Enhancing Interface Design Using Attentive Interaction Design Toolkit

Chia-Hsun Jackie Lee, Jon Wetzel, Ted Selker
Context-Aware Computing Group
MIT Media Laboratory, 20 Ames ST., Cambridge, MA 02139
{jackylee ,jwetzel, selker}@media.mit.edu

Abstract

This paper shows how a software toolkit enables graphic designers to make camera-based interactive environments in a short period of time without requiring experience in user interface design or machine vision. The Attentive Interaction Design Toolkit, a vision-based input toolkit, gives users an analysis of faces found in a given image stream, including facial expression, body motion, and attentive activities. This data is fed to a text file that can be easily understood by humans and programs alike. A four-day workshop demonstrated that some Flash-savvy architecture students could construct interactive spaces (e.g. Eat-Eat-Eat, TaiKer-KTV and ScreamMarket) based on a group of people's body and their head motions.

Keywords: Attentive interaction, design toolkit, camera-based interaction, interactive spaces.

1 Introduction

Visual works of art are often sitting quietly inside galleries or museums. It is important to understand how people react to the artwork and provide feedback at the right time. In [Krueger, 1985], he presented an artificial reality approach for digital art installations in which cameras interact with the viewers. His system took racks of equipment and was tuned to a particular interaction. Can current technology make this easier? Traditionally, providing a dynamic interactivity space has been difficult.

Designing novel visual experiences usually involves understanding how people are paying attention. Since attention is a limited resource [Pashler, 1999], and exhibits in galleries or museums each have their own stories to tell, the way in which viewers perceive these exhibits needs to be designed carefully. Visual attention can be tracked and measured by understanding the patterns of eye gestures [Selker, 2004]. Attention-based augmentations can be deployed to create a digitally-switchable domestic environment [Bonanni, 2005]. In the past eye tracking has been used to measure attention. ScanEval [Weiland, 1998] is a toolkit that processed eye movement and provided a real-time attention assessment and data summary that could be used for a wide variety of purpose, including user interface design.

Monitoring a group of people's attention and behavior gives information about how they are engaged, and is helpful in to providing relevant visual or audio feedback. By providing a system of ways to understanding people's intention and reactions, artists and designers will be able to create works of art that can effectively engage people. The expense and effort to set up systems such as Seeing Machines faceLAB [Web/Seeing Machines] have had limited applications so far. This paper demonstrates a technique which takes available face/eye and head movement software from the Intel Open Source Computer Vision (OpenCV) libraries [Web/OpenCV], and creates a simple interface for Adobe Flash [Web/Flash], which non-programmers can use as a simple develop environment for augmented reality and interactive spaces.

Workshops and educational forums are often created to bring new kinds of techniques and technologies to other communities. The Computer Clubhouse [Resnick, 1998] was a rich environment where mentors, tools, and community made it into an experimental learning place. We brought the Attentive Interaction Design Toolkit to the Asian Reality Design Workshop [Web/Asian Reality]. The participants were undergraduate and graduate students from the departments of arts and architecture, and a few professionals in visual arts or industrial design. Together, we formed an environment consisting of students, designers, software tools (the Attentive Interaction Design Toolkit and Flash), and related computer resources (i.e. desktop PCs, WebCams, video projectors, and internet connection). We expected such a live and resourceful environment to motivate the participants. When participants got a sense of possibility, they could play with their own Big Ideas [Papert, 2000] easily.

After the four-day Asian Reality design workshop, four groups of participants demonstrated interactive installations. Three of them deployed the Attentive Interaction Design Toolkit as a human motion input interface. An exhibition and presentations was held on the fifth day of the workshop. Around two-hundred people came to the exhibition and major part of them had chances to experience the demonstrations.

2 Attentive Interaction Design Toolkit

We present Attention Meter as the Attentive Interaction Design Toolkit. Attention Meter is a Visual C++ program which, as its name implies, measure attention using camera-based input. In Figure 1, different levels of attentive engagement (i.e. passing by, glancing, standing and watching, reading carefully, and engaging) can be observed by monitoring people's behavior patterns using a camera. The input is a video stream from a camera mounted near or within the target of attention. The camera is positioned such that subjects attending to the target are looking almost directly at it, allowing the attention meter to analyze their attention based on cues from their faces.

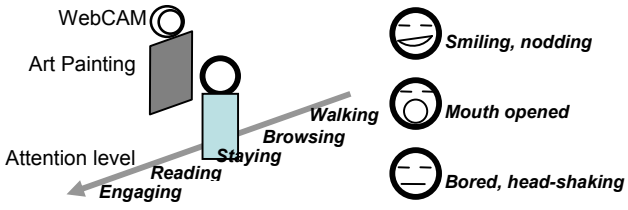

Figure 1: The system measures the attention level of people by using computer vision techniques to monitor their head movement, eye blinking frequency, and proximity.

Face Tracking

Attention Meter monitors and analyzes faces found in the camera view. Each frame taken from the video stream is run through a face detection algorithm from the Intel Open Computer Vision (OpenCV) library, as shown in Figure 2. This algorithm gives us the location and sizes of all faces in the image that are turned

towards the display. Tracking faces from frame to frame is accomplished simply by assuming that faces move very little from frame to frame, and then matching faces with the nearest coordinates within a small threshold. This method can be improved upon, but was found to be sufficient for Attention Meter.

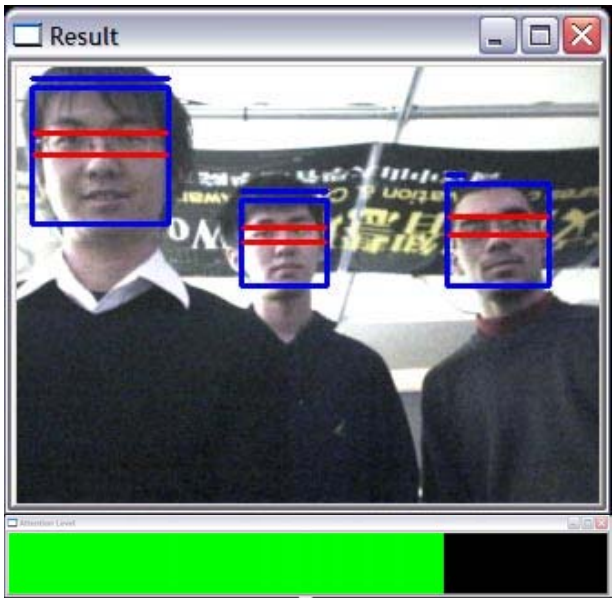

Figure 2: Attention Meter shows the group of attention level as a green bar and reasons human behaviors from head and eye movements.

Head Movements: Large Motion, Nodding, and Shaking

By keeping track of the position of individual faces from frame to frame, the Attention Meter can detect when faces are moving laterally with respect to the target of attention. Using a finite state machine to analyze sequences of small movements, the Attention Meter also recognizes the smaller gestures of nodding and shaking. A further improvement could be to incorporate size, allowing the detection of movement towards or away from the target as well.

Facial Expressions: Blinking and Mouth Position

By using basic knowledge of the structure of the face and looking for the distinctive brightness gradients of the eye, the Attention Meter can quickly find and detect eyes in faces, and over several frames measure the face's blink rate. One new feature in development is detecting expressions of the mouth. In a manner similar to the eyes, the position of the mouth is determined. This mini-frame containing the mouth is then passed to another algorithm (from Sluggish Software) to search for teeth and mouth shape, determining whether the mouth is open wide or smiling.

Attention Scores

Every face being tracked is given an attention score which varies over time. The score starts at 0 and increases up to some predefined maximum as the face exhibits more attention. The individual scores are then summed together to form a group attention score. Remaining still allows the score to increase, while lateral motion halts it. Nodding and/or shaking, moving closer to the target (becoming larger), and blinking less often (eyes visible more often) will also increase the attention score. In the future, expressions of the mouth will be factored in as well.

Various constants affecting the attention score calculation can be set by the designer using the Attention Meter's GUI at runtime. For example, the user may vary the maximum score or the rate at which score changes from blink rate and motion. The user may also change constants affecting the motion recognition, such as the number of times a face must alternate directions to be considered nodding or shaking. Thus users may further customize the Attention Meter to better meet their requirements.

High-Level Activity Recognition

The Attention Meter uses a series of Support Vector Machines (SVMs) [Web/LibSVM] to train and classify inputs to deduce high level information about the people it observes. For instance, facial expressions such as gasps, grins, and yawns can be inferred from the eyes and mouth data. These affects can then be used to discern emotions such as happiness, surprise, or boredom.

Motion can be classified into patterns indicative to the nature of the relationship between the subject and the target of attention. Using the motion data, these patterns of the faces can be classified as behaviors (see Table 1).

Table 1: Patterns of motion are deduced by Support Vector Machines (SVMs), based on values from the image processing in the Attention Meter.

Pattern of Motion	Characteristics
"Just Passing By"	High motion, rarely faces the camera
Casual Browsing	Face visible for a while, but stays in motion
Detailed Look	Face remains still for long periods of time

Combining affect, emotion, pattern of motion, the head movements of nodding and shaking, and the attention score in various ways will also allow us to determine high level activities about the relationship between the subjects and the target of attention. For instance, a low motion and blink rate may imply reading. Smiling and nodding infer agreement. Long periods of open mouths, shaking of the head, and a browsing pattern of motion imply someone was not impressed or even completely bored. The longer a person smiles, the more likely they find the target interesting. Also, the number of faces can be taken into account to get an evaluation the behavior of a group. In this way the Attention Meter can go beyond simply giving an attention score—it can describe the relationships subjects have with the target and with others.

Text Interface to Adobe Flash

The Attention Meter also outputs a summary of its collected data into a plain text file, which can be read by many other programs, including Adobe Flash. This data includes the group attention score, total number of faces, and for each individual face: coordinates attention score, size (proximity) and position, blink rate, and whether the face is moving laterally, nodding, or shaking. An example output might be:

```
wx=0&wy=0&attentionlevel=0&face=1&nodding=0&shaki
ng=0&moving=0&mouthsOpen=0&x0=44&y0=155&width0=55
&height0=55&face_attention0=0&face_age0=0&face_no
dding0=0&face_shaking0=0&face_moving0=0&last_blin
k0=1&mouthOpen0=1
```

A single function call in Flash will read these variable/value pairs into the local environment, allowing programmers to access the input data.

TCP-IP Output Interface

The Attention Meter also streams its data output to a TCP-IP port, so applications can use the sensor remotely over a LAN or the internet.

Limitations

The system needs to go through a calibration routine that includes standing within one meter to camera and inspect if the face track function works in this lighting condition. Distance and camera resolution are also large factors in overall effectiveness, particularly when it comes to analyzing features within the face. Mouth expression and blink detection are best at short distances and/or high resolutions (for example, a 320x240 resolution camera works well at distances under 2.5 meters). Blink detection is sometimes not possible, due to glare from eyewear. The final limitation is inherited to the current design—only a single image stream is used for input. In the future, input could be gathered from multiple sources, such as microphones, proximity sensors, or more cameras.

Empirical parameters

We defined activities and tuned the system parameters through experimentation. Moving means a face is detected and it move greater than 0.5 m/s (around 20 pixel/frame). A valid distance for the system works from 1 meter to 2.5 meter away from the camera

3 User Experience in Asian Reality Design Workshop

Workshops and educational forums are often created to bring new kinds of techniques and technologies to other communities. Asian Reality design workshop 2005, as shown in Figure 3, was used to test if flash programmers with only architecture backgrounds could make cutting edge interactive demonstrations in a few days.

Typically one week workshops will bring techniques such as using digital tools or new ways of thinking about future life or new experience. Also bringing together people and having them think and work together, meet new people and see new talent is often effective. Examples showed how these architecture students transcended their lack of technical background to create a real-time physical interaction with digital arts. In our case, a goal was to see if there is a new approach to design that can work for a group that has very little experience or background in creating computer interactive systems.

Figure 3: The *Attention* + *Interactivity* group in the Asian Reality design workshop.

The workshop required students to have basic Flash techniques. As such, these students were able to come in with a tool that they knew, but were presented with an approach and techniques as well as a new system that has not been available, that is recognizing a human face, its position, and motions.

The Attention Meter system was demonstrated as a toolkit for a four-day design workshop. 23 students who have design-related background (i.e. architecture, design, and art) without any formal training of computer science were divided into 8 groups for quick prototyping ideas. Students were given one three-hour lecture with tutorial for understanding immersive and interactive spaces and how to use the tool to integrate visual attention and multimodal interaction. Their assignment was to explore and implement interactive installations in the context of a night marketplace in Taiwanese culture. Three groups of students quickly integrated the Attention Meter system into their proposals, such as Eat-Eat-Eat, ScreamMarket and TaiKer-KTV. Students believe that they could build interactive that took human figure, shape, number of people into account to build interactive. In the course of this project three interactive working prototypes in big physical spaces with cameras and projectors were demonstrated after four days.

The value of these projects is that these people never had been involved in building prototypes to demonstrate technology and new ways of interacting with computers. They had been involved with classic flash kinds of interactions in which a cartoon or a button interacts. These research-worthy projects done in three to four days with three or four people, including the instruction, are striking.

4 First Example: Eat-Eat-Eat

Eat-Eat-Eat is a game for visually exploring food alternatives in a night marketplace, as shown in Figure 4. The system demonstrates that body motion and audio inputs can be mapped as an avatar inside the projected screen by using the Attention Meter. As moving around to catch the food dropping from the sky, the player needs hold a microphone to yell and speak the name of the food loudly to get the food eaten and counted into scores.

This game was designed under the context of the Taiwanese night marketplace which is full of food, gadgets, toys and clothes. Lots small restaurants and various kinds of food are people's typical impression of a night marketplace. People tend to have lots of different kinds of small dishes during that night. Eat-Eat-Eat collected 20 different typical Taiwanese small dishes.

The game starts after a player loudly speaks "I AM HUNGRAY!" The night market scene begins from small dishes dropping from the sky and moving around the screen. Based on the video input from a WebCAM, the player can control the avatar to move from left to right inside the screen. The player has to yell "EAT" or the name of the dish to catch them and count into scores.

Figure 4: A player is holding a microphone and moving her body from left to right to catch the food dropping from the sky. Around 30 people played this game and some of them do feel hungry after playing.

Eat-Eat-Eat system was well-implemented in Adobe Flash. The demonstration in the workshop was stable and compelling. Around 50 people had played this interactive night market eating

game. Visitors who played this game all agreed that they felt a bit hungry after seeing the delicious food photos and expressing their desire in eating by speaking the name of dishes loudly.

This game showed its interactivity that uses human body motion and voice-input in a context of night marketplace eating experience. Attention Meter allows tracking body movement as simple external input parameters in Flash. This team also implemented voice-input for multimodal interaction that enables players to yell and speak loudly to interact with the game.

5 Second Example: ScreamMarket

ScreamMarket is an interactive night-market show that interacts with audience's attention and feedbacks. This system demonstrated how audience engaged with the performance by monitoring their visual attention and audio feedbacks. The interactive show is implemented in Adobe Flash with Attention Meter as an attention-based triggering mechanism.

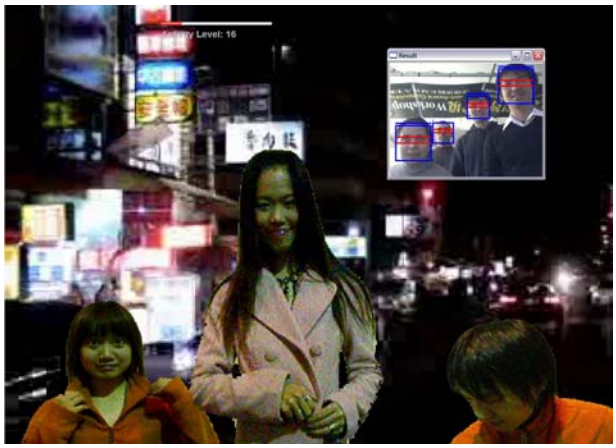

Figure 5: The Scream Market presents an animation of two Taiwanese girls if an audience is paying attention to the stage. When the crowd shows their interest and screams, the virtual girls dance and entertain them.

ScreamMarket transformed the Taiwanese traditional night market experience into a virtual and simulated space. In the beginning, an image of stage in night marketplace is blurred, but it gets clearer when the audiences pay attention to it. If more people are gathering in front of the stage, the dancers will show up, as show in Figure 5. The audience can yell to respond to the stage and get visual feedback.

ScreamMarket is implemented in Flash with Attention Meter. By using a microphone, as the volume of the audience increases, the performers become more active and entertaining.

The process of interaction is similar to the behavior that we watch the interactive show or bargain with the hawker in the night market. According to the method of interaction, the users are not simply viewers, but also performers in their own right. 30 people took turns in an exhibition interacting with the ScreamMarket. People were able to figure out and use it within a minute. It constrains output based on the regular environment noise, so that people may need to scream very loudly to interact with the Flash movie. The atmosphere of the interaction creates a realistic simulation of the night market.

6 Third Example: TaiKer-KTV

TaiKer-KTV enhances the interactivity of the performer and the environment for a more responsive and joyful karaoke space. TaiKer-KTV demonstrates how karaoke players engaged with the song can interact with the whole physical space based on their physical reaction and body movement.

Karaoke (KTV) is very popular in Asia for entertainment and social events. Karaoke TV might reduce people inhibition by focusing people on its screen dance. TaiKer-KTV extends this be requiring people to express themselves with own figures and body movements for on screen performance. KTV is presented from traditional karaoke context, called Tai-Ker KTV (TKTV), as shown in Figure 5, which exploits 'head-shaking dance' to enrich the environmental projection as a way to support group performance. The purpose is to amplify the group activity phenomenon in KTV and to create an interactive way to enhance the joyful and relaxing atmosphere as well as to enrich the KTV experience with fun.

Figure 6: TaiKer characteristics were implemented into this music Flash KTV, allowing people to influence the karaoke environment with their head shaking dance. TaiKer-KTV responds to the party interactively. Whenever people are nodding or shaking their heads, it enhances the visual experience in the party environment with rotating and blinking lights.

'Tai-Ker,' or 'Taiwanese-style guest' in literal translation, is one particular kind of culture on the lower civic level to which native rock stars claim to belong. In Taiwanese dancing circles, Tai-Ker's always have strong visual images and vivid outfits. They like techno music, patterned design shirt, black suits, white socks on black shoes, blue and white slippers, betel nut, and screaming while dancing...etc. Shaking and nodding heads along with the beat of the techno is a common part of the KTV culture.

In the implementation, an interactive music video was created in Adobe Flash and was projected on a wall. The lyrics go with a nodding head indicator to lead singing and dancing. The video is kept still and dull if it doesn't get enough attention, getting more animated only whenever people dance like Tai-Kers. Furthermore, the more people are engaged, the more special visual effects are applied. Some general rules have been defined for the techniques of the music video:

- If one moves his/her body, the image gets clearer.
- If one nods his/her head, image switches faster.
- If one shakes his head, the environmental light flashes more dramatically.
- If there are multiple people participating, the media elements (i.e. symbols, visual effects, texts and recorded

screaming voices) have an additive effect, creating a vivid mix of sound and imagery.

The T-KTV system contains a webcam, a video projector, Flash music video and the Attention Meter system, as show in Figure 6. The webcam is used to observe participants, and a video projector outputs the media for singer-machine-audience interaction. The contextual data, including number of participants, their attention, and whether they are moving, nodding, or shaking their head is interpreted by the Attention Meter system, determining the level of movements, especially for Tai-Ker dancing. A typical rave-party anthem-'Mei-Fay-Se-Wu' by Sammi Cheng is selected as the featured song. A Flash movie was implemented based on the song and receives interaction parameters from the Attention Meter, displaying an appropriate response on the wall projection with environmental visual effects.

This system was completely compelling. Around 100 people came up to it and immediately began making strong movements to make the people on the screen dance. Participants spontaneously tried to get others to join. The ease and success at creating a feeling of inhibition in the user was striking.

7 Discussion

Having a low barrier entry software tool and forming a learning community could help designers to quickly prototype ideas. The participants are mainly graduate and undergraduate students with architectural background. Most of them have Flash experience that they can make Flash-based visual arts quickly. Given instructions on using Attention Meter with Flash, they were able to pick it up quickly. The eating game group requested to track people's position individually.

The Taiker-KTV group were experimenting head motions. People tend to exaggerate their actions when the actions become ways of control, but the vision recognition system was implemented for normal actions like nodding or shaking head naturally. Over-exaggerated actions were usually not as effective as the normal ones. This obstacle was overcome through tuning of the user-side parameters. All groups were able to experiment and tune them so that they could find conditions that made the interactive experience consistent and successful.

The Attention Score made it possible to monitor the visual attention of a group of the audience. In the Eat-Eat-Eat example, they didn't use the attention score for its single player implementation, but they believe it could be useful in a multiplayer extension to their game. In the ScreamMarket example, designers used the scores to give different feedback depending on how many people are facing to the show. The performers are added on the stage if more people present. In the Taiker-KTV example, the overall group attention scores are used for determining the visual interaction between people and the screen. The visual effects get fancier and crazier when more people focus on the screen. Overall, the workshop shows that the Attention Meter's Attention Score can be particularly useful to designers who want to go beyond simply tracking movements.

8 Conclusion

The process of creating interactive art can be intuitive and accessible. Interactive techniques for computer graphics should not only belong to computer scientists. We present the Attention Meter system which allows novice graphic designers to quickly make interactive spaces, not only using analysis of human facial behavior, but also through a calculated measure of attention, the Attention Score.

This experience demonstrates the ability for modern tools to allow visual designers to make innovative arts based on new interface technologies in a very short time. Visual artists and designers usually have limited tools to develop art installations that interact with the audience. Eat-Eat-Eat, ScreamMarket and TaiKer-KTV were built upon the Attention Meter system and were all done in a four-day workshop. These three examples demonstrated the value and opportunity of giving visual designers good understanding and tools, so that even with limited technological background, they can still succeed to make interactive art installations.

The Attention Meter system shows the capability for a single camera to interpret attentive actions and transform the head, face, and eyes movements into computational models. It also demonstrates extensibility in interfacing with other software systems. The Attention Meter system can be extended with modularized sensors as a complete toolkit for designers to quickly prototype ideas. This toolkit demonstrates that research grade user interface tools can be put in a form to allow novices to use them in innovative ways

9 Acknowledgement

We thank Francis Lam, Yang-Ting Shen, Ian Jang, Ding-Han Daniel Chen, Yu-Chun Huang, Wingly Shih, Kristy Liao, Scottie Huang, Yu-Dang Chen, Sheunn-Ren Liou, Mao-Lin Chiu, Sheng-Fen Chien in the Asian Reality workshop 2005 in Taiwan.

References

BONANNI, L., LEE, C.H., SELKER, T., Attention-Based Design of Augmented Reality Interfaces, *Ext. Abstracts CHI 2005, ACM Press (2005).*

KRUEGER, M., GIONFRIDDO, T., HINRICHSEN, K., VIDEOPLACE- An Artificial Reality, *Proceedings of CHI '85*, pp.35- pp.40

PAPERT, S. 2000. What's the big idea: Towards a pedagogy of idea power. *IBM Systems Journal, vol. 39, no. 3-4.*

PASHLER, H., *The Psychology of Attention. Bradford Books. Reprint edition*, 1999

RESNICK, M., RUSK, N., AND COOKE, S. 1998, The Computer Clubhouse: Technological Fluency in the Inner City. *In High Technology and Low-Income Communities, pp. 266-286. Cambridge: MIT Press.*

SELKER, T., Visual Attentive Interfaces. *In BT Technology Journal, Vol 22 No 4, October (2004),* 146-150.

WEILAND, W., STOKES, J., RYDER, J., ScanEval - A Toolkit for Eye-tracking Research and Attention-driven Applications, *Human Factors and Ergonomics Society 1998.*

WEB/SEEING MACHINES- faceLAB, http://www.seeingmachines.com/

WEB/OPENCV- Intel Open Source Computer Vision Library, http://www.intel.com/technology/computing/opencv/

WEB/ADOBE FLASH, http://www.macromedia.com/software/flash/flashpro/

WEB/INTERNATIONAL WORKSHOP ON ASIAN REALITY 2005, http://arch.thu.edu.tw/ar2005/wor_eng_theme.htm

WEB/ LIBSVM, Support Vector Machines (SVMs), http://www.csie.ntu.edu.tw/~cjlin/libsvm