Recording Real Worlds for Playback in a Virtual Exercise Environment

Wei Xu

John Penners

Jane Mulligan

University of Colorado at Boulder

Technical Report CU-CS 1013-06
Department of Computer Science
Campus Box 430
University of Colorado
Boulder, Colorado 80309

Recording Real Worlds for Playback in a Virtual Exercise Environment

Wei Xu, John Penners, Jane Mulligan
Dept of Computer Science,
University of Colorado at Boulder,
Boulder, CO 80309-0430

Abstract

Adhering to an exercise program is a challenge. Augmenting stationary exercise equipment with an immersive Virtual Exercise Environment may make adherence more fun. We describe a system for capturing real bicycle trails for playback in such a VEE. In particular we discuss instrumentation of an adult trike to record distance and tilt angle. Simultaneously we capture image sequences from a cylindrical 5 camera cluster which are merged offline into extended panoramic video sequences. These combined terrain and surround video recording techniques allow us to capture both physical aspects of the terrain and immersive visual information. This information can then be used to "play back" the trail both physically on the exercise equipment and visually on a tracked Head Mounted Display (HMD).

1 Introduction

Sticking to exercise programs is a challenge for both the able bodied and those with disabilities. Our goal is to study the effect of augmenting standard exercise equipment, such as treadmills, stationary bikes or arm ergometers, on adherence to regular exercise programs. Our augmentation takes the form of a Virtual Exercise Environment (VEE) evaluated in phases by able and disabled individuals.

The VEE consists of a target exercise machine, audio and video displays and a workstation to drive the displayed percepts. What distinguishes our system from others that integrate displayed games and graphics for exercise equipment (`http://www.expressofitness.com/`, `http://www.cybextrazer.com`, `http://www.beyondmoseying.com/cateye-game-bike-fitness.html`), is that we are implementing a capture phase which records real trails for playback. We have also focused on low cost solutions because our target community includes persons with disabilities.

The main challenges for a complete recording and playback system for the VEE include:

Physical terrain recording developing instrumentation to capture physical trail features.

Panoramic video recording building a low cost multicamera system which can be driven by a laptop computer at sufficiently high framerates. Undistortion and blending of multi-image frames into coherent panoramas.

Terrain playback modulating recorded sensor data for appropriate playback as resistance or incline on exercise machines. Developing control interface to drive playback.

Immersive video playback display of panoramic sequences in a head mounted display using tracked head position to select the correct view of panorama cylinder, and user speed from the exercise equipment to drive frame rate.

The capture system includes a panoramic head consisting of 5 Unibrain firewire board cameras mounted on a cylinder (Fig. 1(a)), plus special purpose boards and sensors (Fig. 4(b)) integrated with an adult trike which measure tilt and odometry (distance traveled). All measurements are recorded simultaneously by a laptop as the trike is ridden along scenic bike trails. Thus multi-image sequences are labeled with distance and tilt for each temporal frame.

(a) (b)

Figure 1: (a) Low cost cylindrical capture head. (b) Camera head mounted on trike capture vehicle.

Figure 2: Recorded surround video and terrain features are played back on synchronized HMD and exercise bike.

The display system uses the eMagin z800 3D head mounted display which integrates audio earbuds and a 3 DoF head tracker. The exercise equipment itself acts as a physical display, by "playing back" an approximation of the recorded terrain using the speed and incline or resistance settings for the machines. The LifeFitness equipment we are using complies with the Communications Specification for Fitness Equipment (CSAFE) protocol (`http://www.fitlinxx.com/CSAFE/`) which allows us to read and control machine settings. Cylindrical panoramic images computed from the captured multi-image sequences are displayed in the HMD according to head position computed by the head tracker. In this way the user can "look around" the trail surroundings as he moves through the virtual trail world.

2 Related Work

The Virtual Environments group at the University of Utah has produced a number of interesting results on *locomotion interfaces* [18, 9]. These are interfaces which cause the user to expend energy as they simulate unconstrained activities such as walking or running for VR in limited space. The particular system they use combines a special purpose treadmill with immersive visual displays to study perception action couplings. Their locomotion display includes a large Sarcos treadmill with an active mechanical tether. The tether applies inertial forces and emulates slope changes. The standard exercise equipment which forms the locomotion display for the VEE cannot match the devices used in these studies, but concepts such as the importance of matching the visual percept to walking speed [18], are highly relevant

The image processing challenge for the VEE system is to capture and blend extended (45 min. to 1

(a) (b) (c) (d) (e)

Figure 3: Raw images for a temporal cylindrical frame.

hour) trail sequences robustly. Some immersive video systems use mirror-based omnicams to capture a 360° field of view (`http://www.ipix.com/products_video.html`). This approach generally suffers from limited resolution (single CCD) and may be more sensitive to the presence of the sun in outdoor scenes. For panoramas stitched together from multiple camera views many programs work on a single set of images, and require user tuning to achieve the desired quality for the final mosaic (`http://www.panavue.com/index.htm`, `http://www.panavue.com/index.htm`). For the volume of panoramic frames required for the VEE we need an automated system that generates the cylindrical panorama frames with minimal intervention. Many systems also depend on the individual images being relatively high resolution and distortion free. Some work on cylindrical panoramas uses a single camera mounted on a fixed tripod, where the camera is rotated about the vertical axis to capture many overlapping frames [11, 14]. This has the advantage of simplifying the transformation between views to a rotation and allowing as many views as necessary to create a high resolution cylinder. It does not lend itself to real time video capture however.

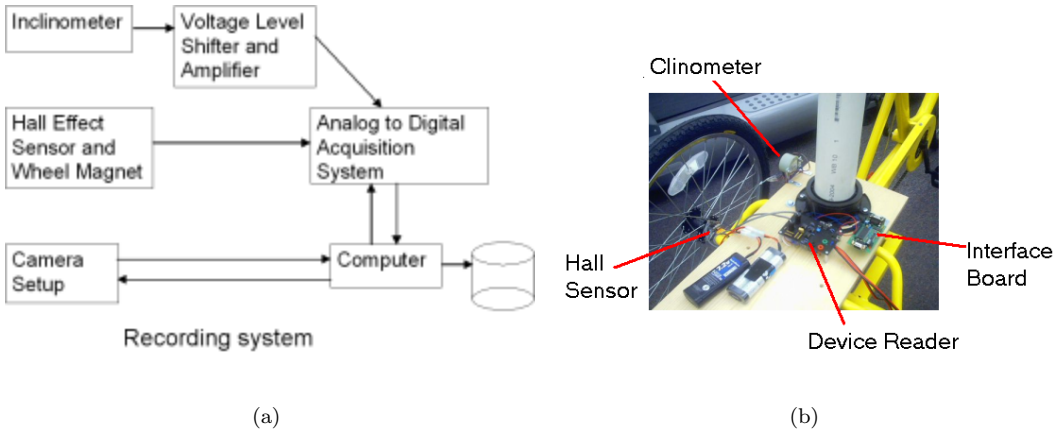
A number of multicamera systems have been proposed for capturing surround video. Point Grey Research packages a six camera proprietary spherical camera head (`http://www.ptgrey.com/products/ladybug2/`). Foote and Kimber [7] and Nanda and Cutler [10] describe 8 and 5 camera systems respectively, applied to recording office meetings. In order to achieve the overlap required to produce cylindrical video frames our system captures from 5 cameras simultaneously, and we use extremely wide angle lenses (1.7 mm focal length). The result is that the camera centres are typically not perfectly aligned on a circle in a single plane and the images have significant radial distortion (Fig. 3).

Another particular challenge of our task domain is that scene distance will range from ground and trail features underfoot and nearby trees and rocks, to distant mountains. Most panorama systems focus on relatively distant scenes which allow the *flat scene* assumption because the depth variation in the scene is small relative to the camera distance [8]. Systems for meeting recording typically calibrate for the approximate distance to participants, and objects much nearer or farther away will have disparity effects, causing a double image in the merged panorama [7].

3 Trail Recording

In order to provide a virtual sensation of strolling down a path, the video image needs to play back at a speed that is consistent with the pace of the person walking down the path [18]. At the same time the amount of effort that is expended needs to be consistent with the grade (incline) of the path. We have developed a measurement system mounted on an adult trike (Fig. 4(b)) which measures tilt using an clinometer and distance using a Hall effect sensor. This is accomplished with a multi-channel analog RS-232 data acquisition system recording electronic measurements while images are being captured so that both the distance traveled and slope can be computed after the recording session is finished. A laptop computer continuously cycles through acquiring images from the camera rig and retrieving data from the sensor acquisition program and writing this data to a file for post processing.

Distance measurements are based on a Hall effect transistor detecting the presence or absence of three strong magnets spaced evenly around one of the rear wheels as the wheel rotates. As the magnets come into proximity to the sensor the it turns on, and as the magnets move away it turns off. This is similar to



(a) (b)

Figure 4: Recording process (a) and recording hardware mounted on capture trike (b).

the way a bike odometer works except that a bike odometer relies on the motion of the magnetic field in order for a signal to be detected. We run our capture system at relatively slow speeds to ensure that the camera rig acquires dense sequences in case a user walks at very slow speeds during playback. At these speeds a bike odometer would not reliably detect the presence of the magnets. With post processing the analog levels collected are converted into on and off levels or *ticks* and the distance traveled is computed and aligned with the images. With three sensors on the 24inch wheel of the trike, each detected sensor level change corresponds to a distance traveled of approximately 2 feet.

The incline measurement utilizes a Schaevitz Accustar Clinometer `http://www.meas-spec.com/myMeas/sensors/schaevitz.asp` attached to the base of tricycle. This is an electronic device that puts out analog voltages based on tilt. When tilted in one direction the voltage goes up and when tilted in opposite direction the voltage goes down. It has a linear range of +/- 45 degrees, which is overkill for this application. As a reference point paved parking lots have a slope of 1-2 degrees for drainage purposes and a steep slope warning is common on freeways when the slope is greater than 7 degrees. The output voltage of the clinometer is centered at zero but our analog data acquisition system only measures positive voltages so an inverting amplifier is used to convert negative voltages into a positive voltage. An additional amplifier is used on the positive signal so that the measurement range of the our data acquisition is +/- 10 degrees with 12 bit resolution. The data acquisition system therefore records two voltages (one for positive tilt and one for negative tilt) and post processing is used to determine whether the slope is positive or negative. Negative voltages on data acquisition system appear as zero voltages and the channel with a positive value indicates the sign of tilt.

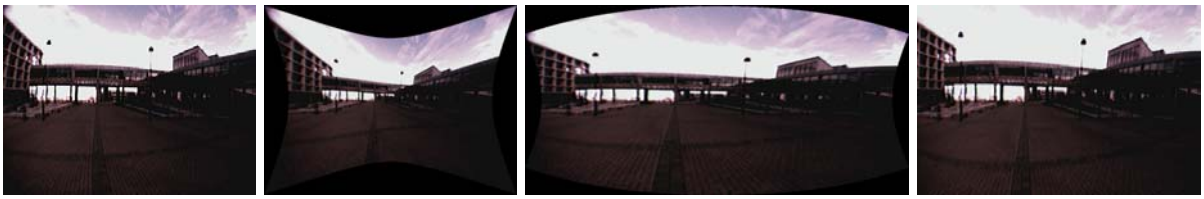
Ideally image and terrain data acquisition would occur simultaneously, where one image set is collected for each incline and distance data measurement. In reality images are acquired more often than tics from the hall sensor, so after data collection distances are interpolated over the number of images between tics to produce a file with a distance and incline value associate with each frame.

4 Panoramic Video

A panoramic video frame or mosaic is built from images captured by our multicamera device according to the following steps: 1) camera calibration and image undistortion; 2) Projection of images to the viewing cylinder; 3) Panoramic image mosaic stitching, which includes image registration, resampling and blending.

4.1 Camera Calibration and Image Undistortion

As we described above, the cameras used for our panorama system are equipped with wide angle lenses and therefore suffer from significant radial distortion. The greatest effect of radial distortion occurs at the



(a) (b) (c) (d)

Figure 5: The undistortion and projection process: a) an original frame from the camera cluster; b) undistorted frame; c) projection onto the cylindrical viewing surface; d) cropped cylindrical frame.

image boundaries, i.e. where the radial distance to the image centre is greatest. Unfortunately these are precisely the regions we wish to merge and blend. Swaminathan and Nayar [16] developed a nonmetric method for calibrating camera clusters composed of wide angle lens cameras. Their method calibrates the images by assuming the distortion of the lens obeys the Brown-Conrady model [4, 5], which includes both radial distortion and tangential distortion:

The radial distortion at a point $q(x, y)$ on the distorted image is approximately modeled as [4]:

$$\Delta r(q) \approx C_3 r^3 + C_5 r^5$$

where $r = \sqrt{(x - x_p)^2 + (y - y_p)^2}$, (x_p, y_p) is the estimated optical center (i.e. the principle point). C_3 and C_5 are the coefficients for radial distortion of 3rd and 5th order.

By defining $\tan\phi = \frac{y-y_p}{x-x_p}$, the tangential distortion component is approximated as [4, 5]:

$$\Delta T_x(q) \approx [P_1 r^2(1 + 2\cos^2(\phi)) + 2P_2 r^2 \sin\phi\cos\phi]$$
$$\Delta T_y(q) \approx [P_2 r^2(1 + 2\sin^2(\phi)) + 2P_1 r^2 \sin\phi\cos\phi]$$

And the total distortion is modeled as:

$$\delta x(q) \approx \delta R_x(q) + \delta T_x(q)\delta y(q) \approx \delta R_y(q) + \delta T_y(q)$$

where $\delta R_x(q) = cos\phi(\delta r(q))$ and $\delta R_y(q) = sin\phi(\delta r(q))$ are the radial distortion terms.

Swaminathan and Nayar's method [16] requires the user to select a set of points from lines that were straight in reality but appear curved in the images due to lens distortion. The optimal lens distortion parameters (including x_p, y_p, C_3, C_5, P_1, P_2) are selected such that undistorted curve points lie on straight lines. They are computed nonlinearly according to least mean squared distance of the undistorted points to the straight lines. They use this method to calibrate their multicamera system composed of four cameras and obtained very good results.

We used Swaminathan and Nayar's method [16] to calibrate our five- camera multicamera system but with slight modification. For many vision applications, it is reported [6, 19, 3] that the tangential distortion component is insignificant relative to radial distortion and can be safely ignored. It is also reported [1] that including both the distortion center and the tangential coefficients in the nonlinear optimization step may lead to instability of the estimation algorithm. Due to these considerations and to save computational cost, we omitted the tangential distortion component in our calibration. The calibration results for the radial distortion component for all the five cameras in our system are very close, all at the level of $C_3 \approx 2.8 \times 10^{-5}$ and $C_5 \approx 2 \times 10^{-11}$.

With the camera distortion parameters known, a backward interpolation technique [21] was used to unwarp the distorted images to their undistorted appearance. Figure 5(a) and 5(b) show an original distorted image captured by one of the five cameras and corresponding undistorted image computed using backward interpolation. It can be seen that curved lines become pretty straight after being unwarped

Figure 6: Computed panorama.

4.2 Projection to the Viewing Cylinder

The calibrated images need to be projected to the viewing cylinder before they are stitched into a panorama since the immersive display system has a conceptual cylindrical viewing surface. Following the projection approach described by Szeliski [17], we map image coordinates $p = (x, y)$ onto 2D cylindrical screen location $u = (\theta, v)$, $\theta \in (-\pi, \pi]$ using

$$\theta = \tan^{-1}(\frac{x}{f}) \text{ and } v = \frac{y}{\sqrt{x^2 + z^2}}$$

where f is the average focal length of the CCD (average of the focal lengths in x and y directions). Figure 5(c) shows the projection of Figure 5(b) onto the viewing cylinder.

4.3 Panoramic Mosaic Stitching

4.3.1 Image Registration

In his book Goshtasby [8] describes a sequence of methods for computing the transformation function between two adjacent views, including both linear and non-linear transformation methods. Generally speaking, different methods should be used for different cases depending on whether the *flat scene* assumption holds and whether the the camera has a wide field of view. Currently we have implemented linear transformation (including affine transformation and projective transformation) based methods, which work for flat scenes that are sufficiently far away (see Figures 3 and 6 for an example). Non-linear transformation based methods are currently under development for close and non-flat scenes.

For building a panoramic image mosaic, Szeliski [17] proved that panoramic images taken from the same view point with a stationary optical center are related by two-dimensional projective transformation. The geometry of our multicamera device approximately satisfies this condition when the scene is far away and we can thus approximate the geometric relationship between any two adjacent camera views by two-dimensional projective transformation. Suppose (x_i^k, y_i^k) and (x_i^{k+1}, y_i^{k+1}) are images of a world point p_i in two adjacent camera views I_k and I_{k+1}. The projective transformation from (x_i^k, y_i^k) to (x_i^{k+1}, y_i^{k+1}) is:

$$\begin{aligned} x_i^{k+1} &= \frac{m_0 x_i^k + m_1 y_i^k + m_2}{m_6 x_i^k + m_7 y_i^k + 1} \\ y_i^{k+1} &= \frac{m_3 x_i^k + m_4 y_i^k + m_5}{m_6 x_i^k + m_7 y_i^k + 1} \end{aligned}$$

where

$$M = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & 1 \end{bmatrix}$$

is the two-dimensional projective transformation matrix. The transformation vector error for p_i is:

$$e_i = \min \Big(\sqrt{\left(x_i^{k+1} - \frac{m_0 x_i^k + m_1 y_i^k + m_2}{m_6 x_i^k + m_7 y_i^k + 1} \right)^2 + \left(y_i^{k+1} - \frac{m_3 x_i^k + m_4 y_i^k + m_5}{m_6 x_i^k + m_7 y_i^k + 1} \right)^2} \Big)$$

The optimal value of m is estimated by minimizing the total transformation error $E = \sum_{i=1}^{n} e_i$ for a set of n ($n \geq 8$) correspondence points which can be easily identified (either by observation or making use of

some feature detection results) in the overlapping region of adjacent camera views. The MATLAB function *LSQNONLIN* is used for this nonlinear optimization process.

4.3.2 Image Blending

Typically colour and brightness vary between cameras even those of the same model. So after the relative geometry of individual views is registered we need to blend them over their overlapping areas to get a smooth transition between the views. In our system, adjacent views have overlapping areas determined by their position and the field of view of the lenses we use. We used the blending algorithm described by Goshtasby [8]: At each overlapping point p, the contribution of each adjacent view is inversely weighted by the distance from the local position of p in that view to the view (image) boundary. That is:

$$I_p = \frac{I_p^k d_p^k + I_p^{k+1} d_p^{k+1}}{d_p^k + d_p^{k+1}}$$

where I^k and I^{k+1} are adjacent overlapping views; I_p^k and I_p^{k+1} are the intensity/colour at the local positions of p in adjacent views respectively; d_p^k and d_p^{k+1} are the distances from p to the boundaries of I^k and I^{k+1}; and I_p is the output intensity/colour of point p in the constructed panorama.

Note that to simplify the computation of the distances to the view boundaries, we crop the cylindrically projected images before registering and blending them. For example, Figure 5(d) is the central rectangular area of Figure 5(c) and it is the actual input to the image registration and blending procedures.

4.3.3 Image Resampling

As mentioned earlier, backward interpolation is used to generate an undistorted image from a distorted image. Actually, not only the image undistortion stage, but all the other image manipulations in our system, including projecting images to the viewing cylinder and stitching the panoramic image mosaic, use backward interpolation as the resampling technique for generating the resulting image. The value of backward interpolation is that it provides a valid source pixel address for each location in the interpolated image and thus eliminates visible seams.

4.4 The Minimum Working Distance

When the objects in the scene are too close and inside the *working distance* of the rig, even a non-linear 2D transformation will fail produce a coherent map to the viewing cylinder. Parallax will cause nearby objects to have a double appearance in regions blended from overlapping views. Swaminathan and Nayar derived constraints to estimate the minimum working distance for a camera cluster [16]. Based on their approach we estimated the minimum working distance for our camera cluster to be about 6.5 meters.

4.5 The look-up table

To produce panoramic video streams in a timely manner, the camera calibration parameters, and the image registration and blending parameters of adjacent views are finally turned into a static lookup table that records the direct map between pixels in the panorama and the acquired images. This kind of direct mapping not only avoids step-wise image transformations, intermediate results storing/loading and backward interpolation, but can also maintain the quality of the resulting panorama if the mapping is carefully designed. The lookup table in our system is designed as follows:

For a pixel p at location (x, y) on the panorama, we want to compute its intensity/colour $f(x, y)$. Due to blending, $f(x, y)$ is computed from the intensity/colour at two positions (x_1, y_1) and (x_2, y_2) in two adjacent views and with weights α_1 and α_2 respectively. That is,

$$f(x, y) = \frac{\alpha_1 f(x_1, y_1) + \alpha_2 f(x_2, y_2)}{\alpha_1 + \alpha_2},$$

where α_1 and α_2 represent the distance between (x_1, y_1) and (x_2, y_2) and their view boundaries. Note that (x_1, y_1) and (x_2, y_2) may fall at subpixel locations. Suppose (x_1, y_1) and (x_2, y_2) are respectively the

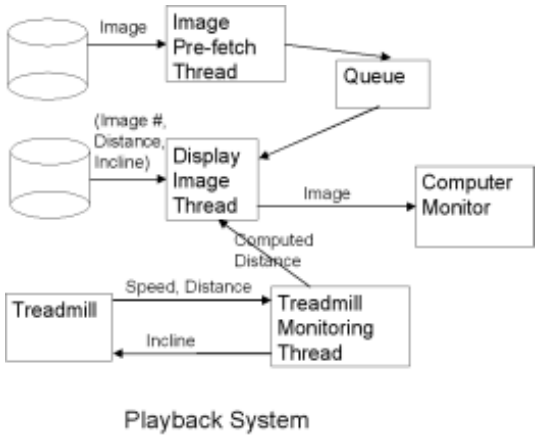


Figure 7: Playback process.

geometric transformation (including undistortion and cylindrical projection) output of two positions (X_1, Y_1) and (X_2, Y_2) in the initial distorted images of two adjacent views. Because geometric transformation does not change pixel intensity/colour, $f(x_1, y_1) = f(X_1, Y_1)$ and $f(x_2, y_2) = f(X_2, Y_2)$. (X_1, Y_1) and (X_2, Y_2) may also fall in positions between pixels, thus the values of their intensity/colour are computed using bilinear interpolation [20]:

$$f(X_1, Y_1) = \det\left(\begin{bmatrix} f(\lfloor X_1 \rfloor, \lfloor Y_1 \rfloor) & f(\lfloor X_1 \rfloor, \lceil Y_1 \rceil) \\ f(\lceil X_1 \rceil, \lfloor Y_1 \rfloor) & f(\lceil X_1 \rceil, \lceil Y_1 \rceil) \end{bmatrix} \begin{bmatrix} \lceil X_1 \rceil - X_1 & X_1 - \lfloor X_1 \rfloor \end{bmatrix} \begin{bmatrix} \lceil Y_1 \rceil - Y \\ Y - \lfloor Y_1 \rfloor \end{bmatrix}\right)$$

$$f(X_2, Y_2) = \det\left(\begin{bmatrix} f(\lfloor X_2 \rfloor, \lfloor Y_2 \rfloor) & f(\lfloor X_2 \rfloor, \lceil Y_2 \rceil) \\ f(\lceil X_2 \rceil, \lfloor Y_2 \rfloor) & f(\lceil X_2 \rceil, \lceil Y_2 \rceil) \end{bmatrix} \begin{bmatrix} \lceil X_2 \rceil - X_2 & X_2 - \lfloor X_2 \rfloor \end{bmatrix} \begin{bmatrix} \lceil Y_2 \rceil - Y \\ Y - \lfloor Y_2 \rfloor \end{bmatrix}\right)$$

where $\lceil \cdot \rceil$ indicates the ceiling of a real value, and $\lfloor \cdot \rfloor$ indicates the floor, and $\det(\cdot)$ indicates the determinant of a matrix. Let

$$A^1 = \begin{bmatrix} a^1_{11} & a^1_{12} \\ a^1_{21} & a^1_{22} \end{bmatrix} = \begin{bmatrix} \lceil X_1 \rceil - X_1 & X_1 - \lfloor X_1 \rfloor \end{bmatrix} \begin{bmatrix} \lceil Y_1 \rceil - Y \\ Y - \lfloor Y_1 \rfloor \end{bmatrix}$$

$$F^1 = \begin{bmatrix} f^1_{11} & f^1_{12} \\ f^1_{21} & f^1_{22} \end{bmatrix} = \begin{bmatrix} f(\lfloor X_1 \rfloor, \lfloor Y_1 \rfloor) & f(\lfloor X_1 \rfloor, \lceil Y_1 \rceil) \\ f(\lceil X_1 \rceil, \lfloor Y_1 \rfloor) & f(\lceil X_1 \rceil, \lceil Y_1 \rceil) \end{bmatrix}$$

$$A^2 = \begin{bmatrix} a^2_{11} & a^2_{12} \\ a^2_{21} & a^2_{22} \end{bmatrix} = \begin{bmatrix} \lceil X_2 \rceil - X_2 & X_2 - \lfloor X_2 \rfloor \end{bmatrix} \begin{bmatrix} \lceil Y_2 \rceil - Y \\ Y - \lfloor Y_2 \rfloor \end{bmatrix}$$

$$F^2 = \begin{bmatrix} f^2_{11} & f^2_{12} \\ f^2_{21} & f^2_{22} \end{bmatrix} = \begin{bmatrix} f(\lfloor X_2 \rfloor, \lfloor Y_2 \rfloor) & f(\lfloor X_2 \rfloor, \lceil Y_2 \rceil) \\ f(\lceil X_2 \rceil, \lfloor Y_2 \rfloor) & f(\lceil X_2 \rceil, \lceil Y_2 \rceil) \end{bmatrix}$$

The lookup table is thus composed of $(x, y), \alpha_1, \alpha_2, A^1, F^1, A^2$ and F^2. The computation of $f(x, y)$ is:

$$f(x, y) = \frac{1}{\alpha_1 + \alpha_2} \begin{bmatrix} \alpha_1 & \alpha_2 \end{bmatrix} \begin{bmatrix} \det(F^1 A^1) \\ \det(F^2 A^2) \end{bmatrix}$$

Figure 6 shows a completed panorama and Figure 3 its source images. The panorama was computed directly using the mapping of the lookup table.

5 Trail Playback

The result of trail acquisition and postprocessing is a dense sequence of panoramic video frames labeled with recorded tilt and distance traveled along the trail. The goal of the VEE is to play back this sequence

Figure 8: Stereoscopic visor with integrated head tracker.

using a tracked head mounted display (Fig. 8) and a locomotion display consisting of a standard piece of stationary exercise equipment such as a treadmill or bike (Fig. 2). The playback on the locomotion display monitors the distance traveled by the person exercising and uses this information to calculate when a set of images should be rendered on the immersive display. The playback effectively runs "on a rail" since currently no navigation interface is provided and only fixed recorded traversal of the trail is available.

The interface with the exercise machine is based on the public Communications Specification for Fitness Equipment (CSAFE) that provides a protocol and a list of commands to talk with exercise machines. The CSAFE protocol is based on a master-slave relationship with a master computer normally instructing the slave machine although under certain rare conditions the exercise machine will issue an unsolicited command. The protocol utilizes a start and stop byte that indicate a full packet with multiple commands possible within a packet. A parity byte is also included for error detection. If an exercise machine detects a packet in error, the packet is silently dropped as if it were never sent, but the machine records the event. If the next packet sent to the machine requests a response, the Status byte, which is included in all responses, will indicate the state of the slave and the status of the previous frame, which would indicate an error occurred. By always requesting a response the master can monitor the success of commands.

The treadmill used in our implementation has a distance resolution that is 0.01miles (52.8 feet) which is not a fine enough resolution. Our implementation uses reported speed to calculate the distance that the exerciser traveled. The playback software continually requests both distance and speed data from the exercise machine. In the present version only speed is used to calculate the distance traveled, but future versions can use the distance measure from the slave to dynamically modify the distance traveled calculation to improve the accuracy if need be.

In order to always have images ready to be displayed they are pre-fetched and stored in a queue in local memory. The computer program is based on three threads running concurrently. One thread pre-fetches images and stores them in a queue, another thread continually requests speed and data from the slave and updates the distance traveled measure, a third thread monitors this distance metric and based on this data retrieves images from the queue to be displayed on the video playback system.

5.1 Immersive Display

Our immersive display device is a single integrated HMD the eMagin z800. It includes a head tracker with 360° in X, Y and Z rotations using micro-electro-mechanical system (MEMS) accelerometers and gyroscopes. It has stereo audio and a microphone which will allow the user to chat with a remote exercise partner or competitor using VoIP in networked exercise environments. The display for each eye is an 800x600 pixel organic light-emitting diode (OLED) micro-display. The visor can also be driven in stereoscopic mode using frame-sequential 3D sequences (alternating right and left eye views) at 60Hz.

There are many known issues with rendering immersive scenes using head tracking, including adequate update speed for the displays and head tracker to match head motion so that the view onscreen does not lag behind the user's viewing direction [15]. Typically these systems will apply smoothing and prediction to the head tracker output to prevent jitter and overshoot in the displayed scene. Smoothing issues and how

they interact with the user's internal image stabilization are particularly important in a treadmill setting where the user may be jogging and moving their head significantly.

Another important factor is what to show the user given observed head motion. Part of this question reflects the difference in the field of view between the cameras used to form the panorama, the HMD and the human visual system. The manufacturers claim the z800 has a 40° diagonal field of view for the display. This translates into about 32° horizontal and 24° vertical field of view. The cylindrical mosaics have 360° horizontal FOV, and about 80° vertical FOV. Various estimates are given for the human eye but the FOV is agreed to be large [12, 13], on the order of 150° horizontal FOV and 120° vertical FOV. Obviously this represents mismatches at every level. HMD's are known to reduce the user's field of view and perception of motion in the periphery [15]. We have not yet sufficient experience with our prototype and the evaluation studies to understand how this will affect users in the VEE. Currently we map FOV angles directly, selecting windows from the panorama for rendering based on the viewing direction with size based on the estimated HMD field of view. Since the sense of presence achieved in a virtual environment is related to large FOV [13, 2] we may have to redesign this mapping in future.

6 Conclusions and Future Work

In this paper we have described a complete system for recording of real outdoor trail environments and playback of these recordings in an immersive Virtual Exercise Environment. The recording phase uses an adult trike instrumented with tilt and odometry sensors which are read by a laptop which simultaneously captures multi-image sequences from a 5 camera cylindrical camera rig. The recordings are postprocessed to create panoramic image sequences labeled with distance and tilt information. The VEE uses stationary exercise equipment as locomotion displays and a tracked Head Mounted Display as the immersive visual display. The recorded sequences are played back in the VEE according to the speed the user generates on the exercise machine.

We have described a set of methods for capturing and merging cylindrical panoramas using a low cost multicamera head. Because we use only 5 cameras mounted on a cylinder to cover 360° field of view, we use very wide angle lenses. The result is significant radial distortion which is most severe at the image boundaries. This makes merging and blending the boundary overlap between adjacent views challenging.

The extreme lighting conditions of the outdoor environments we capture is another challenge for our panoramic video system. We plan to exploit and extend some if the automatic calibration techniques proposed by Nanda and Cutler [10] to automatically adjust exposure and colour settings to improve the uniformity of the captured source images.

We are in the initial stages of evaluating playback of captured trail data on basic locomotion displays. Important issues include the scaling of tilt or resistance information on stationary bikes and treadmills, the scaling and smoothing of head tracking information for acceptable rendering of panoramic data in the HMD, and the relative mapping of the field of view between the rendered window and the HMD.

References

[1] Ahmed, M., and Farag, A. Calibration of camera lens distortion: Differential methods and robust estimation. *IEEE Transactions on Image Processing 14*, 8 (2005), 1215–1230.

[2] Arthur, K. W. *Effects of Field of View on Performance with Head-Mounted Displays.* Department of computer science, University of North Carolina at Chapel Hill, Chapel Hill, 2000.

[3] Brown, D. Close-range camera calibration. *Photogrammetric Engineering 37*, 8 (1971), 855–866.

[4] Brown, D. C. Decentering distortion of lenses. *Photogrammetric Eng. 32*, 3 (May 1966), 444–462.

[5] Conrady, A. Decentering lens systems. *Monthly Notices of the Royal Astronomical Soc. 79* (1919), 384–390.

[6] Devernay, F., and Faugeras, O. Straight lines have to be straight: automatic calibration and removal of distortion from scenes of structured environments. *Machine Vision and Applications 1* (2001), 14–24.

[7] Foote, J., and Kimber, D. Flycam: Practical panoramic video and automatic camera control. In *Proc. IEEE International Conference on Multimedia and Expo* (2000), vol. 3, pp. 1419–1422.

[8] Goshtasby, A. A. *2-D and 3-D Image Registration.* John Wiley & Sons, Hoboken, NJ, 2005.

[9] Hollerbach, J. M. Locomotion interfaces. In *Handbook of Virtual Environments Technology*, K. Stanney, Ed. Lawrence Erlbaum Associates, Inc, 2002, pp. 239–254.

[10] Nanda, H., and Cutler, R. Practical calibrations for a real-time digital omnidirectional camera. In *Technical Sketches, IEEE Conference on Computer Vision and Pattern Recognition CVPR01* (2001).

[11] Ng, K.-T., Chan, S.-C., Shum, H.-Y., and Kang, S. B. On the data compression and transmission aspects of panoramic video. In *Proc. International Conference on Image Procesing* (2001), pp. 105–108.

[12] Owen, G. S. Hypervis - teaching scientific visualization using hypermedia. http://www.siggraph.org/education/materials/HyperVis/hypervis.htm, Oct 1999.

[13] Prothero, J., and Hoffman, H. Widening the field-of-view increases the sense of presence in immersive virtual environments. Human Interface Technology Laboratory HITLab Tech Report R-95-5, University of Washington, Seattle, WA, 1995.

[14] Shum, H.-Y., and Szeliski, R. Systems and experiment paper: Construction of panoramic image mosaics with global and local alignment. *International Journal of Computer Vision 36*, 2 (2000), 101–130.

[15] Stanney, K. M., Mourant, R. R., and Kennedy, R. S. Human factors issues in virtual environments: A review of the literature. *Presence 7*, 4 (Mar 1998), 327–351.

[16] Swaminathan, R., and Nayar, S. Nonmetric calibration of wide-angle lenses and polycameras. *IEEE PAMI 22*, 10 (2000), 1172–1178.

[17] Szeliski, R. Video mosaics for virtual environments. In *Virtual Reality* (March 1996), pp. 22–30.

[18] Thompson, W. B., Creem-Regehr, S. H., Mohler, B. J., and Willemsen, P. Investigations on the interactions between vision and locomotion using a treadmill virtual environment. In *Proc. SPIE/IS&T Human Vision & Electronic Imaging Conference* (Jan. 2005).

[19] Tsai, R. Y. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation RA-3*, 4 (August 1987), 323–344.

[20] Bilinear interpolation. Wikipedia: The Free Encyclopedia, March 2006. `http://en.wikipedia.org/wiki/Bilinear_interpolation`.

[21] Wolberg, G. *Digital Image Warping.* IEEE Computer Society Press, Los Alamitos, CA, 1990.