Brief Introduction to the Special Issue on Behavior Understanding for Arts and Entertainment

ALBERT ALI SALAH, Boğaziçi University
HAYLEY HUNG, Delft University of Technology
OYA ARAN, Idiap Research Institute
HATICE GUNES, Queen Mary University of London
MATTHEW TURK, University of California Santa Barbara

This editorial introduction describes the aims and scope of the special issue of the *ACM Transactions on Interactive Intelligent Systems* on Behavior Understanding for Arts and Entertainment, which is being published in issues 2 and 3 of volume 5 of the journal. Here we offer a brief introduction to the use of behavior analysis for interactive systems that involve creativity in either the creator or the consumer of a work of art. We then characterize each of the five articles included in this first part of the special issue, which span a wide range of applications.

CCS Concepts: ● **Applied computing** → **Fine arts;** ● **Human-centered computing** → *HCI design and evaluation methods; Interaction techniques; Interaction design;*

Additional Key Words and Phrases: Behavior analysis, interactive arts, human–environment interaction, visual arts, affective computing, social and nonverbal behaviors

1. INTRODUCTION

Techniques for understanding human behavior have been growing in prominence as an enabling technology for interactive intelligent systems. For example, the International Workshop on Human Behavior Understanding was initiated in 2010 with the aim of creating a dialog among researchers from different disciplines to investigate issues of computer analysis of behavior in a wide range of settings, incorporating research into sensing, modeling, and analysis of real and virtual behavior, as well as myriad applications. This special issue, the second part of which will be published in the next issue of *ACM Transactions on Intelligent Interactive Systems,* is composed of journal-length articles that focus specifically on the new challenges that arise when the behavior being analyzed involves engagement with arts and/or entertainment.

In this domain, the subject of the behavior understanding may be either a creator, such as a visual artist or a performer, or a person who engages with a (possibly interactive) work of art or entertainment system. These types of interaction can be supported in novel and interesting ways through real-time behavior analysis. The articles in this special issue deal with both sides of the problem. In accordance with the scope of this journal, the articles not only address the theoretical and technical aspects of their respective topics but also present and discuss the implications of relevant interactive systems.

When the application scenario involves creativity, it is less likely that the behavior conforms to known patterns and norms, and thus evaluation is less straightforward. This is one of the most important challenges for behavior analysis in creative settings such as arts and entertainment. The selected articles provide insights into how, in different application domains, novel and rich interaction possibilities are first broadly investigated and then turned into systems that can cope with such variation.

2. ARTICLES IN THIS FIRST PART OF THE SPECIAL ISSUE

2.1. Context-Aware Automated Analysis and Annotation of Social Human–Agent Interactions

The article by Baur, Mehlmann, Damian, Gebhard, Lingenfelser, Wagner, Lugrin, and André deals with problems of labeling and evaluating human social behavior and verbal/nonverbal social signals. The authors present the nonverbal behavior analyzer (NovA), which focuses on bidirectional interactions with a conversational agent. Even though the system is modular, the task of recording, analyzing, and recognizing human behavior in real time is challenging and resource-heavy. A full-blown system is not just about recognition of some social signal in a given modality; it requires tools for continuous analysis of social events, interpretation of scenes and contexts, and automatic annotation. The NovA system is a good starting point for creating interactive agents, even incorporating authoring tools for the user to model and conduct interactions, and it is made available online, with many open source modules.

2.2. A General-Purpose Sensing Floor Architecture for Human–Environment Interaction

If the goal is to understand behaviors of interacting humans, or to provide natural interaction with a system in an ambient setting, it helps to have an environment that is equipped with sensing capabilities beyond the traditional camera and audio array. Vezzani, Lombardi, Pieracci, Santinelli, and Cucchiara describe an architecture that uses a sensing floor for this purpose. A low-cost, resilient, and stable sensor provides pressure field data generated by moving people. These are integrated and used to detect locations and events (like jumping), whose recognition can be made use of in entertainment, virtual reality, and interactive arts applications. The article describes the technical details of the low-level data processing framework, as well as a worked-out example of an application integrated with Google Street View. An important contribution of the article is the way in which the affordances of a novel sensor are explored simultaneously on multiple levels, involving different sets of tools like hardware design, software-based action recognition, and user evaluation.

2.3. The VideoMob Interactive Art Installation: Connecting Strangers through Inclusive Digital Crowds

Grenader, Rodrigues, Nos, and Weibel present an extensive analysis of the challenging problem of understanding the relationship between public artworks and how people interact with them. In this case, the interactive art installation *VideoMob* serves as the focus of study. The development of various incarnations of the work made it possible to trace how users' interaction with the work changed over the different versions. In this

process, a significant amount of human interaction data was accumulated from which to analyze the relationship between the artwork and its audience. The installation enables strangers who interact with different parts of the installation in different locations to interact together remotely. The different iterations of the artwork enabled a study of how different settings, locations, and contexts influenced visitors' reactions to it. The social element in public interactive art is highly complex, but it represents a key issue in human behavior understanding in the context of art installations. In particular, the restrictions imposed by social norms can be dampened in the presence of a shared (and unusual) experience with strangers. The consequences for human behavior are relatively underexplored; this article is one of the few publications in which this phenomenon is analyzed.

2.4. Affective Analysis of Professional and Amateur Abstract Paintings Using Statistical Analysis and Art Theory

The appreciation of an artwork is known to be subjective, regardless of the artwork, its genre, and the time and place of its creation. Yet artists are believed to transmit an emotional message to their audience through the artworks that they create. This article focuses on the analysis of abstract paintings. Sartori, Yanulevskaya, Salah, Uijlings, Bruni, and Sebe employ computer vision techniques to quantify which parts of a painting evoke positive or negative emotions, as well as to predict, on the basis of an analysis of eye movements, on which parts of a painting people prefer to focus. A bag-of-visual-words model is trained, and the classification evidence is backprojected to visualize which parts of the paintings convey positive and negative emotions. Two datasets concerning professional and amateur abstract artworks together with their ground truth (positive or negative emotions) are used for evaluation: a professional one from the MART Museum and an amateur one from the deviantArt Web site. Although the two datasets are visually dissimilar, computer-based analysis demonstrates that similar characteristics are associated with the evocation of emotion: bright colors and smooth lines evoke positive emotions, whereas dark colors and chaotic texture tend to evoke negative emotions. In general, people tend to look longer at parts of a painting that evoke positive emotions.

2.5. In the Mood for Vlog: Multimodal Inference in Conversational Social Video

Conversational social videos contain rich multimodal behavioral information that is conveyed by the speakers and decoded by the viewers. In this article, Sanchez-Cortes, Kumano, Otsuka, and Gatica-Perez focus on the classification and ranking of mood impressions in social videos using conversational *vlogs* from YouTube. First, a dataset of conversational vlogs is labeled by external observers through crowdsourcing using 11 mood categories: *happy*, *excited*, *relaxed*, *sad*, *bored*, *disappointed*, *surprised*, *nervous*, *stressed*, *angry*, and *overall mood* (i.e., overall judgment of positive or negative mood). Both nonverbal and verbal behavioral features are then extracted from audio and video channels. The mood of the vloggers is estimated via supervised learning methods for classification and ranking. The reliability and correlation analyses conducted demonstrate that the problem is challenging, yet several mood categories can be inferred and/or ranked with promising performance. Although results tend to differ depending on the features and machine learning methods used, *happy*, *excited/relaxed*, *disappointed*, and *angry* are found to be easier to model using the extracted audiovisual features. On the other hand, *stressed*, *bored*, *nervous*, and *surprised* are among the most challenging moods for classification and ranking.