

Using the Xbox Kinect Sensor for Positional Data Acquisition

Jorge Ballester^{a)}

Physics Department, Emporia State University, Emporia, Kansas, 66801.

Chuck Pheatt^{b)}

Computer Science Department, Emporia State University, Emporia, Kansas, 66801.

The Kinect sensor was introduced in November 2010 by Microsoft for the Xbox 360 video game system. The sensor unit is connected to a base with a motorized pivot. It was designed to be positioned above or below a video display and to track player body and hand movements in 3D space, which allows users to interact with the Xbox 360. The device contains a RGB camera, depth sensor, IR light source, three-axis accelerometer and multi-array microphone, as well as supporting hardware that allows the unit to output sensor information to an external device. In this article we evaluate the capabilities of the Kinect sensor as a data acquisition platform for use in physics experimentation. Data obtained for a simple pendulum, a spherical pendulum, projectile motion and a bouncing basketball are presented. Overall, the Kinect is found to be both qualitatively and quantitatively useful as a motion data acquisition device in the physics lab.

I. INTRODUCTION

The use of imaging technology to capture motion data in the physics lab has a long history. Intensive pedagogical use dates back to at least the use of strobes and moving objects with blinking lights (e.g. the widely used “blinky”).¹ The visual record of the experiment was created with a Polaroid Land camera using a long time exposure setting. As technologies have been developed and become more affordable, they have been incorporated into the physics lab. The development of videocassette recorders (VCR) and the possibility of advancing the video record frame-by-frame was used for pedagogical investigations of topics such as the underdamped pendulum.² Data extraction techniques included placing a transparent sheet on the screen and marking the sheet as the recording was advanced frame-by-frame. Extensive pedagogical materials for the study of motion graphs, including prerecorded videos, were also developed.³ The development of computer video capabilities and software gradually simplified data capture (Ref. 4), especially with the introduction of point-and-click tools such as VideoPoint™.⁵

Alongside these technological improvements, were significant discussions of the pedagogical effectiveness of video techniques in improving student understanding of particle motions.⁶ Video analysis techniques were also gradually adapted for use in studying intermediate physics concepts.⁷ The aforementioned imaging technologies are limited to providing a record of one or two-dimensional motion in a plane perpendicular to the line of sight. Scaling image distances to real world distances may be done with a reference object of known size in the image. Other motion tracking technologies have been developed in parallel with imaging technologies. For example, ultrasonic motion detectors have been used extensively in introductory physics labs. In some ways, motion detectors challenge video analysis in terms of pedagogical effectiveness.⁶ The limitations of video analysis include the 30 frames per second (fps) video standard. This temporal resolution is adequate for video playback, but it can be a limit in motion studies. As has been noted by users, 30 fps can make precise numerical differentiation to obtain velocities and acceleration difficult.² Alternative techniques providing much higher temporal resolution but lacking video images have been available for some time.⁸ Affordable high-speed cameras capable of up to 1000 fps, such as the Casio EX-FH20 have become available recently.⁹

As has been the case with many previous technological innovations, The Kinect sensor for the Xbox 360 video game system has potential applications in the physics laboratory. The Kinect sensor was introduced in November 2010 by Microsoft. The sensor unit is connected to a base with a motorized pivot. It was designed to be positioned above or below a video display and to track player body and hand movements in 3D space, which allows users to

interact with the Xbox 360. The Kinect contains a RGB camera, depth sensor, IR light source, three-axis accelerometer and multi-array microphone, as well as supporting hardware that allows the unit to output sensor information to an external device. In this article we evaluate the capabilities of the Kinect sensor as a data acquisition platform for use in physics experimentation. Several sample experiments demonstrating the sensor's use in acquiring positional data are provided.

II. SENSORS

The RGB and depth images from the unit are of greatest interest for the purposes of this paper. The RGB and depth hardware used in the Kinect were developed by PrimeSense.¹⁰ Both the RGB and depth images have a resolution of 640×480 pixels. The unit generates a RGB 8-bit color graphics video stream. The unit's depth image sensing hardware provides 11-bit depth data for each pixel. A PrimeSense patent (Ref. 11) notes that the "technology for acquiring the depth image is based on Light CodingTM. Light Coding works by coding the scene volume with near-IR light." A near-IR light source and diffuser are used to project a speckle pattern onto the scene being assessed. An example image of the speckle pattern and a discussion of its properties are available.¹² The image of the speckle pattern that is projected onto an object is then compared with reference images to identify the reference pattern that correlates most strongly with the speckle pattern on the object. This process provides an estimate of the location of the object within the sensor's range. Cross-correlation between the speckle pattern on the object and the identified reference pattern is then used to map the object's surface.¹¹ Note that the RGB and depth sensors are offset from one another in the Kinect unit by approximately 2.5 cm, yielding offset view-points. A viewing transformation must be applied to allow the images generated to have the same point of view.

Estimates of the Kinect depth sensor's ranging limit vary from 0.8-3.5 m (Ref. 13) to 0.7-6.0 m (Ref 14). The angular field of view for both the RGB and depth sensors' is approximately 57° horizontal by 43° vertical.¹⁴ Both sensors acquire data at the rate of 30 frames per second (fps). The Kinect uses a USB type A connector which may be attached to a personal computer (PC) with USB input capability.

Software that allows the device to be connected to a PC has been available since December, 2010.^{15,16} These software suites allow data to be acquired by the sensor and manipulated independently of the Xbox gaming unit. The authors have utilized and modified the aforementioned software to process Kinect output and acquire the 3D positional data discussed in this paper.

III. DEPTH SENSOR EVALUATION

Depth sensor data was first evaluated by collecting raw data using the Ajax software suite.¹⁵ Raw depth data (D_{raw}) from the Kinect is provided in an 11-bit form, with a potential range of values 0-2047. A test range consisting of a composite grid of 0.15 m square targets was evaluated. Measurements collected from the Kinect unit on the targets verify that the Kinect distance measurements are aligned in a rectangular grid with respect to the horizontal, vertical and depth planes. From these measurements, a regression equation relating raw depth sensor values (D_{raw}) and actual depth (D_{meters}) was developed:

$$D_{\text{meters}} = \frac{\beta_1}{\beta_0 - D_{\text{raw}}}, \quad \beta_0 = 1090.7 \pm 0.2, \quad \beta_1 = 355.1 \pm 0.5 \quad (1)$$

A plot of observed depth data values and the equation above is presented in Fig 1. It is of note that the raw depth values for depths of 0.6 m to 4.3 m are approximately 500 to 1000, utilizing considerably less than one-half of the available 0 to 2047 range. The sensor vendor's documentation (Ref. 13) specifies a 1 cm depth resolution at 2 m. Our evaluation confirms this statement. However, as the depth increases, the device's resolution decreases as the reciprocal of the distance. At a depth of 4 m, the resolution is reduced to approximately 2.7 cm. This change in resolution at greater depths is directly due to the relationship between D_{raw} and D_{meters} as defined in Eq. (1).

The unit's performance was also evaluated with the OpenNI software suite which uses software components from PrimeSense.¹⁶ This software provides processed depth ($D_{\text{processed}}$) information, reporting depth values over a range of 0 to 10,000. Using an evaluation technique similar to the one used in the previous evaluation for assessing the raw data, the processed depth data was found to have somewhat poorer resolution characteristics than those found using the raw data from the Ajax suite. As depth increases, the resolution in the horizontal and vertical planes is also reduced. This is illustrated in Fig. 2. As one moves further away from the Kinect sensor (from depth A to depth B in Fig. 2), the area represented by each of the 640×480 sensor pixels increases as a function of distance, reducing resolution. Any two objects appearing within a single sensor pixel-cell are indistinguishable. Based on $D_{\text{processed}}$ measurements, a regression equation relating processed depth sensor values versus depth was generated as:

$$D_{\text{meters}} = \beta_1(\beta_0 + D_{\text{processed}}), \quad \beta_0 = 0.004 \pm 0.003, \quad \beta_1 = 0.001007 \pm 0.000001 \quad (2)$$

Processed depth values reported by the OpenNI suite essentially present depth in mm, so a user transformation to a linear depth scale is not required. Therefore, the OpenNI suite defines a depth range of 0-10 m relative to the Kinect unit. Provided in Fig. 3 is a plot of the resolution values for the horizontal and vertical planes as well as for the depth measurements. Depth resolution values for 2 m and 4 m distances were found to be 1.1 cm and 4.6 cm respectively. Based on depth information and the depth sensor field of view, resolution in the horizontal and vertical planes was calculated. Both horizontal and vertical resolutions vary from 0.35 cm to 0.70 cm at 2 m and 4 m, respectively.

Although the OpenNI suite exhibits poorer depth resolution than using the raw data values from the Ajax suite, several features built into the OpenNI suite make it more desirable as a data collection tool. These include individual image time-stamp information as well as a built-in viewing transformation allowing the depth and RGB images to have the same point of view. The OpenNI software was used in all subsequent evaluations.

IV. MAPPING THE KINECT TO A 3D EXPERIMENTAL REGION

The Kinect output naturally lends itself to the use of a 3D rectangular coordinate system (see Fig. 2). Given the aforementioned depth range of 0-10 m and a sensor field view of 57° horizontally and 43° vertically, this naturally defines an experimental space of 12 m in the x -direction, 9 m in the y direction, 10 m in the negative z direction. For convenience, the coordinates of the Kinect sensor unit are defined to be ($x = 6$ m, $y = 4.5$ m, $z = 0$ m) with respect to the experimental space. The following equations map Kinect outputs (depth and pixels) to the following right-handed 3D coordinate system:

$$z_{\text{meters}} = -\frac{D_{\text{processed}}}{1000}, \quad (3)$$

$$x_{\text{meters}} = 6 - 2 \left(\text{Pixel}_x - \frac{639}{2} \right) \tan \left(\frac{57^\circ}{2} \right) \frac{z_{\text{meters}}}{640}, \quad (4)$$

$$y_{\text{meters}} = 4.5 - 2 \left(479 - \text{Pixel}_y - \frac{479}{2} \right) \tan \left(\frac{43^\circ}{2} \right) \frac{z_{\text{meters}}}{480}. \quad (5)$$

Pixel_x and Pixel_y represent a single pixel in the xy -plane associated with a depth measurement. Note that by convention, the pixels are numbered horizontally from 0 to 639 and vertically from 0 to 479, with the origin positioned in the upper left hand corner of the images. The 3D coordinates calculated in this way are in a single octant.

The Kinect's 3D rectangular coordinate system translates and rotates rigidly with the Kinect unit's orientation. In our experimentation, alignment in all three axes was accomplished using an $1 \text{ m} \times 1 \text{ m}$ target in the shape of a

Greek cross with equal sized arms mounted on a standard (see Fig. 2). First, the target is aligned plumb to the Earth's surface using a simple bubble level. The Kinect unit is then positioned so that the target is level, centered and at a constant depth when viewed by the Kinect. Although not necessary, the Kinect's xy and yz -planes may be made to coincide with the walls of a rectangular experimental space by initially positioning the target squarely in the experimental space.

V. TIMING CONSIDERATIONS

Using a PC for real-time data acquisition can be problematic. Data acquisition devices are typically implemented using dedicated real-time processors that have been designed to acquire data in a lossless fashion based on the desired rate of data capture. PC operating systems on the other hand were not designed to minimize timing variability (timing jitter) with respect to completing tasks. From a typical PC users' perspective, variations in timing of multiple milliseconds are of little concern. However, these timing variations in data acquisition may have a significant effect on the interpretation of results.

The Kinect output is organized into frames consisting of RGB and depth images. The OpenNI software suite provides captured frame numbering as well as frame timestamp information (reported in units of 10^{-5} s). One thousand data frames were captured and the timestamp information analyzed. The average frame to frame capture time was calculated as 0.033338 s, with an expected value of $0.03\bar{3}$ s based on a 30 fps acquisition rate. Maximum frame to frame jitter was found to be 10^{-5} s. Based on this analysis, timing corrections can be made, but may not be necessary for most experimentation.

PC processor speed also plays a role in the ability to acquire data in a lossless fashion. When a data frame is collected from the Kinect, it must be fully processed before the information from the next data frame can be acquired. Failure to keep pace with the Kinect's data generation rate will result in lost or corrupted data. In our experimentation, both RGB and depth images were captured and stored on disk for later post-processing. We found that buffering Kinect output was necessary in order to minimize data loss over extended periods of time (greater than several seconds). It was also noted that a PC with multiple cores and a processor speed of greater than 2.8 GHz assured that data loss would not be an issue for most experimentation.

VI. EXAMPLE EXPERIMENTS

The Kinect was used to digitize several motions commonly studied in physics. The purpose of these experiments is to assess the real-world effectiveness of the Kinect in gathering motion data. Data obtained from the unit was evaluated with regard to its ability to produce qualitative motion patterns and quantitative results that can be compared to values generated by commonly used techniques.

The first experiment was the simple pendulum which is constrained to swing in a plane. The pendulum bob consisted of a metal ball with a radius 2.6 cm and mass 0.5 kg. The length L of the pendulum is (2.30 ± 0.01) m. In the first trial, the plane of the pendulum was perpendicular to the line of sight from the Kinect, i.e. in the xy -plane as previously defined. In this type of experiment, the majority of the motion is along the x -axis with some motion in the y -axis direction as well. The x coordinate as a function of time is presented in Fig. 4. This transverse motion is what is commonly digitized in a video clip using Logger Pro (Ref. 17) or similar software. The Kinect data clearly demonstrates the periodic motion of the bob with well-defined maxima and minima.

In a second trial of the simple pendulum experiment, the plane of the pendulum was in the yz -plane. In this case, the Kinect is providing motion data along the z axis, data that is not readily obtainable from video digitizing software. Fitting a sinusoidal curve to the data shows the period for the xy -plane and yz -plane pendulums to be (3.0635 ± 0.0004) s and (3.0624 ± 0.0004) s, respectively. This compares favorably with a period of (3.042 ± 0.007) s calculated from the small-angle simple pendulum formula:

$$T = 2\pi \sqrt{\frac{L}{g}}. \quad (6)$$

The angular amplitude of the oscillations in both trials was $(15 \pm 1)^\circ$. Including the finite amplitude correction (Ref. 18) to Eq. (6), results in a predicted period of (3.055 ± 0.008) s.

The second experiment consisted of a spherical pendulum wherein the bob is free to move anywhere on a spherical surface defined by the length of the string. For small amplitudes, the pendulum bob's motion is predicted to consist of separate sinusoidal oscillations along the x and z directions with the same period as given by Eq. (6). The two oscillations generally have different amplitudes, resulting in an elliptical orbit in the xz -plane. In practice, deviations from the idealized small-amplitude pendulum produce an approximately elliptical orbit that does not quite close on itself and precesses. For example, in the case of a Foucault pendulum, the pendulum precesses due to the Earth's rotation. With traditional video capture techniques, a camera would be mounted above or below the spherical pendulum. Space limitations would make data acquisition difficult and prone to parallax distortions. With careful alignment, the Kinect can provide motion data for this pendulum as well. A plot of the motion in the xz -plane is presented in Fig. 5. The pendulum bob orbits clockwise in the figure. Two complete orbits separated by 4 minutes are displayed, with the larger ellipse representing the earlier orbit. The decreasing size of the orbit demonstrates the decay of the amplitudes. The precession is in the clockwise direction in Fig. 5 with an average rate of $(78 \pm 2)^\circ$ per minute.

A third experiment that lends itself to qualitative and quantitative analysis is projectile motion. The projectile was a wooden ball with a radius 3.6 cm and a mass 0.158 kg which was tossed several times across a distance of approximately 3 m. The motion was executed in the xy -plane to evaluate the Kinect's ability to analyze transverse motions. In a second trial, the motion was then executed in the yz -plane to evaluate the Kinect's unique ability to analyze motion with a normal component, i.e. along the sensor unit's line of sight.

Projectile motion can be described either as independent horizontal (x or z) and vertical (y) motions in terms of time or as a trajectory through space with the vertical motion plotted versus the horizontal motion. In the first trial the motion is in the xy -plane, it is predicted to follow the standard equations:

$$x = x_0 + v_{x0}t \quad (7)$$

and

$$y = y_0 + v_{y0}t - \frac{1}{2}gt^2. \quad (8)$$

The motion can also be analyzed directly in terms of the trajectory through space, i.e. when time is eliminated the trajectory is predicted to follow:

$$y = y_0 + \frac{v_{y0}}{v_{x0}}(x - x_0) - \frac{g}{2v_{x0}^2}(x - x_0)^2. \quad (9)$$

The dependence of both the horizontal x and vertical y motions on time is plotted in Fig. 6. The x versus t plot displays the linear behavior predicted by Eq. (7) with an estimate of $v_{x0} = (402 \pm 2)$ cm/s. The y versus t plot displays the quadratic behavior predicted by Eq. (8) with $g = (980 \pm 30)$ cm/s². This compares favorably to data collected using video digitizing software. The projectile motion in the yz -plane can also be described using Eqs. (7)-(9) by replacing

x with z . As before, analyzing the z versus t linear relationship produces an estimate of $v_{z0}=(364\pm3)$ cm/s. Analyzing the y versus t quadratic relationship produces an estimate of $g=(960\pm20)$ cm/s².

The final experiment consisted of tracking a series of bounces executed in the yz -plane using an extremely overinflated basketball. The positional data is presented in Fig. 7. The motion between bounces is expected to follow a standard parabolic trajectory as described by Eq. (8). The individual trajectories were each fit with a quadratic equation which is displayed in Fig. 7. The individual fits were then used to estimate the maximum height reached between consecutive bounces. The dotted curve in Fig. 7 is fit to connect the successive maxima. Comparing successive heights can be used to calculate the coefficient of restitution e for the basketball according to the relation

$$e = \sqrt{\frac{h_{n+1}}{h_n}}, \quad (10)$$

where h_{n+1} and h_n are successive heights. Based on the data, $e=0.977\pm0.006$ for all the bounces. As expected, this is higher than the coefficient of restitution of 0.78 ± 0.03 associated with regulation basketballs.¹⁹

VII. CONCLUSIONS

The Kinect provides an inexpensive option for acquiring 3D positional data with a time base. The experimentation performed shows that meaningful qualitative and quantitative data can be acquired and analyzed. The device has limitations with respect to spatial and temporal resolution (data acquisition rate of 30 fps), and object size and speed need to be carefully considered when designing an experiment. Overall the Kinect has been found to be an effective device with the potential for rapidly acquiring data in many diverse situations.

The software used in generating the results presented here is available from the authors.

ACKNOWLEDGEMENT

The authors would like to thank the reviewers for their insightful and thoughtful comments. The authors would also like to thank Andrew Wayman for his work on assessing the Kinect system.

- a) Electronic mail: jballest@emporia.edu
- b) Electronic mail: cpheatt@emporia.edu
- ¹ Physical Science Study Committee, *Physics*, 2nd. Ed. (D. C. Heath and Company, Lexington, MA).
- ² M. S. Greenwood, "Using videotapes to study underdamped motion of a pendulum: A laboratory project," *Am. J. Phys.* **55**, 645-648 (1987).
- ³ D. A. Zollman and R. G. Fuller, "Teaching and Learning Physics with Interactive Video," *Phys. Today*. **47** (4), 41-47 (1994).
- ⁴ M. Marcuso and R. M. Webber, "Kinematical measurements using digital image capture," *Am. J. Phys.* **64**, 1080-1083 (1996).
- ⁵ Videopoint[®], <http://www.lsw.com/videopoint/>.
- ⁶ R. J. Beichner, "The impact of video motion analysis on kinematics graph interpretation skills," *Am. J. Phys.* **64**, 1272-1277 (1996).
- ⁷ W. M. Wehrbein, "Using video analysis to investigate intermediate concepts in classical mechanics," *Am. J. Phys.* **69**, 818-820 (2001).
- ⁸ P. A. DeYoung and B. Mulder, "Studying collisions in the general physics laboratory with quadrature light emitting diode sensors," *Am. J. Phys.* **70**, 1226-1230 (2002).
- ⁹ Casio EX-FH20, http://www.casio.com/products/archive/Cameras/High-Speed/EX-FH20/content/Technical_Specs/.
- ¹⁰ PrimeSense, Tel-Aviv, Israel. <http://www.primesense.com>.
- ¹¹ J. Garcia and Z. Zalevsky, U.S. Patent No. 7,433,024 B2 (7 October 2008).
- ¹² Technical description of Kinect calibration, http://www.ros.org/wiki/kinect_calibration/technical.
- ¹³ PrimeSensor Reference Design, <http://www.primesense.com/?p=514>.
- ¹⁴ Kinect, <http://www.businessinsider.com/blackboard/kinect>.
- ¹⁵ Ajax Software Suite - The software for acquiring unprocessed (raw) data from the Kinect consisted of the V16 OpenKinect for Windows driver (<http://ajaxorg.posterous.com/kinect-driver-for-windows-prototype>) using and libusb 1.0 driver (<http://www.libusb.org>).
- ¹⁶ OpenNI Software Suite - The software for acquiring processed data from the Kinect consisted of interaction software from OpenNI, OPENNI-Win32-1.0.0.23.exe (<http://openni.org/downloadfiles/2-openni-binaries>), PrimeSense NITE 1.3 middleware, NITE-Win32-1.3.0.17.exe (<http://openni.org/downloadfiles/12-openni-compliant-middleware>) and SensorKinect drivers, SensorKinect-Win32-5.0.0.exe (<https://github.com/avin2/SensorKinect/tree/master/Bin>).
- ¹⁷ Vernier Logger Pro 3 software, <http://www.vernier.com/soft/lp.html>.
- ¹⁸ J. B. Marion and S. T. Thornton, *Classical Dynamics of Particles and Systems*, 4th. Ed. (Saunders College Publishing, Philadelphia, PA.).
- ¹⁹ International Basketball Federation (FIBA), *Official Basketball Rules 2010: Basketball Equipment*, San Juan, Puerto Rico, 17th April 2010.

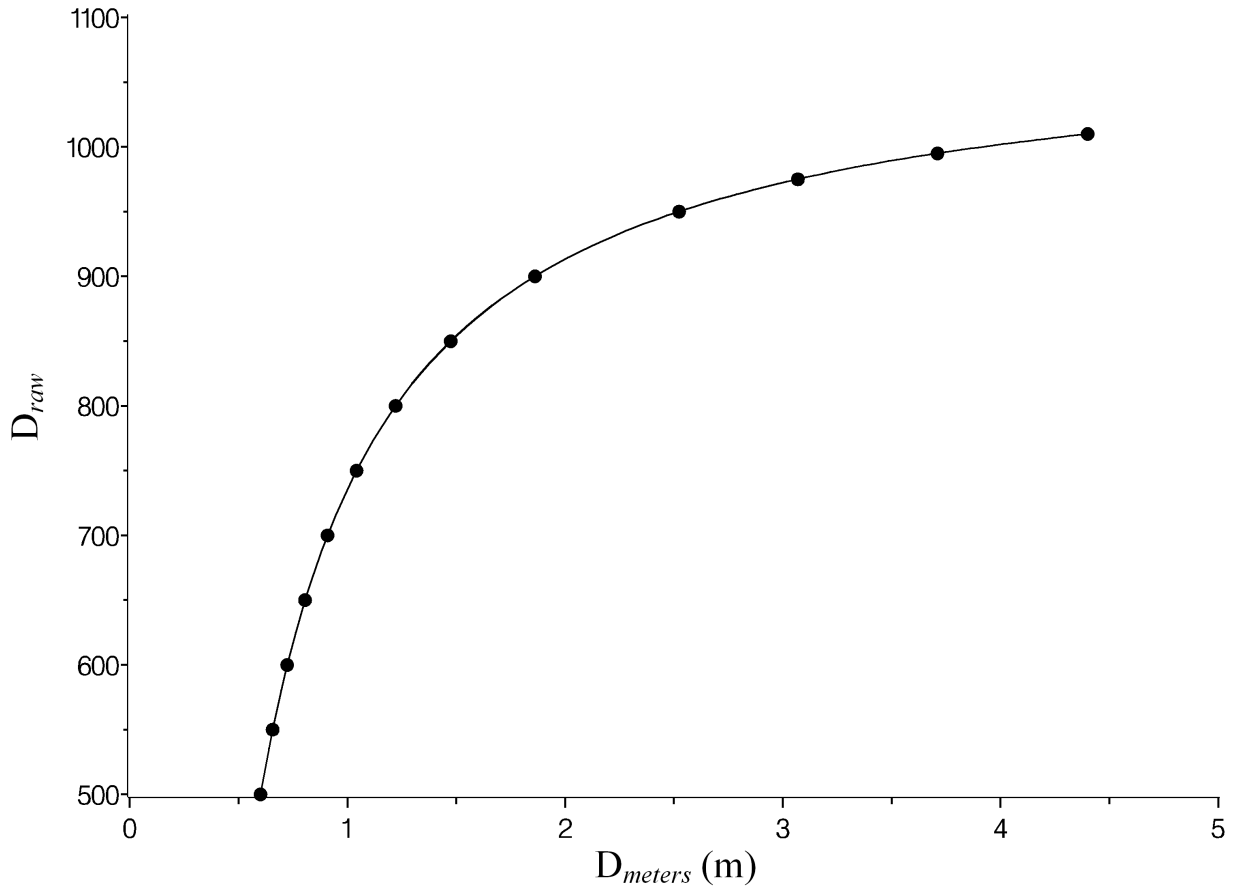


Fig. 1. Calibration curve for raw depth sensor values (D_{raw}).

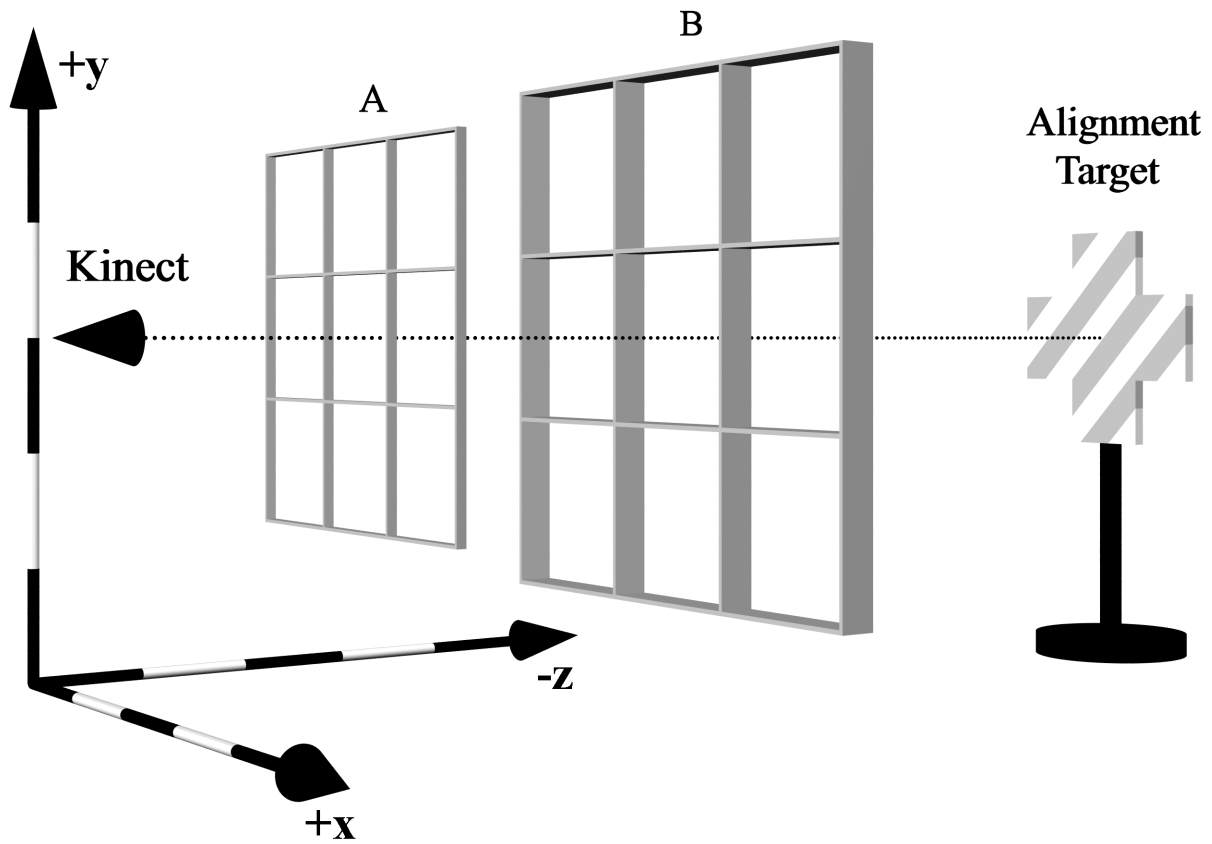


Fig. 2. Kinect alignment and resolution in the experimental space. The coordinates of the Kinect sensor unit are defined to be $(x = 6 \text{ m}, y = 4.5 \text{ m}, z = 0 \text{ m})$ with respect to the experimental space.

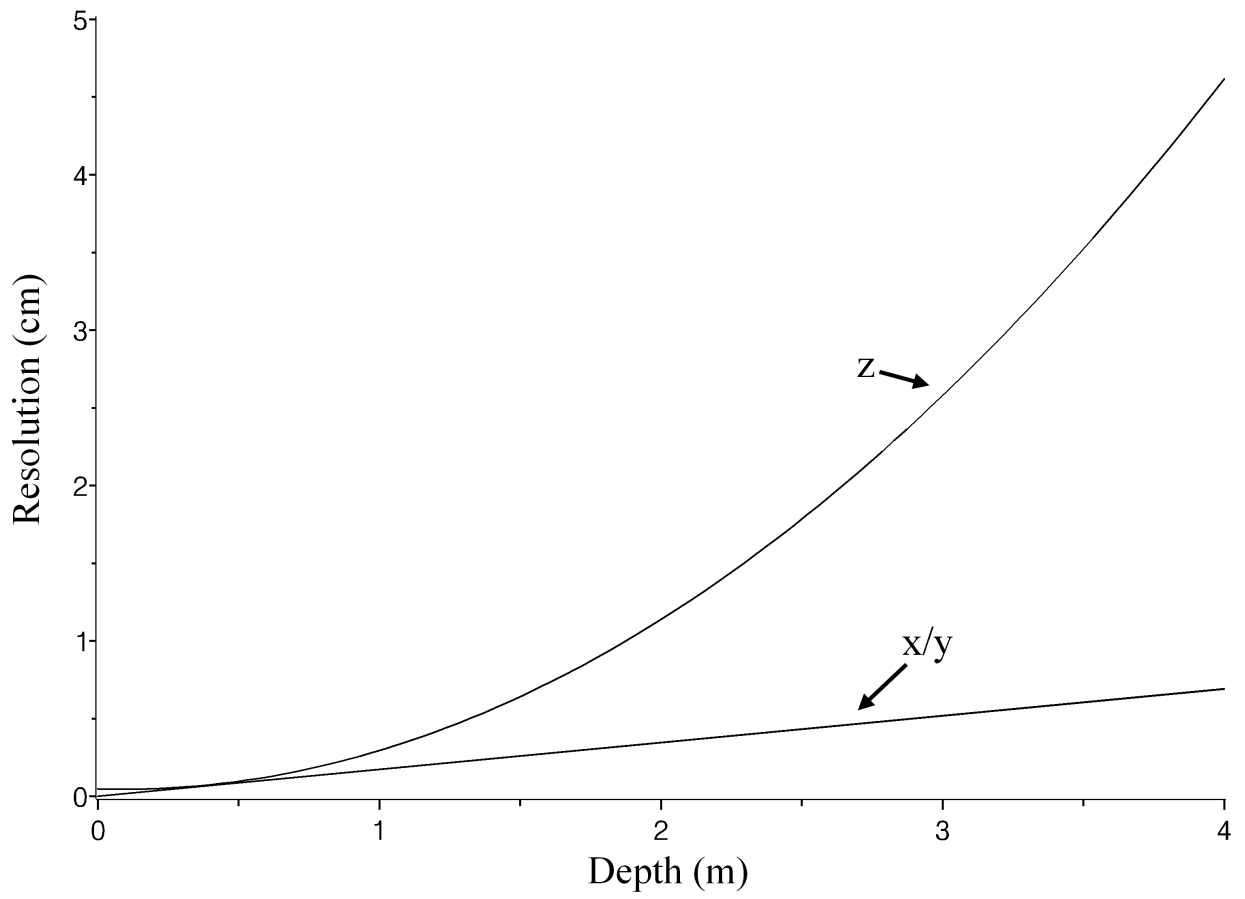


Fig. 3. Kinect spatial resolutions for the horizontal/vertical (x/y) plane and depth (z).

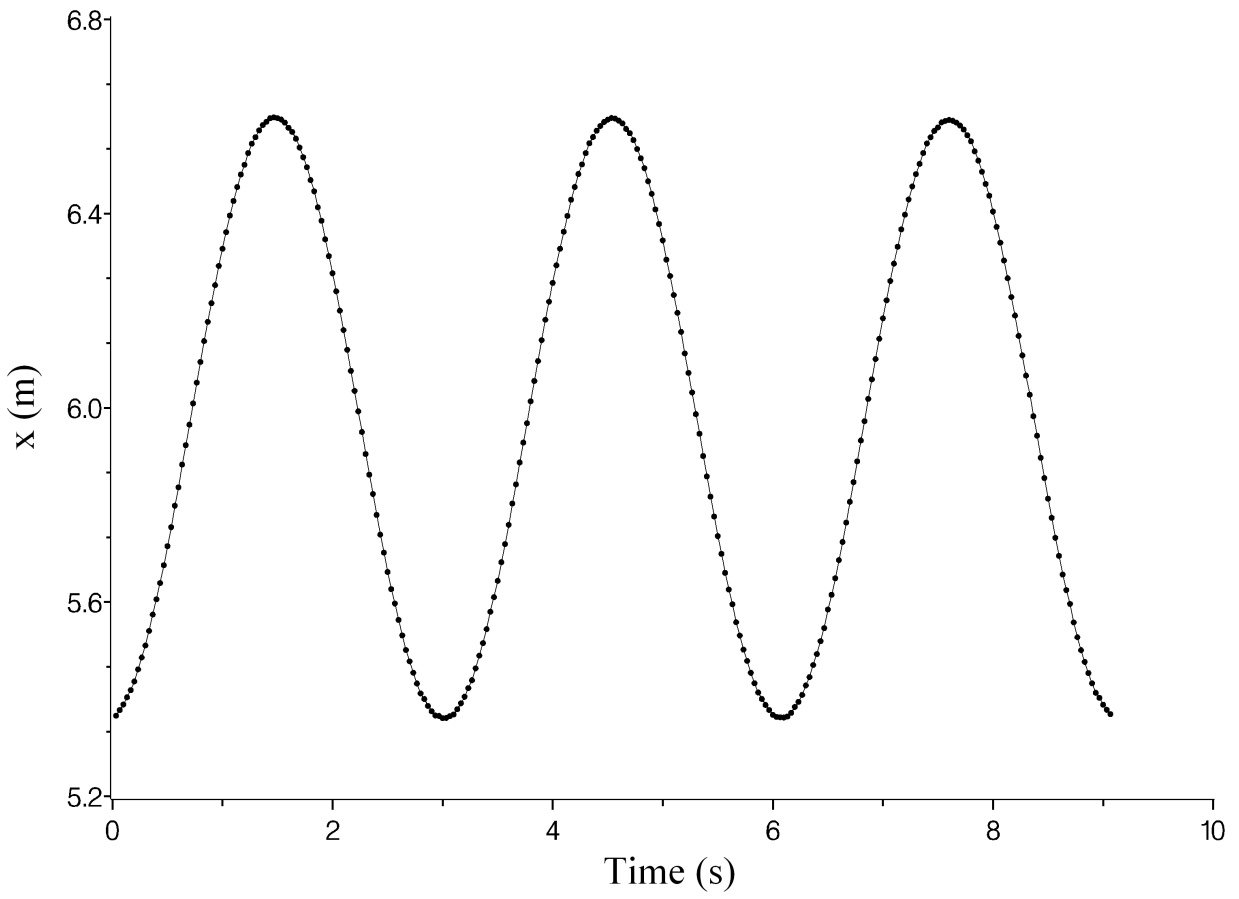


Fig. 4. Transverse oscillations of the simple pendulum.

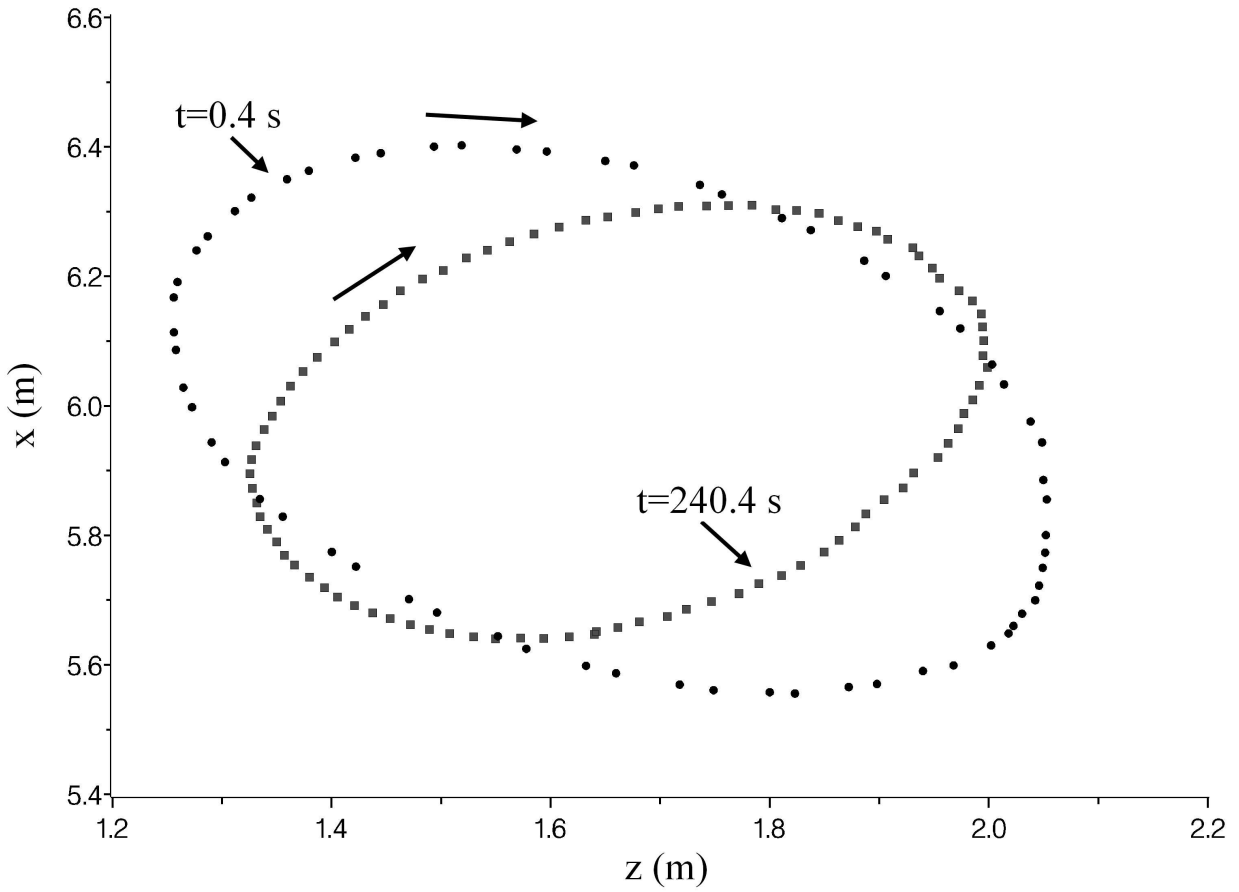


Fig. 5. Two orbits of the spherical pendulum separated by 4 minutes. The precession is in the clockwise direction with an average rate of $(78 \pm 2)^\circ$ per minute.

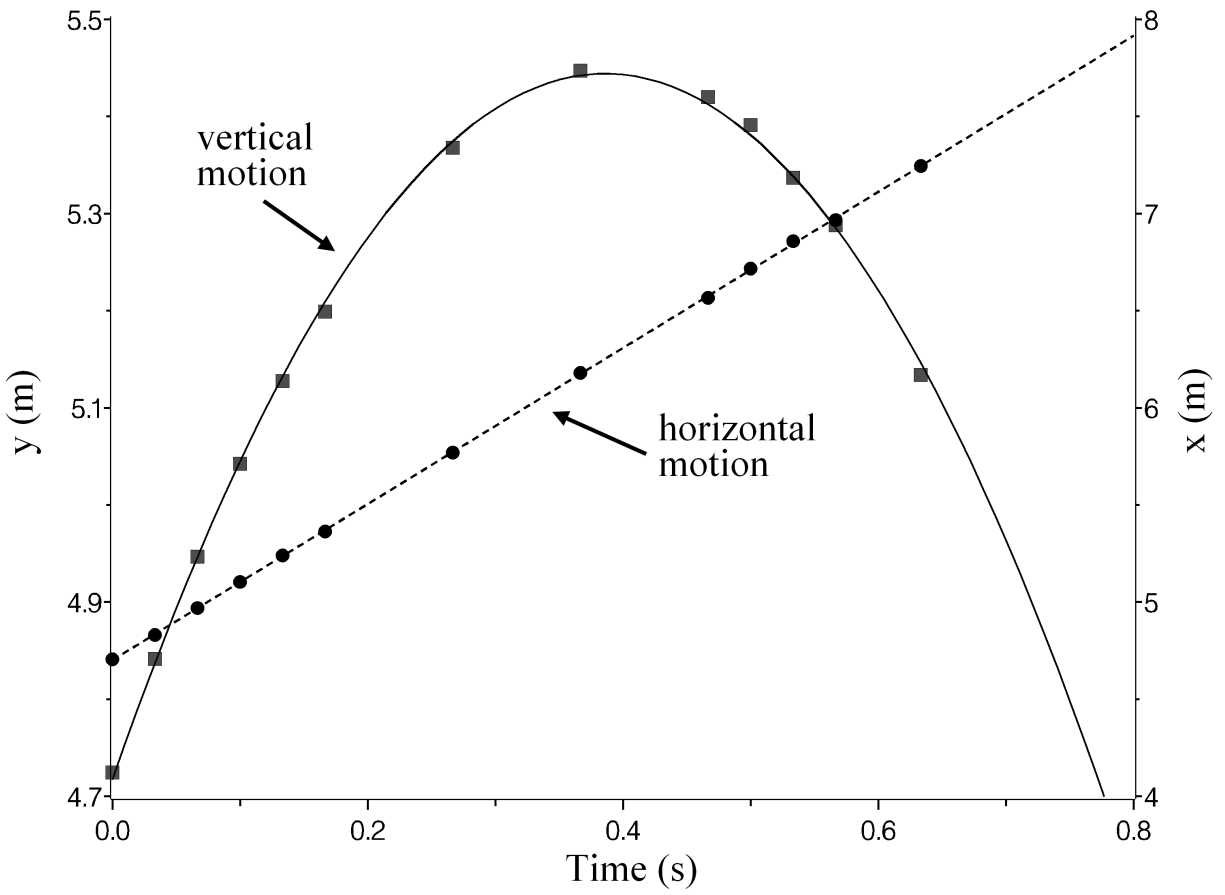


Fig. 6. The vertical (y) and horizontal (x) motions of the projectile.

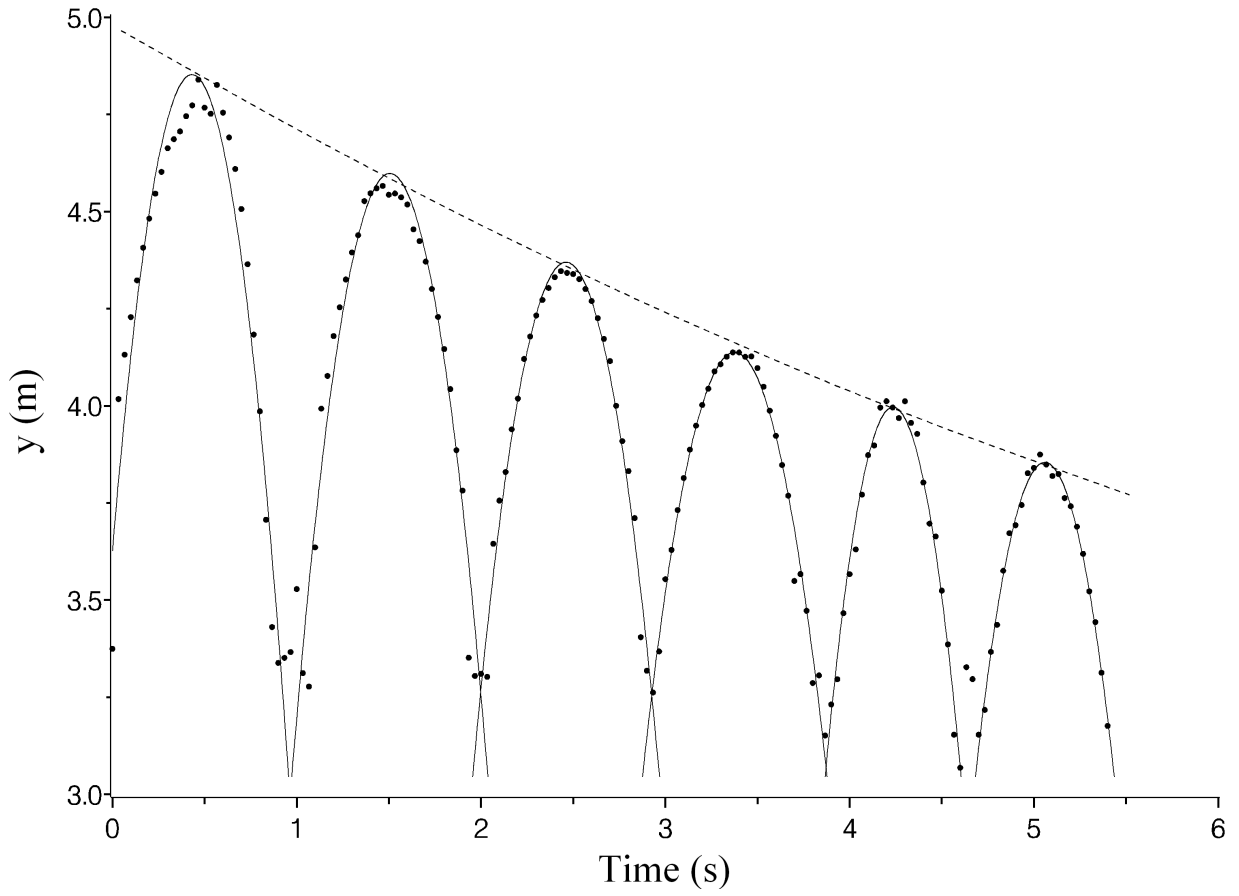


Fig. 7. The vertical motion of a bouncing basketball. The floor is at approximately $y=3.1$ m.