Augmenting Exercise Systems with Virtual Exercise Environment

Wei Xu, Jaeheon Jeong and Jane Mulligan

Department of Computer Science, University of Colorado at Boulder, Boulder,
Colorado 80309-0430 USA
{Wei.Xu, Jaeheon.Jeong, Jane.Mulligan}@Colorado.edu

Abstract. Adhering to an exercise program is a challenge for everybody who wants to build a healthier body and lifestyle through physical exercise. We have developed an Virtual Exercise Environment (VEE) that augments stationary exercise equipment with virtual reality techniques to make exercising more enjoyable. Our VEE system consists of a recording system to capture video, distance and incline data about real trails, and a playback system which "displays" both video and terrain data in the form of video speed and resistance. Trails are played back according to the speed the user generates on the stationary exercise equipment. The system uses commodity capture and display devices and supports standard interfaces for existing exercise equipment. User studies have shown that users enjoy the ability to guage their progress and performance via their progress through trail playback in the VEE.

1 Introduction

Sticking to exercise programs can be challenging for most of us, especially using stationary exercise machines at home or at the gym . Studies show that an external focus of attention helps improve enjoyment and intensity of workouts [1], but typically the only distraction available is a TV or mp3 player. To address this deficiency we have developed an add-on virtual reality system to augment standard exercise equipment, such as treadmills, stationary bikes or arm ergometers. Our augmentation takes the form of a Virtual Exercise Environment (VEE) that simulates real, vivid, outdoor exercise trails for exercising in an indoor environment.

The VEE system consists of a target exercise machine, immersive 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 (e.g., [2]), is that we have developed a capture phase which records both appearance and terrain of outdoor real trails using a panoramic camera head and special purpose boards and electronic sensors. We have also focused on low cost solutions to make the system affordable to most of people.

2 Related Work

To the best of our knowledge, there is no previous work that addresses all the features of our VEE system. However, there is separate work in various related research communities, including computer vision, image processing, computer graphics and virtual reality community, that gives us hints on how to build our system.

The Virtual Environments group at the University of Utah has produced a number of interesting results on *locomotion interfaces* [3, 4]. These are interfaces which cause the user to expend energy as they simulate unconstrained activities such as walking or running in virtual reality (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 in our VEE system can not match the devices used in these studies, but concepts such as the importance of matching the visual percept to walking speed [3], are highly relevant.

Some navigation-based VR systems such as the early Aspen Movie Map [5] and the more recent Google Street View [6] have also some similarity to our VEE system in the aspect that they also visually record and play back panoramic frames of real outdoor trails. However, these systems do not have a locomotiion display part comparing to our VEE system — there is no monitoring and association of the user's walking speed to the playback of recorded panoramic frames. And the terrain changes of the trails are not recorded and thus there is no way to play them back via a locomotion display in these systems.

A number of multi-camera systems have been proposed for capturing surround video. Point Grey Research packages a six camera proprietary spherical camera head [7]. Foote and Kimber [8] and Nanda and Cutler [9] describe 8 and 5 camera systems respectively, applied to recording office meetings. These multi-camera systems are carefully designed, manufactured and calibrated, but are usually inaccessible or unaffordable to our target community. We instead built our system upon low-cost commodity board cameras that are affordable to most of people. In order to achieve the overlap required to produce cylindrical video frames while using as few cameras as possible, we also used extremely wide angle lenses (1.7mm focal length) that have significant radial distortions (see Figure 4(a)). We will describe how we solve these practical problems in detail in the following sections.

3 The VEE System

Our VEE system is composed of a trail recording phase, a data processing and integration phase, and a trail playback phase. The trail recording sub-system can record both appearance and terrain of a real outdoor trail. The records are then processed and integrated by the data processing and integration sub-system, and

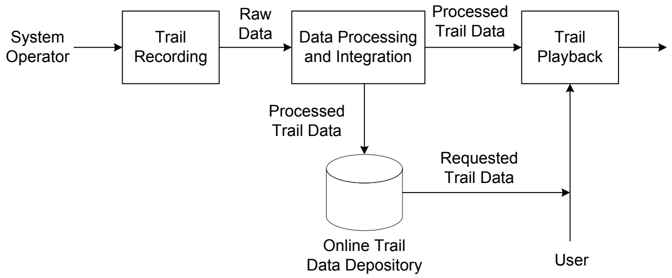

Fig. 1. VEE system overview.

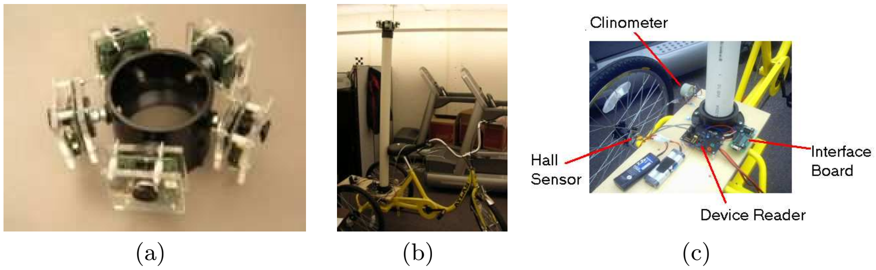

Fig. 2. (a) Low cost cylindrical capture head. (b) Camera head mounted on trail capture vehicle. (c) Terrain recording hardware mounted on capture trike.

the processed data are uploaded into a trail data distribution website from which the public can download the data for free. With downloaded trail data, a user can play the trail back using an immersive display device and standard exercise equipment. Figure 1 gives an overview of the VEE system.

3.1 Trail Recording

The trail recording sub-system includes a panoramic camera head consisting of five Unibrain firewire board cameras mounted evenly on a cylinder (Figure 2(a)), an altered adult trike used to carry the camera head to move around (Figure 2(b)), plus special purpose boards and electronic sensors (Figure 2(c)) integrated with the trike which measure tilt and odometry (distance travelled). All measurements are recorded simultaneously by a control program running on a laptop computer as the trike is ridden along scenic bike trails.

Distance measurements are provided by a Hall effect transistor which detects the presence or absence of three strong magnets spaced evenly around one of the rear wheels of the trike as the wheel rotates. As the magnets come into proximity to the Hall sensor it turns on, and as the magnets move away it turns off, which generates a tick. The distance traveled D_{curr} is computed as

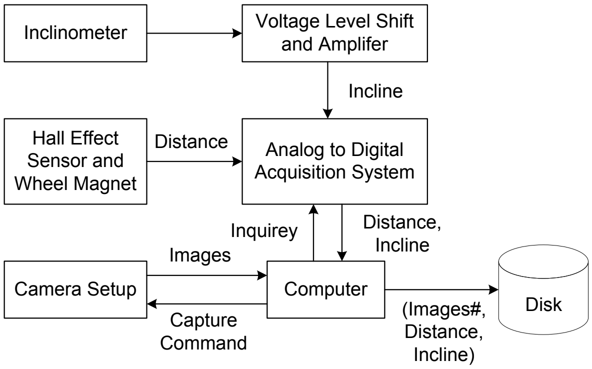

Fig. 3. Trail recording and data processing and integration process.

$D_{curr} = (N_{tick} - 1) * U_{dist}$, where N_{tick} is the count of ticks and U_{dist} is the unit distance traveled between two successive ticks. With three sensors on the 24" wheel of the trike, U_{dist} is approximately two feet.

The incline measurement utilizes a Schaevitz Accustar clinometer [10] attached to the base of of the trike. This is an electronic device that puts out analog voltages based on tilt — when tilted in up-hill direction the voltage is positive and when tilted in down-hill direction the voltage is negative. Since the analog data acquisition system we use to collect the clinometer signals only measures positive voltages, an inverting amplifier is used to convert negative voltages into positive voltage. Another thresholding amplifier is used to collect the positive signals only (by converting all of the negative voltages into zero). By comparing these two serials of positive signals, the original serial of signals that contains both positive and negative values can be restored:

$$T^{(t)} = \begin{cases} (T_1^{(t)} + T_2^{(t)})/2 & \text{if} \quad T_2^{(t)} > 0 \\ -T_1^{(t)} & \text{if} \quad T_2^{(t)} = 0 \end{cases}$$

where $T^{(t)}$ is the original signal of tilt at a time tick t, $T_1^{(t)}$ and $T_2^{(t)}$ are corresponding signals collected vis the inverting amplifier and via the thresholding amplifier respectively.

Procedure of the trail recording process is shown in Figure 3.

3.2 Panoramic Video Stitching

The individual videos captured by the five cameras need to be stitched into a 360° panoramic video for immersive playback in future. This includes several operations considering the special capturing device we use. First, the cameras used in our system are equipped with low-grade wide angle lenses (with $\approx 75°$ horizontal FoV) which suffer from substantial radial distortions. We use the

Brown-Conrady model [11, 12] to model the radial distortions by a low-order polynomial: $x_d = x_u(1 + \kappa_1 r^2 + \kappa_2 r^4)$ and $y_d = y_u(1 + \kappa_1 r^2 + \kappa_2 r^4)$, where $r^2 = x^2 + y^2$ and κ_1 and κ_2 are called the radial distortion parameters [13], (x_d, y_d) is distorted (image) coordinate and (x_u, y_u) is corresponding undistorted coordinate. Here we omit the tangential (decentering) distortions [11], since they are usually negligible for consumer-level stitching [14, 13]. Our calibration using a checkerboard(Figure 4(a)) estimates that $\kappa_1 \approx 1.6 \times 10^{-5}$ and $\kappa_2 \approx 2.0 \times 10^{-11}$. This estimate is applicable for all of the five cameras.

Next, we register the calibrated images and stitch them into a panorama. We assume the five cameras consisting our panoramic camera head approximately follow a rotational geometrical model, i.e. they are co-centered and their poses are only different to each other by a 3D rotation. Under this model, the stitching is simply warping the images into cylindrical coordinates and then using a pure translational model to align them [15]. The formula for converting a rectangular coordinate (x, y) into the corresponding cylindrical coordinate (x_c, y_c) is as follows: $x_c = s\theta = s \arctan \frac{x}{f}$ and $y_c = s\rho = s\frac{y}{\sqrt{x^2+f^2}}$, where s is an arbitrary scaling factor that determines the size of the resulting image. Then, the parameters of the translational model between the i'th pair of adjacent views, $(\Delta x, \Delta y)$, are estimated from robust SIFT feature matches (where RANSAC is used to remove the outlier matches). Also, because what we are stitching is a 360° panorama, we need to make sure the left and right ends of the panorama matches with each other. This is achieved by minimizing the global registration error E_G:

$$E_G = \sum_{i=1}^{5} err_i(\Delta x_i, \Delta y_i) = \sum_{i=1}^{5} \sum_{j=1}^{n_i} (\hat{x}_{ij} - x_{ij} - \Delta x_i)^2 + (\hat{y}_{ij} - y_{ij} - \Delta y_i)^2$$

where $err_i(\cdot)$ is the re-projection error and n_i is the number of feature matches between the i'th pair of adjacent views. $(\hat{x}_{ij}, \hat{y}_{ij})$ and (x_{ij}, y_{ij}) are a pair of matched points. After the views are registered in cylindrical surface, a linear image blending technique [16] is used to remove appearance difference and seams in the overlap areas of adjacent views.

Figure 4 illustrates the image calibration and projection operations through an example. And figure 5 shows an example panorama as the final output of the whole processing. In practice, we use direct mapping based on look-up tables to speed up the processing. We learn the geometric transformation and image blending parameters from the stitching of several carefully selected frames. We then compute the direct mapping from pixels on the original images to pixels on the final composite panorama and save the mapping into a static look-up table which is used to stitch all the remaining frames. Direct mapping avoids both step-wise image warping and backward interpolation [17] and saves the time and space for storing/loading intermediate results. Details about the design and implementation of the look-up table can be found in [18].

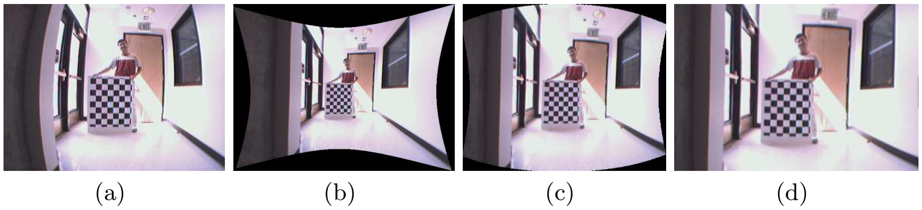
(a) (b) (c) (d)

Fig. 4. Image calibration and projection process. (a) Original image frame captured by one of the five cameras. (b) Frame after removal of radial distortions. (c) Calibrated frame projected onto the cylindrical surface. (d) Cropped cylindrical frame.

Fig. 5. An example panorama.

3.3 Data Integration

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 [3]. At the same time the amount of effort that is expended needs to be consistent with the grade (incline) of the path. This means ideally image and terrain data acquisition would occur simultaneously, where one image set is collected for each distance and incline data measurement. However, 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 one distinct distance and incline mark associated with each frame. The timestamps of the data records were used to guide the association.

4 Trail Playback

The result of trail recording and data processing and integration is a dense sequence of panoramic video frames labeled with recorded incline and distance traveled along the trail. The last phase of the VEE system is to play back this sequence using an immersive video display (either a tracked HMD or a surrounding monitor cluster) and a locomotion display consisting of a standard piece of stationary exercise equipment such as a treadmill or bike. We have developed a multi-threaded control program to coordinate the playback on these two kinds of displays. The program is based on three threads running concurrently. One thread pre-fetches images and stores them in a queue, another thread continually

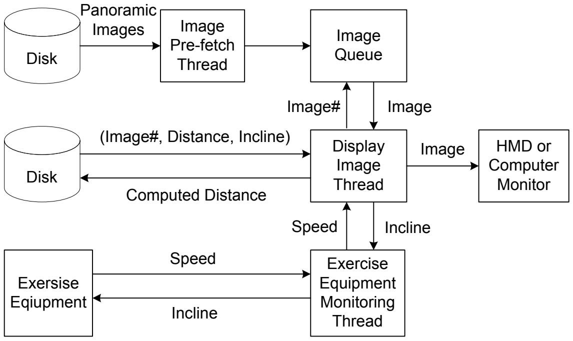

Fig. 6. Trail playback process.

requests speed and data from the exercise equipment and updates the distance traveled measure, and a third thread monitors this distance metric and based on this data retrieves images from the queue to be displayed on the immersive video display. This multi-threaded processing could smoothly play back a trail at a higher speed than single-process progressive processing and it never got any lag problems in its practice with the several testing trails that we have under proper parameter settings (e.g., the image queue size). Figure 6 is a sketch map of the playback process.

4.1 Immersive Video Display

The immersive video display device used in our experiments is a single integrated HMD, the eMagin z800(Figure 7(a)). 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 (Figure 7(b)). To avoid lags due to loading image files into memory, panoramic images are pre-fetched and stored in a queue in local memory for display as needed. Figure 7(c) shows an early non-immersive playback system using a single forward-facing monitor, but including the locomotion display playback on the treadmill.

4.2 Locomotion Display

The locomotion display devices used in our system include various kinds of exercise equipment, such as treadmills, stationary exercise bikes and arm ergometers. The exercise equipment itself acts as a physical display, by "playing back" an approximation of the recorded terrain using the speed and slope or resistance settings of the equipment. The thread that plays back the trail on the exercise equipment calculates the distance traveled by the user that is exercising and uses

(a) (b) (c)

Fig. 7. (a) Head Mounted Display (HMD) (b) Recorded surround video and terrain features are played back on synchronized HMD and exercise bike. (c) Playback using synchronized computer monitor and treadmill.

this information to determine which frame of the panorama sequence should be rendered on the immersive video 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. Some of the exercise equipment does not have a fine enough distance resolution for retrieving continuous image frames for rendering. For example, the distance resolution of the treadmill used by our system (Table 1) is 0.01miles (52.8 feet) which causes jumping artifacts in the rendered video. So, instead our system uses the instantaneous speed reported by the exercise equipment to calculate the accumulated distance and uses this distance to retrieve the appropriate panorama frame for rendering:

$$curState.dist = curState.dist + rd * \mathrm{mph2fps} * curState.speed$$

where $curState$ is a global structural variable that records the run-time status of the machine. $rd = dispTime.runningDuration()$ records the running duration from last computation of $curState.dist$ to now, and $dispTime$ is a global timer whose count resets after each update of $curState.dist$. $\mathrm{mph2fps} = 1.4667$ is a scaling factor that converts the unit from mph (mile per hour) to fps (frame per second). In our current implementation only the speed information reported by the exercise equipment is used to calculate the distance traveled, but in future versions reported distance might also be used to dynamically calibrate the calculation.

The incline information is physically played back on the locomotion display. Different exercise equipment has different usage of this information. In our system, we use the incline information to control the slope of the deck of treadmills, and the resistance level of arm ergometers or stationary exercise bikes. For machines without controllable incline or resistance, we emulate a low gear on a bicycle, where the user must pedal faster (proportional to the measured trail incline) to travel the same distance. Table 1 shows a list of exercise equipment onto which we can successfully display the incline information. We use either the

Name	Category	Manufacturer	Property	Interface
FlexDeck Shock Abso. Sys.	treadmill	Life Fitness	slope	CSAFE
LifeCycle 9500 HR	exercise bike	Life Fitness	resistance	CSAFE
SciFit Pro I	arm ergometer	Sci-Fit	resistance	CSAFE
PCGamerBike	arm ergometer	3D Innovations	resistance	FitXF SDK

Table 1. The exercise equipment experimented in our system, their properties used to play back the incline information, and their communication interfaces.

Communications Specification for Fitness Equipment (CSAFE) protocol [19] or machine associated SDK to control these equipments. The CSAFE protocol is a public standard on exercise equipment manufacturing that is supported by most of today's exercise equipment manufactures. Our playback control program has fully implemented it and thus can communicates with any exercise machines that supports this protocol.

5 Pilot Study and Results

The VEE system we developed was evaluated by researchers at Assistive Technology Partners in Denver, CO, USA. A male participant was recruited for a pilot study of the technology using a treadmill. This individual completed a week long baseline phase in which he logged his daily step count as usual, followed by an intervention phase during which he exercised on our VEE-outfitted treadmill a total of 11 times over a period of four weeks. At the end of the study, the participant was asked to provide feedback on his VEE experience. He described the technology as compelling, stating that he had found it interesting and enjoyable to use. The participant's average daily number of steps during the baseline phase of the study 2065. His average daily step count during the intervention phase was 6151. This included a low of 4433 per day (third week of intervention) to a high of 7926 (fourth week of intervention). Throughout the intervention phase, the participant completed an average of 4563 steps per intervention session. A final pleasant outcome was the patient's body weight, which decreased 5 lbs from the baseline phase to the end of the intervention phase.

6 Concluding Remarks

In this paper we have described a complete system for recording of real outdoor trail environments and playing these recordings back in an immersive Virtual Exercise Environment (VEE). We have developed a low-cost capture device which records both appearance and terrain of outdoor real trails using a panoramic camera head and special purpose boards and electronic sensors. We have also developed a collection of computer programs for controlling the trail recording and playback phases, for rapidly stitching panoramic videos, and for driving and monitoring various kinds of immersive video display and locomotion display devices. Efficacy of the VEE system has been verified by human pilot study.

References

1. Porcari, J.P., Zedaker, J.M., Maldari, M.M.: Real increased exercise performance results when members have more to engage their attention than their own effort and fatigue. Fitness Management Magazine **14** (1998) 50–51
2. Cybex: Cybex Trazer. `http://www.cybextrazer.com` (2007)
3. Thompson, W.B., Creem-Regehr, S.H., Mohler, B.J., Willemsen, P.: Investigations on the interactions between vision and locomotion using a treadmill virtual environment. In: Proc. SPIE/IS&T Human Vision and Electronic Imaging Conference. Volume 5666. (2005) 481–492
4. Hollerbach, J.M.: Locomotion interfaces. (In: Handbook of Virtual Environments Technology)
5. Naimark, M., Clay, P., Mohl, B.: Aspen Movie Map. `http://www.naimark.net/projects/aspen.html` (1978)
6. Google: Google Street View. `http://maps.google.com/help/maps/streetview` (2007)
7. PointGreyResearch: Ladybug. `http://www.ptgrey.com/products/spherical.asp` (2006)
8. Foote, J., Kimber, D.: Flycam: practical panoramic video and automatic camera control. In: Proc. IEEE International Conference on Multimedia and Expo. Volume 3. (2000) 1419–1422
9. Nanda, H., Cutler, R.: Practical calibrations for a real-time digital omnidirectional camera. In: Technical Sketches IEEE International Conference on Computer Vision and Pattern Recognition. (2001)
10. Measurement Specialties, I.: Schaevitz accustar clinometer. `http://www.meas-spec.com/myMeas/sensors/schaevitz.asp` (2007)
11. Brown, D.C.: Decentering distortion of lenses. Photogrammetric Engineering **32** (1966) 444–462
12. Conrady, A.: Decentering lens systems. Monthly Notices of the Royal Astronomical Society **79** (1919) 384–390
13. Brown, D.C.: Close-range camera calibration. Photogrammetric Engineering **37** (1971) 855–866
14. Devernay, F., 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
15. Chen, S.E.: Quicktime vr c an image-based approach to virtual environment navigation. In: Computer Graphics (SIGGRAPH'95). (1995) 29–38
16. Goshtasby, A.A.: 2-D and 3-D Image Registration. John Wiley & Sons, Hoboken, NJ (2005)
17. Wolberg, G.: Digital Image Warping. IEEE Computer Society Press, Los Alamitos, CA (1990)
18. Xu, W., Penners, J., Mulligan, J.: Recording real worlds for playback in a virtual exercise environment. Computer Science CU-CS 1013-06, University of Colorado at Boulder, Boulder, CO (2006)
19. FitLinxx: Communications SpecificAtion for Fitness Equipment (CSAFE). `http://www.fitlinxx.com/CSAFE/` (2004)