

RESPONSIVE PORTRAITS

Flavia Sparacino, Glorianna Davenport, and Alex Pentland

Introduction

Modern techniques for high resolution, still-image display offer new expressive possibilities for photographic portraiture and exhibition." Responsive portraits" challenge the notion of static photographic portraiture as the unique, ideal visual representation of its subject.



Editors are usually confronted with choosing ONE ideal portrait from a limited set of pictures which represent poses, gestures, and expressions which ALL contribute to defining the character. In our view the entire set of a subject's typical portraits should be kept for interactive exhibits.

A responsive portrait consists of a multiplicity of views whose dynamic presentation results from the interaction between the viewer and the image. The viewer's proximity to the image, head movements, and facial expressions elicit dynamic responses from the portrait, driven by the portrait's own set of autonomous behaviors. This type of interaction reproduces an encounter between two people: the viewer and the character portrayed.

The experience of an individual viewer with the portrait is unique, because it is based on the dynamics of the encounter rather than on the existence of a unique, ideal portrait of the subject.

The sensing technology that we used is a computer vision system which tracks the viewer's head movements and facial expressions as she interacts with the digital portrait; therefore, the whole notion of "who is watching who" is reversed: the object becomes the subject, the subject is observed.

Background

When compared to film, photography seems to carry an intrinsic narrative poverty, because of its static nature. In the case of portraiture, portraits are usually read not as stories but as symbols, short visual poems that describe a unique and immediate perception of reality. Moreover editing single photographs for magazines or exhibits can be a frustrating experience for the artist as it requires discarding a number of photographs which all contribute to define the story or personality of the portrayed subject.



Responsive Portraits seem to be filling some of these gaps by incorporating a story behind the photographic portraits, by letting the photographs tell the viewer their own story through the interaction. Here the meaning of a photograph is enriched by its relationship to the other photographs in the set, and to the story line attached to them.



The uniqueness of the portrait is instead transferred to the uniqueness of the encounter between the viewer and the portrayed character. In this sense the viewer and the artist cooperate in creating an artistic experience similar to that which happens in an exhibition gallery or museum. Another shift then happens which leads from the reproducibility of a work of art to the uniqueness of the experience happening in the exhibition gallery or art museum. Following on the Bauhaus concept of a "Modern Exhibition", exhibited art should not retain its distance from the spectator. It should be brought close to him, penetrate and leave an impression on him. It should explain, demonstrate, and even persuade and lead him to a planned reaction. In this sense exhibit design can borrow from the psychology of advertising.

The Responsive Portraits

A Responsive Portrait consists of a multiplicity of photographs virtually layered on a high-resolution digital display. The image which is shown at a given time depends on how the viewer approaches and reacts to the portrayed subject. An active computer controlled camera is placed right above the display. By using real-time computer vision techniques we are able to determine how close/far the viewer is to the portrayed character, her viewing angle, and we can interpret some of her facial expressions like smile, laughter, surprise, or disappointment. We then feed this information back into our system which uses a behavior-based AI technique (Media Creatures) to display the "response" of the portrayed subject to the viewer's approach.



Responsive Portraits are created in two steps. First the photographer goes on assignment and shoots an extended set of portraits of her subject in a variety of poses, expressions, gestures, and significant moments. We feel that is important that at this stage the artist concentrates on connecting with its subject and postpones editing choices to the next step.



Later, editing happens. In the case of Responsive Portraits the photographer can choose at this stage not only *what* the public will experience but also *how* it will be experienced. The artist can edit a set of pictures which map her experience approaching the subject. Otherwise she can choose another set which represents a landscape of portraits of a person which changes according to the point of view of the observer. It is important to notice that at this stage the artist does not do a final edit of what the viewer is going to

see. The artist only sets up the terms of the encounter between the public and the portrayed character by choosing a basic content set and a mapping.

Mapping is done by autonomous agent based modeling of content. In this work we make use of our previous research and implementation of Media Creatures [1]. Media Creatures are autonomous agents with goals, behaviors, and sensors. A Media Creature knows whether its content is text, image, movie clip, sound, or graphics and acts accordingly. It also has a notion of its role and "personality attributes".

Traditional digital content presentation uses passive content and a separate program that coordinates the presentation and creates a mapping between input and output based on the user's input. This model is analogue to that of an orchestra director who conducts musicians following a given score. In our view this leads to a fixed, repetitive mapping and limited interaction modalities. Behavior-based design adopts instead the "jam session" model where musicians, each with its own personality and instrument, meet to create a musical experience with no previous program or score. This interactive design approach implies that there is no separation between content and choreography of content. It leaves more space for interactivity with the public because of the improvisational nature of the experience.

When using this design strategy the metaphor for the interaction between the user and the virtual world is not that of an *exploration* but that of an *encounter* with a Responsive Portrait. By encounter we mean a two-way movement. One by the viewer in search of an aesthetical or learning experience and the other by the responsive portraits looking for someone interested in their story or performance. We have so far successfully applied this type of content modeling to build an Improvisational Theater Space with a Text Actor [2][1], an Interactive Dance Space [3], a City of News [4] which organizes information in a 3d architectural space, and a digital circus.



According to the type of chosen mapping we are gathering content and implementing three types of Responsive Portraits: 1. The Extended Portrait; 2. The Responsive Hologram; 3. The Photographic Essay.

The Extended Portrait maps single aspects of the personality of the portrayed subject to the "personality" of a Media Creature. Extended Portraits include: "The Chronological Portrait", which layers photographs of a person across time; "The Expressive Portrait" which sets up a communicative facial expressions game between the portrayed character and the viewer; and the "Gestural Portrait" which uses a wider framing of the subject, including hands, to engage the public in the interactive experience.

Responsive Holograms are portraits which react as a function of the viewing angle of the observer as certain well-known holograms. These holograms [5], [6] show a sequence of an action as the viewer moves her head horizontally across

the display. In our system the portrayed subject changes her pose/expression according to the observation point of the viewer. The metaphor here is that we tend to see people according to our own emotional and experiential perspective coordinates: as these coordinates change we acquire new knowledge and understanding of the people surrounding us.



Lastly the "Photographic Essay" addresses the challenge of letting the public edit a photographic narrative piece through the interactive feedback of distance from the subject, point of view and facial expression.

Although we have not as yet to date produced a public installation of this piece we would like to see if the proposed model of interactivity can generate a communication dynamics among the public. We are interested in observing not just how viewers interact with the responsive portraits but also if they exchange knowledge about ways of interacting with the portrayed characters or if they enjoy watching each other interacting with the photographs. Such a dynamics would certainly add a new dimension to exhibit design.

Interactive Techniques

In our view, as artists, designers and programmers of this art piece, the technical tools that enable the experience are an integral part of our work. This implies that the software tools used for the interactive interface (LAFTER), display (MEDIA CREATURES), content selection, and design of the mapping between content and tools, are given equal importance in the description and realization of this piece.

The interactive interface is a real-time computer vision system named LAFTER [7]. LAFTER is an active-camera, real-time system for tracking, shape description, and classification of the human face and mouth. By using only an SGI INDY computer it is able to provide a wide range of information about the person appearing in the frame, such as: the center of the bounding box of the head and mouth, the rotation angle of the face and mouth about the axis given by the standing body, size of face and mouth, distance of the viewer from the camera, head motion, facial expression recognition – the person is: surprised, smiling, laughing, sad, neutral.



The system runs at a speed which varies from 14 to 25 Hz on a 200MHz R4400 indy, according to whether or not parameter extraction and mouth detection are activated in addition to tracking.

To estimate the location of the face and the lips in the image, the LAFTER system makes use of 2-D blob features, spatially-compact clusters of pixels that are statistically similar in terms of low-level image properties. It uses examples of lip and skin pixels to build models of the probability distributions of each class in

color space. The distributions are modeled as mixtures of Gaussians and are estimated using statistical estimation techniques (EM algorithm).



Feature vectors are computed at each pixel by concatenating the (x,y) spatial coordinates and the color components at that point. These features are then clustered so that image properties such as color and spatial similarity combine to form coherent connected regions, or "blobs," in which all the pixels have similar image properties.

By training the general model on thousands of skin color samples, we have obtained a model which is valid for a broad spectrum of users (Indian, Asian, Caucasian, South American, etc.). In addition LAFTER uses adaptive statistical modeling of the blob features to narrow the general model, so that its parameters are closer to the specific user's characteristics.

Patterns of behaviors, e.g., facial expressions and head movements are classified in real time using Hidden Markov Models (HMM) methods.

LAFTER has been successfully used as the base for several different applications with hundreds of naive users in several physical locations showing extremely reliable and accurate performance.

Conclusions

Responsive Portraits challenge our notion of the photographic portrait as a unique image that captures the essence of the subject. By layering a multiplicity of images of the portrayed person on the same interactive display and offering a natural interactive interface and mapping modalities, an extended set of expressive communication abilities is available to the artist photographer. Also, through this artwork, new venues are described for the design of interactive photo exhibitions for galleries and museums.



Bibliography

[1] Flavia Sparacino, *DirectIVE: Choreographing Media for Interactive Virtual Environments*, Master Thesis, MIT Media Lab, 1996.

[2] Public Performance. Description in: Flavia Sparacino, Kristin Hall, Christopher Wren, Glorianna Davenport, and Alex Pentland. *Improvisational Theater Space*. The Sixth Biannual Symposium on Arts and Technology. Connecticut College, February 27-March 2, 1997.

[3] Christopher R. Wren, Flavia Sparacino et al. *Perceptive Spaces for Performance and Entertainment: Untethered Interaction using Computer Vision and Audition*, *Applied Artificial Intelligence*, vol. 11, n. 4, June 1997.

[4] Flavia Sparacino, Alex Pentland, Glorianna Davenport, Michal Hlavac, Mary Obelnicki. City of News. Ars Electronica, 1997.

[5] Lloyd Cross & Pam Brazier, Kiss I, 1974, multiplex hologram, MIT Museum.

[6] Patrick Boyd, Train arriving to the station, 1989, holographic stereogram, MIT Museum.

[7] Nuria Oliver and Alex Pentland. Lafter: Lips and face tracking. In CVPR97. IEEE Computer Society, 1997.