METAMORPHOSIS: Towards Immersive Interactive Film

Barnabas Takacs[1,2]
[1] Digital Elite Inc., Los Angeles, California, USA.
www.digitalElite.us.com , www.PanoCAST.com

Abstract

This white paper describes the methodology of a novel entertainment experience that combines the benefit of movie quality film production with interactive and immersive capabilities available on the DVD version or over the Internet. Our system is based on Panoramic Broadcasting *(PanoCAST) that utilizes spherical video to capture the entire scene and subsequently turn this footage into an interactive experience with the help of tracking and advanced interaction modalities, including virtual-reality-on-demand. We present a case study on a short film based on Franz Kafka's Metamorphosis.*

Keywords: Panoramic Broadcasting (PanoCAST), Interactive Film, Immersive Media, Virtual Reality On-Demand, Virtual Human Interface.

1. Introduction

Entertainment technology, in recent years, has been increasingly dominated by new forms of interactive experiences, such as games, that harness the power of computing power and place users in a world they control themselves. From the perspective of entertainment computing this global phenomena shifted high-tech interests away from traditional media, most notably film, and created a new wave of opportunities that focus on *shared experiences* and as a result on *social media*. To keep up its hi-tech appeal and unsurpassed visual and production quality the film industry invested in digital cinema *[DigitalCinema]* and most recently 3D movies. While these steps successfully brought audiences back to movie theatres around the world, the need to combine film, classically considered as a passive experience, with interaction still remains a challenge.

We propose herein a methodology that combines the benefit of movie quality production that result in a film projectable in theatres, with interactive capabilities invoked on the DVD version or over the Internet. Our system is based on *Panoramic Broadcasting* (PanoCAST) that during production captures a *spherical video* of the entire scene and delivers the above mentioned outputs simultaneously as the result of the same production process.

The remainder of this paper is organized as follows: First we briefly review panoramic broadcasting architecture our current solution is based upon. Next we discuss how to bring interaction to the realm of film without hurting the interest of artistic expression followed by a description of advanced interaction modalities. The methodology to turning panoramic film into a truly interactive experience using tracking and clickable content is followed by a brief case study and our conclusion.

2. PanoCAST – A Novel Architecture for Interactive Entertainment

PanoCAST or Panoramic Broadcasting *[PanoCAST10]* is a technology originally designed to deliver telepresence-based entertainment services over the Internet or mobile networks and thereby allow viewers to experience the very feeling of being somewhere else from their physical location *[Takacs07]*. The basic concept of the architecture is as follows: We first capture a stream of high fidelity spherical video (often called immersive media) with the help of a special camera system with six lenses packed into a tiny head-unit *[POINTGREY10]*. The images captured by the camera head are compressed and sent to our server computer in real-time delivering up to 30 frames per second, where they are mapped onto a corresponding sphere for visualization. The basic recording and server architecture, then employs a number of virtual cameras and assigns them to each viewer thereby creating their own, personal view of the events the camera is capturing or has recorded. The motion of the virtual cameras is controllable via TCP/IP with the help of a script interface that assigns camera motion and pre-programmed actions to key codes on the mobile device. The host computer then generates the virtual views each users sees and streams this information back to their location using RTSP protocol. This concept is demonstrated in Figure 1 in the context of a rock concert. The spherical camera head (left) is placed at the remote site in an event where the user wishes to participate. The camera head captures the entire spherical surroundings of the camera with resolutions up to 3K by 1.5K pixels and adjustable frame rates of maximum 30 frames per second (fps). These images are compressed in real-time and transmitted to a remote computer over G-bit Ethernet connection or using the Internet, which decompresses the data stream and re-maps the spherical imagery onto each viewer's display device locally.

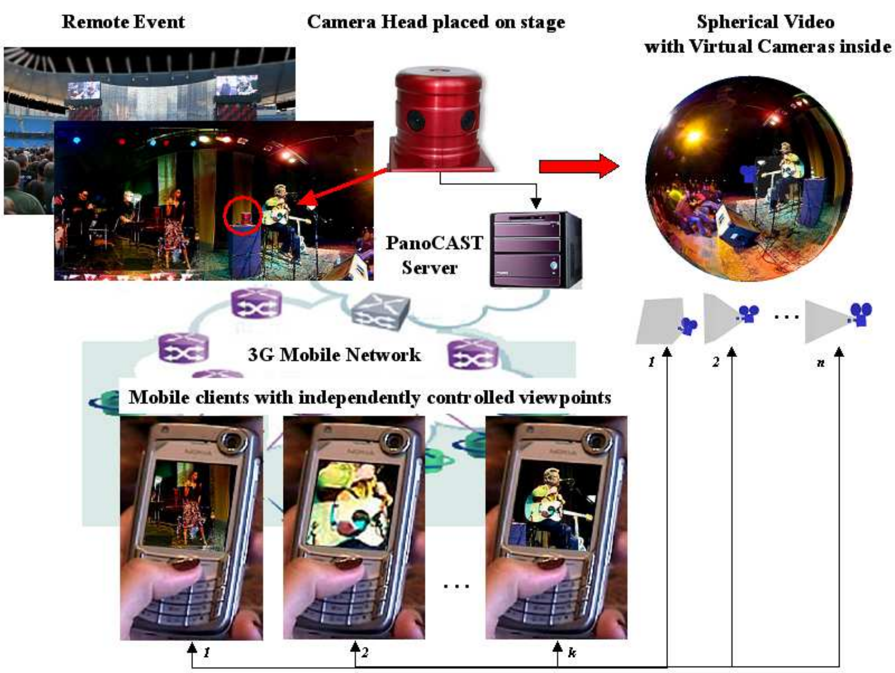

Figure 1: Functional overview of the Panoramic Broadcasting (PanoCAST) system for interactive entertainment.

The key idea behind the *PanoCAST* architecture, as opposed to just simply streaming the entire spherical video, is to distribute "just-in-time" views created by the virtual cameras. This forms the foundation of the proposed interactive film technology as discussed in the following sections.

3. Film vs. Interactive Experience – Retaining Artistic Intent

For many directors, photographers and camera crew, *interaction* is a word that goes against the artistic intention of the creators. Specifically, the very concept of a frame that can guide our eyes to bring our attention to the smallest details as well as showing sweeping overall views that spur engulfing emotions, allow artists to include what is of momentary interest and exclude everything else. By this definition reality is "boring" and artistic talent is precisely what allows us to focus us on what is truly important. If we simply allowed viewers to look around (interactively) how would this be possible?

To answer this challenge one is required to bring *interaction* into the picture without interfering with the artistic process itself while still serving the underlying message of the creators themselves. To achieve this goal we set up a two-stage production where in the first leg we record the entire film sequence using a panoramic video head but used almost as if it was a regular film camera aimed directly at the scene. The camera crew watches the output of one of the six camera modules to see the composition and the action, but in the second leg, during post production, we sit down with them again to create a second pass using virtual cameras placed inside the spherical film itself. It is at this stage where the final shape and look of the film takes place. By aiming the virtual camera at different portions of the spherical recording, changing the field of view and moving the virtual camera the director decide upon the final imagery that will eventually be rendered in high resolution (4Kx4K frames) to be finished in the film lab and printed on film. Due to the spherical nature of the recording, however, if needed the field of view may be extra large thereby introducing new visual ways of representing scenery or the emotional as well as the mental state of a character. The result is stunning imagery that is iconic and instantly recognizable, yet it can be projected in theatres using the traditional channels of film distribution.

During a film shoot the neither the director nor any member of the staff are present on the scene. Only a remote controlled robot moves the camera and interacts with the actors. Thus, to see how good the take was, on-set the director uses a virtual reality HMD to place himself in the scene and to see if he/she got what wanted from the actors. This pre-visualization tool is based on the Virtual Human Interface (VHI) module of the system *[DigitalElite10]*. Off-set, during post production, the director works with physical camera controls (*KORG Nano Kontrol*) that use *MIDI* interface to communicate camera commands to the rendering system (see Figure 2). This allows natural interaction and quick review of the different visual possibilities offered by the *PanoCAST* system.

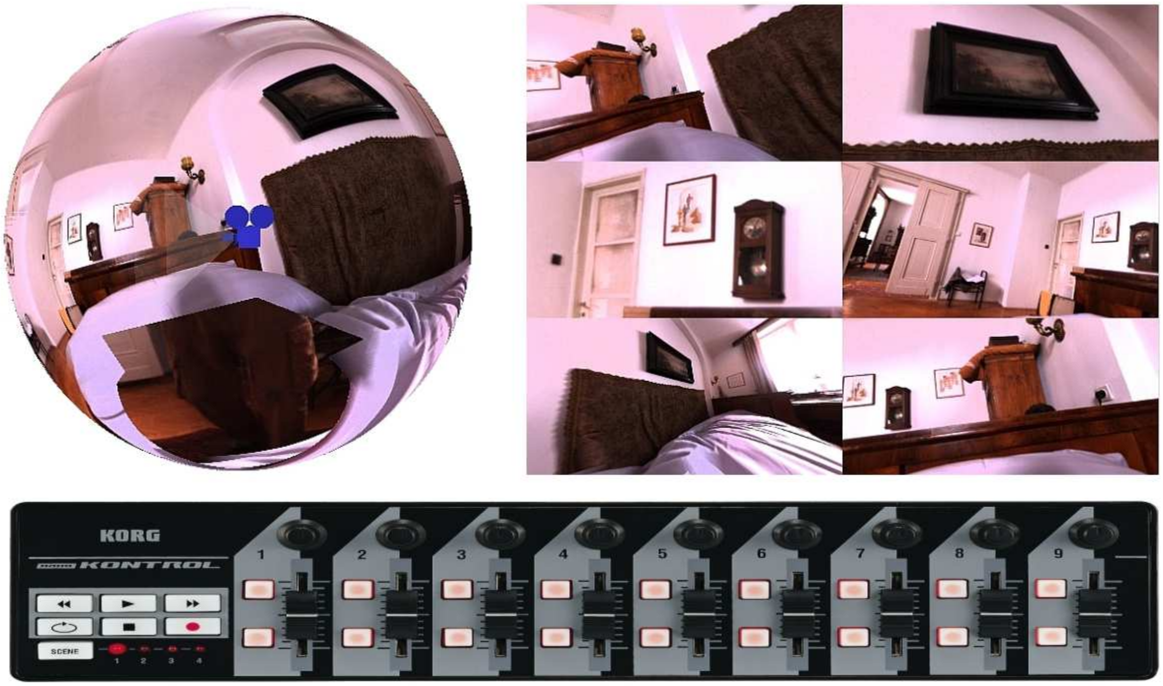

Figure 2: The PanoCAST pre-visualization and editing system uses MIDI-channels to control a set of virtual cameras with the help of physical controls. Upper left: The spherical image mapped shown from outside with the virtual camera in the center. Upper Right: Six independent camera views of the same scene shown on the left. Below: KORG Nano Kontrol used to direct multiple virtual cameras and edit key frames.

While in most cases the panoramic camera was moving constantly on the set with the help of a remote controlled robotic camera (see sections below), in some instances the entire output of the spherical lenses was used to produce the final shot. This is demonstrated in Figure 3 showing how the final scene, called "bubble sequence" in the film was made.

To create an immersive interactive experience using the same panoramic footage that complements the film and provides additional insight we created a spherical video player that may run on a computer installed locally or accessible via the Internet *[Metamorphosis10]*. In this mode viewers are allowed to control camera parameters interactively and individually thereby exploring the content of the film in a different manner. However, to represent the original concept and the artistic intention of the creators the same *virtual camera track* used for rendering the movie version is retained in the interactive mode as the default behavior of the viewer. Thus, in passive mode, i.e. when viewers do not want to take control they see exactly how the film was originally meant to be seen, yet when they decide to turn their attention to somewhere else, they can literally "enter that world" and feel as if they are in the very center of the action *[PanoCAST10]*.

Figure 3: Creating the "bubble sequence". A static camera surrounded by actors (upper left) was used to create a stretched full circle panorama image (upper right) and subsequently mapped onto the surface of a deforming sphere through the wall of which the camera passes (lower left) to reveal the soap bubble floating in space (lower right).

3. Advanced Modes of Interaction – Virtual Reality on Demand

The PanoCAST system offers several interaction modalities, the first one of which is the viewers' ability to freely look around. While the film rolls the motion of a virtual camera assigned to them may be controlled by each viewer independently using the mouse, the keyboard or even advanced game controllers, like the *Wii* or tracked orientation data obtained from a *head mounted display* (HMD). It is this later mode, that virtually allows users to enter the film and feel almost as if they were there together with the actors. The on-line version of our system has been implemented in Adobe Flash. To handle input streams from a variety of I/O streams we created a tiny program (called *VHI DeviceManager*) that upon installation connects to all available physical sensors (see webpage for a list of supported devices *[PanoCAST10]*) and represents this information to the player for interactive purposes. The simplest of this interaction is change in head orientation (i.e. looking around). More complex rules are defined in a *dynamics* descriptor file that defines the onset and offset of *triggers* via simple sets of rules and applies these rules to the scene accordingly. Triggers may detect simple click over a region of interest in the image (see section on *Clickable Content*) start events and even physical simulation to seamlessly integrate animated 3D models into the original recording.

Therefore we argue that the infrastructure that forms the basis of on-demand virtual reality and novel forms of interactive entertainment over the Internet as well as on mobile platforms already exists and may be utilized to create *interactive films* (Figure 4). Specifically, with the advent of low-cost computer accessories, peripheral devices and even HMDs with built in trackers that cost below $500 *[Takacs08]* inspired by and deployed in the use of VR offers more intuitive interfaces and applications areas than ever before *[Takacs05, Takacs06]*. Thus, our *Panoramic Broadcasting* system combines three key elements which, a new form of interactive film that advances the state-of-the-art and radically changes the way we entertain ourselves and consume media. We extend the boundaries of video-on-demand to *Immersive Intreactive Reality (IIR)* where not only the content, but also the personalized point of view of the user may be changed interactively. Moreover, we move away from 3D models and traditional VR and use i) *spherical film* in combination with ii) tracking technology to create *Clickable Content*, and iii) *physical simulation* for advanced interaction. This is the subject of the following section.

Figure 4: User experiencing full VR immersive interactive reality programming over the Internet using our PanoCAST player, VHIDeviceManager, and built in Physics Engine implemented in Adobe Flash. A range of additional motion sensitive devices, such as a Wii controller or even IPhone/IPad may also be used to control the view point and click on elements in an IP-TV setting.

4. Clickable Content and Tracking Interface

To make interactive film more exciting and appealing to large audiences a set of extra features needed to be developed that offer more than just the mere ability to look around and being in the center of the action. First, we wanted to allow viewers to learn more about the film itself, being able to click on actors and recall the role of that character, the plot itself ort any other linked information that may be of value and interest. To do this, we developed a pipeline to produce what we call *Clickable Content (CC)*.

In simple terms *CC* means that whenever the user clicks on the scene the rendering engine "fires a search ray" from the viewing camera and orders each visual elements as a function of their distance along this ray to find the closest one. Then the algorithm returns the object's name, the exact polygon it is intersected at, and the texture coordinates of the intersection point itself. This information allows the system to assign high level actions, such as displaying a callout image and/or text message when clicking on an object of interest in the interactive film or alternatively to directing the user to a website where further information may be found. For easier handling individual texture coordinates can be grouped by regions and tracking algorithms are used to annotate each element of the scene during a broadcast. The final output of such algorithms is a set of tracked regions with actions, visual and text information and web pages assigned on each of which *PanoCAST* viewers can click on during a film show and learn more about the event or simply educate themselves.

Tracking may be difficult as lighting often changes drastically, thus features do not necessarily retain their low level image characteristics. In addition, tracking on the live plate is also hampered by the slow frame rate (25 fps for film) and the fast movements characteristic of the particular film we at hand. Furthermore, since both the Panoramic camera as well as the actors may move in any given scene, we needed to combine *automated tracking* to insert key frames and filter the motion of touch-sensitive panels with post-production tools to correct the output of these algorithms. The automatic solution employs a modified version of our hierarchical tracker (see *[Takacs99]* for more details)**.**

Low-level tracking was implemented using a variety of algorithms including region-based tracking algorithms, such as normalized correlation, texture-based trackers, and optical flow-based estimators. As mentioned above, due to the lack of stable image context low-level trackers exhibit serious limitations in a real-world environment therefore the use of higher level implicit structural models always requires for accurate registration. In our case, what makes this process possible is the large tolerance to errors, as our goal is not to track the tiny details of each actor, but rather to create panels (generalized rectangles) that cover large areas and move along with the actors. So to overcome these problems a hybrid tracking algorithm is used to combine the advantages of each low-level method while minimizing the overall sensitivity to noise and variations in imaging conditions. We start with a *Point Tracker and* adaptively select the best method for each tracking point in a given frame in the region selected by an operator. The image is then processed multiple times and the results are integrated, minimizing the error and creating a smooth transition. As stated above, there is not much emphasis placed on achieving 100% accuracy in the first iteration. As time goes by and the tracking process progresses, more and more evidence emerges supporting a consistent interpretation of the underlying data. Higher level 2D/3D trackers (driven partly by the known motion of the camera) use these raw inputs and impose a simple geometric model on the tracking process and subsequently correct for inaccuracies from lower layers. This geometric correction layer operates on top of region-based, *Gabor jet*, and *optical flow* tracking algorithms. The second layer in this architecture implements a set of s*tructural trackers* that, instead of following individual points in the image plane take advantage of an underlying model geometry to which these (individual) points belong to. Thus, one can track point groups, 2D structures, and/or splines as well as true 3D objects imported in the form of a geometry description file.

One advantage of tracking in a panoramic video is the system's ability to keep objects of interest always in the center of the camera. In other words, when following a single point in the scene, the context and window size in which the tracking algorithm operates is kept constant by moving the virtual camera's target proportionally to the displacement measured on that feature. In other words, the when the raw tracking algorithm outputs a displacement in pixel space, that information is mapped onto the 3D geometry of the spherical video and its relation to the virtual cameras relative geometric arrangement and the target of the camera is displaced such that the feature being track remains is brought back to the center of the image. As the process continues this allows our solution to employ larger than usual tracking windows without explicitly loosing context around the edge of an image.

The output of this process is a set of "invisible" panels the viewer may click on to get access to more information. Tracked panels are *named* and as such they may also be selected from a list and used to follow any given person or object with the help of the virtual camera. When selecting this mode viewers may enjoy looking always at the same actor or actress irrespective of the original camera track. Finally, we briefly mention that the Internet version of our system supports a multitude of Community features as well, thereby allowing on-line users to discuss and share their experiences. As an example, a *Chat* interface was created based on the scene the viewers are looking at and general statistics on camera moves, clicks and interaction data is collected anonymously for subsequent analysis and user characterization. By recording this information it also enables the system to generate a novel camera track that reflects what the majority of on-line viewers were most interested in to look at. This voting scheme or similar alternatives may also be used in large venues during interactive presentations of the film itself, thus maintaining the original concept of the artists but bringing an element of audience participation in as well.

Figure 5 shows a screenshot of the *tracking interface* and *callout editor* in on-line mode. With the help of these tools production staff may track visual elements on-line, review and edit keyframes and assign callout information in multiple languages. When viewers view the same scene, a tiny hand appears discretely above clickable areas encouraging them to learn more if interested. Once these tracks are added they may be referred by name and viewers can direct their virtual camera to follow e.g. "Grete" in all scenes. For more details the interested reader is referred to *[Metamorphosis10]* or for a similar interactive cultural application *[VirtualMuseum09,VirtualMuseum10]*.

Figure 5: Left: On-line tracking interface to create clickable content and manage callouts in multiple languages. Right: Callout editor interface, below Callout as shown in the application.

5. Kafka's METAMORPHOSIS

To validate our concept of interactive Immersive Interactive Film we produced a short film based on Franz Kaffka's famous story, Metamorphosis *[Metamorphosis09]*. The original novel tells the story of a traveling salesman who one day waking up finds himself turned into a insect *[Wikipedia10]*. While the spherical nature of the camera head is ideally suited to the subject of the film, the key challenge we first faced was to maintain an extra large field of view while excluding everything unwanted in the picture. More specifically, because panoramic video records and maps the entire 360 scene with only a minimal blind spot, production crew, lighting and stuff, even the director needed to remain hidden entirely.

To achieve maximal visual coverage and also to create an extremely low- angle view of the set as envisioned from the insect's first person perspective, we devised a remote controlled robotic camera platform that was directed and programmed to interact with the actors. This is demonstrated in Figure 6. The recording system (shown on the left) was connected wirelessly (via WiFi) to a hand help computer (*ASUS R2H Ultra Mobile PC*) that in turn controlled the motors and sensors of an *IRobot Create* platform the camera head was mounted on (right). Using the UMPC's touch screen driving the robot became a simple task. We also enabled the camera system with macro capabilities, frequently used to “teach” the robot a series of actions to carry out without human intervention. As an additional convenience we also included a Wii controller, that was connected to the UMPC and the robot via a Bluetooth link, the accelerometers and programmable buttons of which offered further convenience. As an example, by simply tilting the Wii in one or another direction, the robot and thus the camera would start to move. The key features of this programmable camera interface are shown in Figure 7.

Figure 6: Remote controlled robot camera used to record the panoramic videos for the Metamorphosis (see text for details)

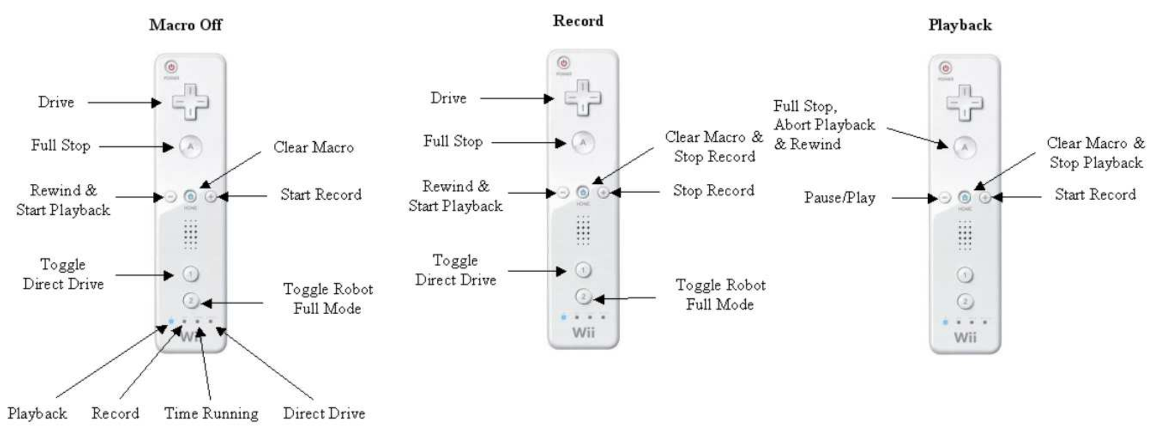

Figure 7: Using a Nintendo Wii controller and Bluetooth connection to control the motion and create programmed action sequences of the robotic camera.

Perhaps one of the most novel aspects of our 360 Immersive Interactive Kafka film is the several layers through which it may be experienced. These are briefly described as follows:

- **Theatrical version:** From the original spherical recordings, with the help of a second production pass using our virtual camera, we have created a final rendered sequence that is projected in movie theatres.
- **Virtual Camera Tracks:** These virtual camera tracks are retained/matched for use in the *Immersive DVD* and *On-line versions*. In particular, viewers may choose to follow the original camera track (see Figure 8) and lay back to simply experience what the director wanted them to see, or alternatively, they can "enter Gregor's world" and look around by turning their virtual camera to any direction using the mouse and keyboard or via the several supported interactive devices, such as *virtual reality glasses*, a *Wii controller* or even their own *IPhone/IPod*. Interaction patterns of different viewers are recorded and used to generate a "viewers choice" camera track that eventually may differ from what the director wanted the audience to see. Through out the entire length of the move each actor and several objects were tracked. In every frame of the film the actors currently present or visible may be selected by the viewer (see Figure 8) and followed as they move during the scene. If they disappear, the camera turns back to the original track, but when they reappear again, it will track them again.
- **Clickable Content:** The tracked actors may also be clicked on to learn more about the character itself, the dramaturgical context or how the film was produced. By moving the mouse over an actor a short label appears. This label may change over time, thus the system is not only capable of expressing mere information, but to follow the evolution of the character, their moods and reactions, or their current state-of-mind. When an actor is clicked, an information window with a short description, photos, videos, etc. appear. This feature allows users to learn more about the story, e.g. several layers of the original book may be accessed, or sometimes simply just to get a "behind-the-scenes" glimpse of how the movie was created.
- **Community / Chat Interface / Share:** On-line users may also enter *chat rooms* related to the film or different scenes of it further fostering interactive experiences, education and learning. With the help of the *DeviceManager* and *VHI Virtual Camera Driver* (need to be separately downloaded) they can also share their own virtual views with others via Skype, Yahoo Messenger, MS or any application that can use a webcamera.
- **Large format interactive version:** Due to the spherical nature of the original footage it is also possible to show the movie in a *large format artistic installation* as demonstrated in Figures 9 & 10. Here the audience is surrounded by a U-shaped projector screen with 4 independent virtual camera views to provide full immersion and enhance the overall experience using peripheral vision. Again, here the original camera tracks as intended by the director are used to control the motion of the virtual camera, but a set of buttons placed on the floor (see Figure 9) allow visitors to jointly rotate it and explore various angles. In the tiny blind-spot of the spherical camera a

Barnabas Takacs/ Digital Elite Inc

web-camera images showing the audience appears conveying the message that Gregor is none else, but ourselves.

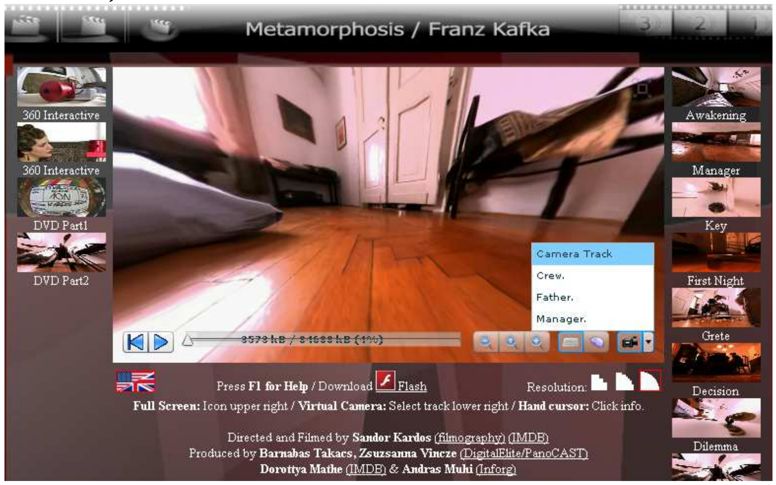

Figure 8: Selecting virtual camera tracks to follow characters using the on-line interface.

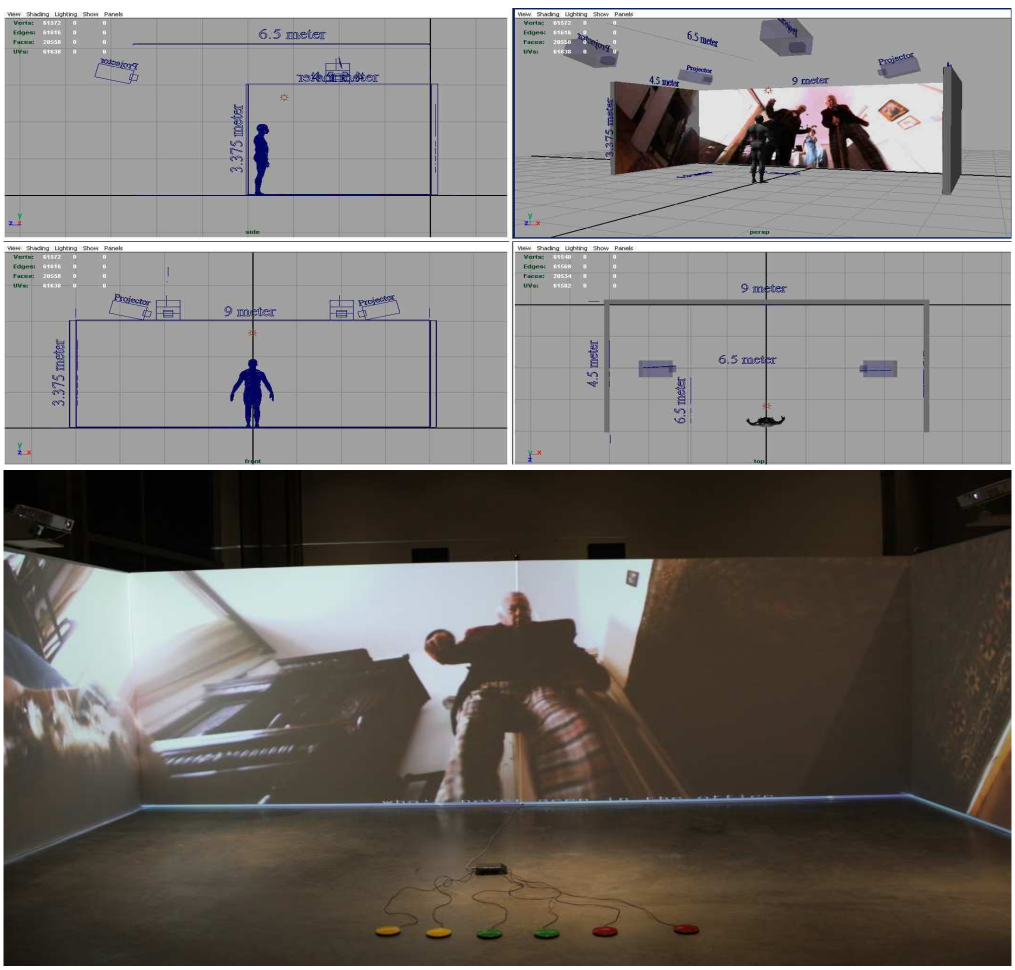

Figure 9: Projection design and actual implementation of a large format U-shaped projection screen to showcase Kafka's immersive interactive film as a museum art installation.

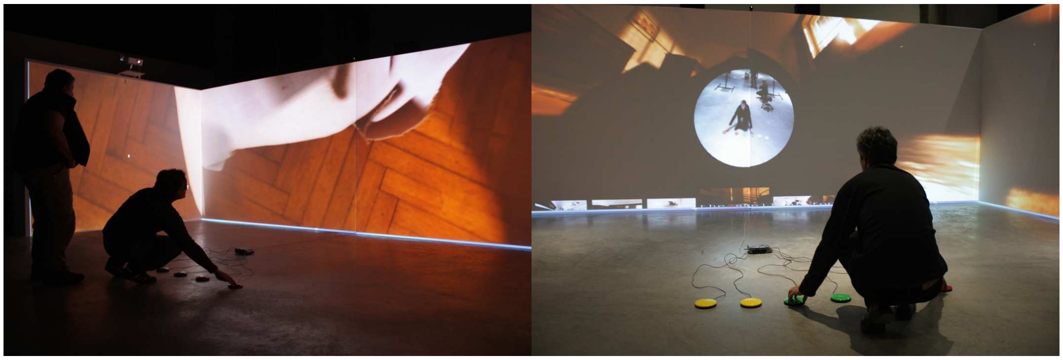

Figure 10: Interactive controls (six buttons on the floor) allow visitors to rotate the virtual camera and change its field of view as a shared museum experience.

6. Conclusion

In this whitepaper we described a novel content production and distribution architecture for Immersive Interactive Film, called *PanoCAST*, that is capable of simultaneously producing a traditional film sequence and turn that same footage into an interactive experience for DVD or on-line viewing. The system employs panoramic video as a starting point and create clickable content with the help of feature tracking to turn every element of a panoramic film scene into a link for further information. We discussed the details of this architecture and demonstrated how it is used in the context of a film project of 40 minutes, entitled Immersive Kafka: The Metamorphosis. Based on our results and the production experience we argue that immersive interactive film as it relates to virtual reality represents a viable path to produce engaging and immersive media content and as such it may serve as the foundation of new generation media.

References

[DigitalElite10] Digital Elite Inc. (2010), www.digitalElite.US.com
[DigitalCinema] Charles S. Swartz (2004), Understanding Digital Cinema: A Professional Handbook, Elsevier.
[Metamorphosis10] Immersive Interactive Film of Franz Kafka's Metamorphosis, http://www.digitalelite.us.com/Metamorphosis2010_Fullversion.html
[PointGrey10] Point Grey Research, LadyBug2 Camera (2010), http://www.ptgrey.com/products/ladybug2/
[PanoCast10] PanoCAST Inc. (2010), www.PanoCAST.com
[Takacs99] Takács, B., T. Fromherz, S. Tice and D. Metaxas (1999), Digital Clones and Virtual Celebrities: Facial Tracking, Gesture Recognition and Animation for the Movie Industry, *ICCV'99*, Corfu, Greece.
[Takacs05] Takacs B. (2005), "Special Education and Rehabilitation: Teaching and Healing with Interactive Graphics", *IEEE Computer Graphics and Applications*, **25**(5): pp.40-48.
[Takacs06] Takács B. (2006), Cognitive, Mental and Physical Rehabilitation Using a Configurable Virtual Reality System, Intl. Journal of Virtual Reality, **5**(4): pp. 1-12.
[Takacs07] Takács, B. (2007) "PanoMOBI: A Panoramic Mobile Entertainment System", *ICEC07, Sept 15-17, Shanghai, China.*
[Takacs08] Takacs B. (2008) "How and Why Affordable Virtual Reality Shapes the Future of Education"**,** *The International Journal of Virtual Reality, 2008, 7(1):53-66.*
[VirtualMuseum09] Virtual Museum in Volos, Greece (2009), http://www.digitalelite.us.com/Pages/PanoCAST/VirtualMuseum.html
[VirtualMuseum10] Takacs, B. (2010), Immersive Interactive Reality: Internet-based On-Demand VR for Cultural Presentation, Springer Virtual Reality, accepted for publication.
[Wikipedia10] Wikipedia (2010) Franz Kaffka: The Metamorphosis,

http://en.wikipedia.org/wiki/The_Metamorphosis