# PanoMOBI: Panoramic Mobile Entertainment System

Barnabas Takacs[1,2]

[1] MTA SZTAKI, Virtual Human Interface Group, Hungarian Academy of Sciences,
Kende u. 11-13, 1111 Budapest, Hungary.
Btakacs@sztaki.hu
[2] Digital Elite Inc.
415 Washington Blvd, 90292 Marina del Rey, CA, USA.

**Abstract.** This paper presents a panoramic broadcasting system for mobile entertainment using 3G network or WIFI where multiple viewers share an experience but each having full control of what they see independent from other viewers. Our solution involves a compact real-time spherical video recording setup that compresses and transmits data from six digital video cameras to a central host computer, which in turn distributes the recorded information among multiple render- and streaming servers for personalized viewing over 3G mobile networks or the Internet. In addition, using advanced computer vision, tracking and animation features, our architecture introduces the notion of Clickable Content (CC) where each visual element in the image becomes a source for providing further information, educational content or advertising. Therefore the PanoMOBI system offers a low-cost and economical solution for personalized content management and it can serve as a unified basis for novel applications.

**Keywords:** PanoMOBI, PanoCAST, Clickable Content, Telepresence, Immersive Spherical Video, Mobile Broadcast Architecture (MBA)

## 1 Introduction

The dream of achieving panoramic broadcasting and subsequently delivering immersive content for tele-operators of robotic- or surgical equipment as well as for security and defense purposes, has been in the focus of research for many decades. In addition, the notion of telepresence as implemented with the tools of virtual-reality for entertainment and education (edutainment) have also long intrigued scientists and developers of complex systems alike. Our current research focuses on presenting visual and auditory stimuli to multiple viewers at the same time and allowing them to share their experience. Video-based telepresence solutions that employ panoramic recording systems have recently become an important field of research mostly

deployed in security and surveillance applications. Such architectures frequently employ expensive multiple-head camera hardware and record data to a set of digital tape recorders from which surround images are stitched together in a tedious process. These cameras are also somewhat large and difficult to use and do not provide full spherical video (only cylindrical), a feature required by many new applications. More recently new advances in CCD resolution and compression technology have created the opportunity to design and build cameras that can capture and transmit almost complete spherical video images [1], but these solutions are rather expensive and can stream images only to a *single viewer*. For the purposes of entertainment, however, many of these systems are too complex for the every-day user and also costly for operators and content providers to deploy. Therefore in our current research we focused on presenting visual and auditory stimuli to multiple users who share the same experience using their mobile-phone or IP-based digital delivery mechanisms. To address this technical challenge we have designed and architecture that can redistribute spherical video to multiple independent viewer each having control over their own respective point of view with the help of an advanced virtual reality environment, called the *Virtual Human Interface (VHI)* [2][3].

## 2 PanoMOBI System Architecture

PanoMOBI stands for *Panoramic Mobile Broadcasting*. The technology is the extension of our earlier solution for telepresence and Internet-based services, called *PanoCAST* (Panoramic Broadcasting). To record and stream high fidelity spherical video we employ a special camera system with six lenses packed into a tiny head-unit. 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 server architecture, then employs a number of virtual cameras and assigns them to each user who connects from a mobile phone, thereby creating their own, personal view of the events the camera is seeing 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 process 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 remaps the spherical imagery onto the surface of a sphere locally. Finally, the personalized rendering engine of the viewer creates TV-like imagery and sends it to a *mobile device* with the help of a *virtual camera* the motion of which is directly controlled by the actions of the user.

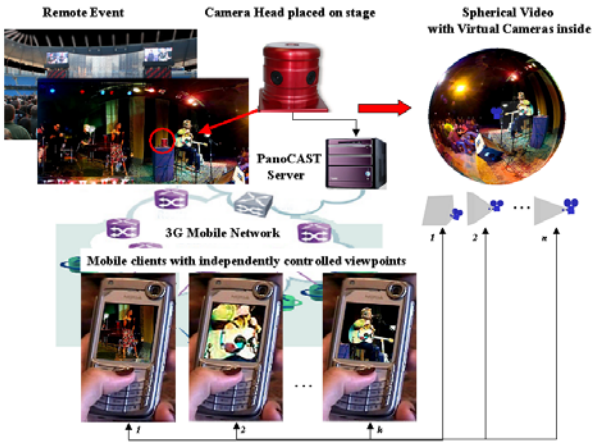


**Fig. 1.** Overview of the PanoMOBI system used for interactive mobile entertainment.

The key idea behind our *PanoMOBI* solution is based on distributing each user only what they currently should see instead of the entire scene they may be experiencing. While this reduces the computational needs on the receiver side (essentially needing only to decode streamed video and audio data) and send tracking information and camera control back in return, it places designers of the server architecture in a difficult position. To overcome these limitations we devised an architecture as follows: The *panoramic camera head* is connected via an optical cable to a *JPG compression* module, which transmits compressed image frames at video rates to a distributions server using IEEE firewire standard. The role of the *distribution server* is to multiple the data video data and prepare it for broadcast via a server farm. To maximize bus capacity and minimize synchronization problems, the distribution server broadcasts its imagery via *UDP protocol* to a number of *virtual camera servers*, each being responsible for a number of individually controlled cameras. Video data is then first encoded in MPEG format and subsequently distributed among a number of *streaming servers* using RTSP (Real-time Streaming Protocol) before sent out to individual clients over 3G mobile networks (or the Internet for WIFI service). Assuming 3Gbit/sec connection a streaming server is capable of servicing up to 100 simultaneous clients at any given moment. Again, the number of streaming servers can be scaled according to the need of the broadcast. Finally, independent image streams are synchronized with audio and arrive at the user site ready to be decoded and displayed. In the *PanoMOBI* entertainment system interaction occurs by key action on the mobile device whereas the user controls the orientation and field of view of the camera to observe the remote concert event taking place.

## 3 Panoramic Recording and Broadcasting

One of the key elements of our mobile broadcasting solution is the *PanoMOBI* recording system that comprises of a compact and portable 360 degree panoramic video recording system that was designed to minimally interfere with the scene while providing maximal immersion for the viewer. Since practically the entire spherical surround environment is recorded working with such a camera is rather difficult from a production's point of view. Specifically, the basic rules and the concept of frames here become obsolete, both lighting, microphones as well as the staff remains visible. To provide as much immersion as possible, the camera head is placed on the end of a long pole carried by the cameraperson. To enhance the functionality of our broadcasting solution we have enabled our server-architecture to create multiple live streams of the panoramic imagery and redistribute them in real-time (or from recorded content) to a large number of mobile viewers, each controlling their own point of view. These multiple receivers on the client side were created in the form of a *PanoMOBI* player and programmed in *Symbian* for maximum speed and performance. In addition, a simpler Java-based client is also available, which is capable of receiving video data for applications where synchronized sound and high speed image transmission are not as critical. In addition to the mobile player different Web-based streaming solutions make this personalized content available to even broader range of audiences. This is shown in Figure 2. In this example the camera was placed on stage to provide an unusual perspective of the band playing. The six raw images from the panoramic recording head are shown in the upper left corner. From these streams a full spherical view of the scene was created and a number of virtual cameras image the scene each controlled independently by a different user. In the upper right hand corner with six independent views stacked up serving as examples. The streaming server then encodes and distributes these streams to individual viewers in different formats, such as RTSP on 3G mobiles (shown lower left) or web-based clients (lower right). In the former case, the rotation and field of view of the camera may be controlled from the keypad of the mobile phone, while in the latter case HTML-embedded keys and Java script commands serve the same purpose. Finally, in both situations the user may click on the screen the result of which the *PanoMOBI* system computes which object was selected and sends back information and web-links for later use. This provides added value for services based on this technology not only for entertainment, but e-commerce, education and many other fields of application. In the next section we briefly discuss how the functionality of *Clickable Content* was achieved using resources of the GPU and our advanced real-time image processing pipeline.

## 4 Real-Time Image Processing and Clickable Content

To enhance the functionality and the utilization of the basic panoramic viewing experience we have incorporated a set of real-time image processing algorithms that help with compression, correct artifacts, produce effects and finally, but most importantly allow for tracking different elements as they move in the scene or

measure displacements as a result of camera motion (Handy-Cam mode). The high resolution raw image that contains six independent images) enters the stream and first passed through a real-time image processing module. This module was built on top of *OpenCV*, a computer vision library that offers basic image processing and enhancement algorithms as well as advanced tracking solutions and 3D reconstruction methods. These processes run on the CPU and are used for non-linear noise filtering (not only to improve overall image quality, but also to allow the better compression of the image stream). Due to the very high resolution of the raw images, the average performance rate (for decoding the video frames and processing them) drops render performance to approx. 15 to 24 fps. As a function of lighting conditions and scene complexity, which is still sufficient to stream images live to the mobile devices at a rate of 15 fps. These CPU-based algorithms are also useful for extracting the information needed for *Clickable Content*, as automated face detection algorithms and other object recognition modules are used to identify and track the apparent motion of elements of interest. The processed images are mapped onto surfaces (a sphere in our case) using a dynamic texture mapping algorithm that was optimized for minimizing the bottleneck between the CPU and the Graphics Card (GPU) present on the computer bus that connects them. Finally, once the textures are uploaded onto the graphics card, the parallel image processing algorithms may be constructed to further enhance the image itself in the form of pixel shaders. We use a single pass method whereas the shader code and parameters correct color, find edges and/or filter out noise using kernels. The main advantage of processing images at this stage comes from the distributed nature of the architecture. According to our experiments the GPU-based image enhancement algorithms we implemented caused practically no slow down in the overall rendering performance.

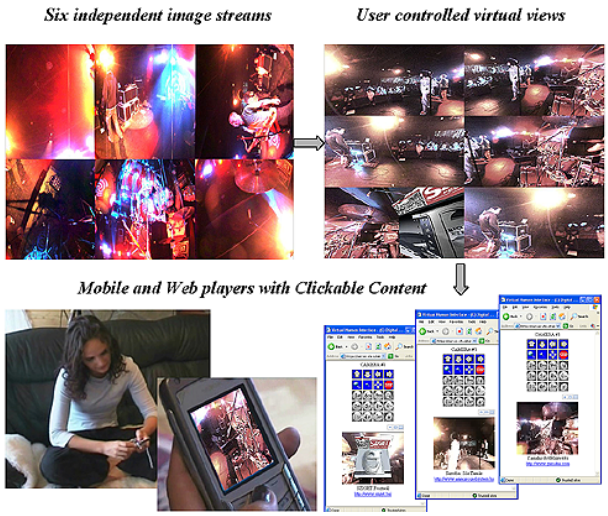


**Fig. 2.** Panoramic imagery is passed from the camera head to individual users who view and interactively control their viewpoint as well as the content using mobile- and web clients.

*Clickable Content* (*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. The final output of such algorithms is a set of tracked regions with actions, text information and web pages assigned on each of which *PanoMOBI* viewers can click on during a performance and learn more about the event or simply educate themselves.

## 5 Conclusion and Future Work

In this paper we have introduced a multi-casting application for mobile phones where each viewer is allowed to individually control their own perspective while sharing this experience with others. We used a high resolution panoramic camera head placed at an entertainment event to create a real-time spherical video. We devised a rendering architecture to map virtual views of this scenery with multiple cameras and augmented the basic capabilities of our solution with real-time image processing pipeline running both on the CPU and the graphics card. This allows for balanced performance and personalized effects viewers may want. Finally, with the help of built-in tracking algorithms a content management solution, called *Clickable Content* (CC) was developed in order to turn an every day video into a rich source of information and commercial tool. Using this *PanoMOBI* architecture we developed intuitive user controls and multiple applications that involve this special form of telepresence. Specifically, we recorded music concerts, real-estates, scenery and travel-sites to demonstrate the usability of our system in real-life applications. The broadcasting system been tested in a number of digital networks including 3G mobile with multiple carriers for phone devices and PDA's, WIFI connection and even wired-Internet for desktop solutions. Test results showed that a single server computer can deliver services to up to 50 clients with reasonable 1-2 seconds delay. We argue that such a technical solution represents a novel opportunity for creating compelling content.

**Acknowledgments.** The research described in this paper was partly supported by grants from MTA SZTAKI, *VirMED Corporation*Budapest, Hungary (www.VirMED.net), and *Digital Elite Inc.*, Los Angeles, California, USA (www.digitalElite.net).

## References

1. Immersive Media http://www.immersive-video.eu/en
2. Takacs B. Special Education and Rehabilitation: Teaching and Healing with Interactive Graphics, IEEE Computer Graphics and Applications, **25**(5): pp.40-48 (2005).
3. Takacs B. Cognitive, Mental and Physical Rehabilitation Using a Configurable Virtual Reality System", International Journal of Virtual Reality, **5**(4): pp.1-12 (2006).