in terms of various attributes: semantics, discrimination variable, display media, and interactivity. Determining the optimal presentation is then dependent upon the relationship between the presentation $\mathbf{V_k}$ and the following – (a) semantics of the media elements and the (b) user profile – (interests, behavior, goals).

Media presentation synthesis: Media elements needs to synthesized within an optimal media presentation model, as the *media* and *method* do not specify an *effective presentation*. Like our earlier solution [13], we plan on developing *a joint* optimization framework that determines the correct media, presentation method and the synthesis simultaneously.

Stylistic Understanding

An artist's painting style is often the defining aspect of a work – often beyond the content of a particular painting. One means of understanding a painting style is through examination of the artist's brushstroke. Style can also be understood in the context of meta-level form and content. Repetition of shape, angularity, and subject distortion are style characteristics for artists like Pablo Picasso. This aspect is critical to understanding a given artwork, yet is an exceedingly difficult concept to convey to a general audience that may not have the language to articulate such concepts. A short text description is inadequate to communicate the complexities of style, and we are investigating appropriate means for the dynamic creation of experiences for museum visitors that are both informative and demonstrative.

There has been prior work that deals with the non-photorealistic rendering leading to rendering of painterly styles [2]. In particular techniques for texture-transfer results have good results as they can provide real-time synthesis – however, in many cases the texture transfer algorithm needs multiple passes, or the parameters of the algorithm need to be manually tuned. There has also been much work on auditory display to convey and navigate information [10]. Though such auditory displays are practical in the context of simple interactions with data, they are not intended or capable of musically illustrating meta-level information.

We are developing a real time, gesture based interface that enables users to understand paintings through a visually and sonically augmented exploration of painting style. In the visual domain, there are two problems – (a) development of a gesture driven interface that allows the users to explore the different painting styles, and (b) real-time, non-parametric artistic rendering, based on texture-transfer. In our approach we plan to couple the speed offered by the non-parametric methods with machine learning to determine optimal texture transfer parameters.

Our prior work has dealt with the creation of generative structures for music [3,4] which respond to user interactions. In the auditory domain we are developing algorithms that are capable of automatically generating music that captures the stylistic and formal structures of a given piece of art. Based on a visitor's interaction with the proposed style browser, we intend to musically convey information about painting style at two levels – the low level of the painting technique (eg. Brush-stroke, color, etc.), and the higher level of meta-level attributes (eg. Fluid motion, dark tone, violence, etc.). Secondly, we are using sound and music to convey meta-level

information about stylistic features such as tone, mood, and energy.

Imaginative Soundscape Feedback

Through the introduction of imaginative sound collages and soundscapes in the Next Generation Museum, we propose to enrich the experience of the museum exhibit and to encourage users to reflect on the time, place, and cultural references which are suggested by individual paintings. Through the use of sound we will sonically place museum visitors in different, reflective environments that both complement and contrast the visual context of a given painting.

Soundscapes and sound collages are capable of creating a strong sense of time, place, and reference. This has been documented in analytical work from musicians theorists [9] and ethnomusicologists [6]. Soundscape collages from the real world have been captured and studied, and have been artificially synthesized for artistic purposes [14]. However, most of these techniques require painstaking field recording or offline composition. Work such as [10] propose an approach to the automated generation of soundscapes, however are limited by the small sonic palette and the focus on information display.

We are developing generative mechanism to enrich the imaginative space of the Next Generation Museum through the inclusion of sound collages which will relate to the time, space, and references of pieces in the exhibit. These sound collages utilize information from the user's context model, but they do not require active control from the user. Rather, the sound collages are dynamically generated by the system and utilize a database of annotations, sounds, and concept networks provided by a museum curator for paintings contained in the exhibit. These objects might refer to particular subjects in the painting, refer to cultural aspects of the work, or might more abstractly relate to the connotation space of the painting. We propose to systematically explore the influence of soundscape attributes such as density, amplitude, the diversity of sounds, and sonic familiarity for users in the museum environment. By presenting museum patrons with these sonic environments as they reflect on a given painting, we hope to enrich and expand the ability of patrons to understand and imagine these works in a non-textual, intuitive sense.

CONCLUSION

Through our work described in this paper we are developing a framework that will allow museum visitors to experience exhibits of traditional artworks in an enriched information and imagination space. Through the implementation of a dynamic user context model, the museum exhibit will present information that is relevant to the particular needs and desires of individual users. Through the implementation of sophisticated generative mechanisms for multimedia presentation, the museum will greatly expand the communicative tools available to museum curators.

REFERENCES

[1] *Merriam Webster Dictionary* http://www.m-w.com.

[2] M. ASHIKHMIN (2003). *Fast Texture Transfer.* IEEE Computer Graphics and Applications **23**(4).

[3] D. BIRCHFIELD (2001). *Evolving Intelligent Musical Materials*. Second Iteration: Emergence, Generative Systems in the Electronic Arts

[4] D. BIRCHFIELD (2003). Genetic Algorithm for the Evolution of Feature Trajectories in Time-Dependent Arts. 6th International conferene on Generative Art. Milan, Italy.

[5] A. K. DEY and G. D. ABOWD (1999). *Towards a Better Understanding of Context and Context-Awareness*, Proceedings of the 3rd International Symposium on Wearable Computers, San Francisco, CA, pp. 21-28, October 20-21, 1999.

[6] B. GOODEY (1974). Images of Place: Essays on Environmental Perception, Communications and Education. Birmingham, England, University of Birminghan, Centre for Urban and Regional Studies.

[7] N. KOSHIZUKA and K. SAKAMURA (2000). *Tokyo University Digital Museum*, In Proceedings of the 2000 Kyoto International Conference on Digital Libraries: Research and Practice, Kyoto University, Nov. 2000.

[8] K. MASE, R. KADOBAYASHI and R. NAKATSU (1996). *Meta-Museum: Supportive Augmented-Reality Environment for Knowledge Sharing*, Proceedings of International Conference on Virtual Systems and Multimedia'96 in Gifu (VSMM'96), pp.107-110, Sep. 1996.

[9] S. MCADAMS and E. BIGAND (1993). Thinking in Sound: The Cognitive Psychology of Human Audition, Clarendon.

[10] E. MYNATT, BACK, M., WANT, R., BAER, M., ELLIS, J.B. (1998). Designing Audio Aura.

[11] J. SPALDING (2002). The poetic museum : reviving historic collections. Munich ; New York, Prestel.

[12] H. SRIDHARAN, H. SUNDARAM and T. RIKAKIS (2003). *Context, memory and Hyper-mediation in Experiential Systems*, 1st ACM Workshop on Experiential Telepresence, in conjunction with ACM Multimedia 2003,, Berkeley CA, Nov. 2003.

[13] H. SUNDARAM, L. XIE and S.-F. CHANG (2002). *A utility framework for the automatic generation of audio-visual skims*, ACM Multimedia 2002, Juan-les-Pins, France, ACM Press, 189-198, Dec. 2002.

[14] B. TRUAX (2001) *Acoustic Communication*