

DRIFT CORRECTION AND OPTICAL-INERTIAL SENSOR DATA FUSION FOR VIRTUAL REALITY

Cross-Reference to Related Applications

The present application claims priority under 35 USC 119(e) to US Provisional Application No.
5 63/447,657 filed February 23, 2023, the entire contents of which is incorporated by reference.

Field of the Embodiments

The field of the invention and its embodiments relate to improving accuracy and reducing
latency when calculating user positional data and correcting for drift in virtual reality systems
10 realized using head-mounted displays (HMDs).

Background

Challenging issues arise when calculating user positional data and correcting for drift in
virtual reality systems realized using head-mounted displays (HMDs). Issues involving accuracy
15 and latency can be particularly challenging.

Review of related technology:

US11244509B2 – This reference describes drift correction for industrial augmented
reality (AR) applications. Specifically, this reference provides a method for increasing the
20 accuracy of AR information presented to a user navigating a real-world, mapped process plant
environment. The method includes numerous process steps, such as: registering, at a server
containing a map database, a mobile device of the user at a landmark within the mapped process
plant environment; in response to registering at the landmark, transmitting, from the server, an
authorization to gain physical access to the mapped process plant environment; and identifying,
25 as the user moves through the mapped process plant environment, a specific object within a field
of view of a camera of the mobile device. The process step of identifying the specific object
includes: predicting, by a machine learning (ML) model processing one or more images captured
by the camera, one or more object types, and identifying a respective node, of a plurality of
nodes within the map database representing the mapped process plant environment, that
30 corresponds to a respective object type for each of the one or more object types. At least some
nodes of the plurality of nodes represent field devices within the process plant environment. A

first respective node corresponds to the specific object. The method also includes: retrieving, from the map database, a location associated with the first respective node; using the retrieved location to update or confirm one or more estimated locations of the mobile device; and causing digital information including each of the one or more object types and an avatar of the first
5 respective node to be superimposed on a real-world view presented to the user via a display of the mobile device. The digital information is selected based at least in part on the updated or confirmed one or more estimated locations of the mobile device. The digital information includes alert information corresponding to the specific object. Further, selection of the avatar enables the user to control operation of the specific object by issuing a voice command.

10 US11036284B2 – This reference describes methods for tracking and drift correction. Further, this reference describes improved user interfaces for interacting with a virtual environment. The virtual environment is presented by a display of a first device having an image sensor. The first device uses the image sensor to determine a relative position and orientation of a second device based on a marker displayed on a display of the second device. The first device
15 uses the determined relative position of the second device to display a representation of the second device including virtual content in place of the marker.

US10809795B2 – This reference describes six degree of freedom tracking with scale recovery and obstacle avoidance. According to this reference, a virtual reality or mixed reality system is configured to preform object detection using a monocular camera. The system is also
20 configured to make the user aware of the detected objects by showing edges or lines of the object within a virtual scene. Thus, the user the user is able to avoid injury or collision while immersed in the virtual scene. In some cases, the system may also detect and correct for drift in the six degree of freedom pose of the user using corrections based on the current motion of the users.

WO2016187760A1 – This reference describes sensor fusion using inertial and image
25 sensors. In particular, this reference provides systems, methods, and devices for controlling a movable object using multiple sensors. A method for calibrating one or more extrinsic parameters of a movable object having a plurality of sensors in an initial configuration is provided. The method comprises numerous process steps, such as: detecting that the initial configuration of the plurality of sensors has been modified; receiving inertial data from at least
30 one inertial sensor during operation of the movable object; receiving image data from at least two image sensors during the operation of the movable object; and estimating the one or more

extrinsic parameters based on the inertial data and the image data in response to detecting that the initial configuration has been modified. The one or more extrinsic parameters comprise spatial relationships between the plurality of sensors in the modified configuration.

5 US10007329B1 – This reference describes drift cancelation for portable object detection and tracking. The technology disclosed in this reference can provide capabilities such as using motion sensors and/or other types of sensors coupled to a motion-capture system to monitor motions within a real environment. A virtual object can be projected to a user of a portable device integrated into an augmented rendering of a real environment about the user. Motion information of a user body portion is determined based at least in part upon sensory information
10 received from imaging or acoustic sensory devices. Control information is communicated to a system based in part on a combination of the motion of the portable device and the detected motion of the user. The virtual device experience can be augmented in some implementations by the addition of haptic, audio and/or other sensory information projectors.

US8957909B2 – This reference describes a system and method for compensating for drift
15 in a display of a user interface state. Specifically, this reference describes a system, a non-transitory computer readable storage medium including instructions, and a method for adjusting a displayed user interface in accordance with a navigational state of a human interface device. For each measurement epoch, a base set of operations are performed, including: determining an unmodified user interface state in accordance with the navigational state, and generating current
20 user interface data. Upon detecting an error introducing state, additional operations are performed, including: determining a modified user interface state; adjusting the current user interface data in accordance with the modified user interface state; and determining a user interface state error. Upon detecting an error compensating state, additional operations are performed, including: determining a compensation adjustment and adjusting the current user
25 interface data and user interface state error in accordance with the compensation adjustment. The current user interface data enables a current user interface to be displayed.

Roberto A. Montano-Murillo, et al., “Drift-Correction Techniques for Scale-Adaptive VR Navigation,” UIST '19: Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, 2019, Pages 1123–1135, doi: 10.1145/3332165.3347914. – This
30 reference describes that scale adaptive techniques for virtual reality (VR) navigation enable users to navigate spaces larger than the real space available, while allowing precise interaction when

required. However, due to these techniques gradually scaling displacements as the user moves (changing user's speed), they introduce a drift effect. That is, a user returning to the same point in VR will not return to the same point in the real space. This mismatch between the real/virtual spaces can grow over time, and turn the techniques unusable (i.e., users cannot reach their target locations). This group characterizes and analyzes the effects of drift, highlighting its potential detrimental effects. Further, this group then proposes two techniques to correct drift effects and then uses a data driven approach (using navigation data from real users with a specific scale adaptive technique) to tune them, compare their performance and chose an optimum correction technique and configuration. This reference shows that the correction technique can significantly reduce drift effects and extend the life-span of the navigation techniques (i.e., time that they can be used before the drift draws targets unreachable), while not hindering the users' experience.

John David Prieto Prada, et al., "Enhanced Location Tracking in Sensor Fusion Assisted Virtual Reality Micro-Manipulation Environments," PLoS ONE, 2021, 16(12): e0261933, doi: 10.1371/journal.pone.0261933. – This reference describes that virtual reality (VR) technology plays a significant role in many biomedical applications. These VR scenarios increase the valuable experience of tasks requiring great accuracy with human subjects. Unfortunately, commercial VR controllers have large positioning errors in a micro-manipulation task. Here, this group proposes a VR-based framework along with a sensor fusion algorithm to improve the micro-position tracking performance of a microsurgical tool. To the best of our knowledge, this is the first application of Kalman filter in a millimeter scale VR environment, by using the position data between the VR controller and an inertial measuring device. This study builds and tests two cases: (1) without sensor fusion tracking and (2) location tracking with active sensor fusion. The static and dynamic experiments demonstrate that the Kalman filter can provide greater precision during micro-manipulation in small scale VR scenarios.

Thomasz Ursel, et al., "Displacement Estimation Based on Optical and Inertial Sensor Fusion," Sensors, 2021, 21, 1390, doi: 10.3390/s21041390. – This reference aims to develop a system capable of estimating the displacement of a moving object with the usage of a relatively cheap and easy to apply sensors. There is a growing need for such systems, not only for robots, but also, for instance, pedestrian navigation. In this paper, the theory for this idea, including data postprocessing algorithms for a MEMS accelerometer and an optical flow sensor (OFS), as well as the developed complementary filter applied for sensor fusion, are presented. In addition, a

vital part of the accelerometer's algorithm, the zero velocity states detection, is implemented. It is based on analysis of the acceleration's signal and further application of acceleration symmetrization, greatly improving the obtained displacement. A test stand with a linear guide and motor enabling imposing a specified linear motion is built. The results of both sensors' testing suggest that the displacement estimated by each of them is highly correct. Fusion of the sensors' data gives even better outcomes, especially in cases with external disturbance of OFS. The comparative evaluation of estimated linear displacements, in each case related to encoder data, confirms the algorithms' operation correctness and proves the chosen sensors' usefulness in the development of a linear displacement measuring system.

Vitor Reus, et al., "Correcting Drift, Head and Body Misalignments between Virtual and Real Humans," SBC Journal on 3D Interactive Systems, 2013, 4(2), Pages 55-65. – This reference describes that head-mounted displays (HMD) allow a personal and immersive viewing of virtual environments, and can be used with almost any desktop computer. Most HMDs have inertial sensors embedded for tracking the user head rotations. These low-cost sensors have high quality and availability. However, even if they are very sensitive and precise, inertial sensors work with incremental information, easily introducing errors in the system. The most relevant is that head tracking suffers from drifting. In this paper, this group presents important limitations that still prevent the wide use of inertial sensors for tracking. For instance, to compensate for the drifting, users of HMD-based immersive VEs move away from their suitable pose. This group also proposes a software solution for two problems: prevent the occurrence of drifting in incremental sensors and avoid the user from move its body in relation to another tracking system that uses absolute sensors (e.g. MS Kinect). This group then analyzes and evaluates the proposed solutions experimentally, including user tests. Results show that the groups comfortable pose function is effective on eliminating drifting, and that it can be inverted and applied also to prevent the user from moving their body away of the absolute sensor range. The efficiency and accuracy of this method makes it suitable for a number of applications in immersive VR.

Sharmin Majumder, et al., "Vision and Inertial Sensing Fusion for Human Action Recognition: A Review," IEEE Sensors Journal, 2020, 21(3), Pages 2454-2467. – This reference describes that human action recognition is used in many applications such as video surveillance, human-computer interaction, assistive living, and gaming. Many papers have appeared in the literature showing that the fusion of vision and inertial sensing improves recognition accuracies

compared to the situations when each sensing modality is used individually. This paper provides a survey of the papers in which vision and inertial sensing are used simultaneously within a fusion framework in order to perform human action recognition. The surveyed papers are categorized in terms of fusion approaches, features, classifiers, as well as multimodality datasets considered. Challenges as well as possible future directions are also stated for deploying the fusion of these two sensing modalities under realistic conditions. Various umbrellas exist. However, their means of operation are substantially different from the present disclosure, as the other inventions fail to solve all the problems taught by the present disclosure.

10

Summary

Method and system for improving accuracy and reducing latency when calculating user positional data and correcting for drift in virtual reality systems realized using head-mounted displays (HMDs). An optical motion capture system consists of a plurality of stationary cameras that track a device on board a head mounted display (HMD). Position and orientation data are obtained from the optical motion capture system, and full rotational data is obtained from a sensor on-board the HMD. Yaw refers to vertical rotational data for rotations of a neck of a wearer. Thus yaw data is a component of the full rotational data from the HMD, which is used in embodiments for space matching. The yaw data coordinate system, a coordinate system specifically focused on vertical rotational aspects, which is useful for inertial corrections and calculations, is matched to vertical rotational components of the optical motion capture coordinate system. But because inertial data is subject to drift, the offset between the coordinate spaces of the two systems must be continuously updated to avoid the HMD from going out of sync with the spatially consistent environment. Focusing on the vertical component is sufficient to account for the majority of drift and to perform space-matching corrections in an efficient manner. Drift correction is made while the user is moving their head to avoid an apparent jump.

15

20

25

Brief Description of the Drawings

No drawings

Detailed Description

30

Embodiments of the present invention will now be described. These embodiments are examples provided by way of explanation of the principles of the present invention, which is not

intended to be limited thereto. Rather, the scope of the invention is defined not by the example embodiments, but by the claims. In fact, those of ordinary skill in the art may appreciate upon reading the present specification that various modifications and variations can be made to the example embodiments.

5 As used herein, the singular forms “a,” “an,” and “the,” are intended to include the plural forms as well, unless the context clearly indicates otherwise.

 The phrase “and/or,” as used herein in the specification and in the claims, should be understood to mean “either or both” of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Thus, as a non-
10 limiting example, a reference to “A and/or B”, when used in conjunction with open-ended language such as “comprising” can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

 As used herein in the specification and in the claims, the phrase “at least one,” in
15 reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the
20 list of elements to which the phrase “at least one” refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, “at least one of A and B” (or, equivalently, “at least one of A or B,” or, equivalently “at least one of A and/or B”) can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally
25 including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

 It will be further understood that the terms “comprises,” “comprising,” “includes,” and/or
30 “including,” when used herein, specify the presence of stated features, integers, steps, operations, elements, and/or, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, and/or groups thereof.

Disclosed herein is an optical motion capture system comprising many stationary cameras that track one or more head mounted displays (HMDs) that contain IR emitting LEDs.

Optical tracking (specifically, outside-in optical tracking) has the benefits of being spatially consistent and does not suffer from drift or accumulated error, and tracks both position and orientation. However optical tracking suffers from latency at least because the optical tracking information (OTI), including optical data, is typically resolved on a remote machine. Therefore, OTI data needs to be transmitted from the cameras to the remote machine. The remote machine performs calculations based on the OTI data, and sends imagery derived from the OTI data to the HMD. Optical tracking is also less precise than inertial methods.

Inertial methods, which perform measurements at the HMD, are lower latency than OTI methods, but suffer from drift between the imagery derived at the HMD and imagery of the environment. The HMD, and sensors therein, captures full rotational data. The yaw or vertical component is key for efficient space matching..

The herein disclosed method combines these two types of tracking data such that a user wearing an HMD gains an optimal immersive experience, or at least an immersive experience better than that which may be obtained using either the optical tracking data or the inertial tracking data alone.

The method comprises using positional and orientation data from an optical motion capture system, and combining it with rotational (yaw) data from an inertial sensor on-board the HMD. The inertial yaw data is derived in an arbitrary coordinate system, with special attention paid to the vertical axis. Accordingly, the coordinate system of the yaw data needs to be matched to the coordinate system of the optical motion capture data.

In order to do this it must be known, for a given inertial sample, which optical sample corresponds to the inertial sample. Because the optical systems incur latency but inertial systems do not, the inertial sample data need to be stored so that when data of an optical sample is received, it can be determined which inertial sample it corresponds to. For this to be accomplished, the stored optical data must contain a measurement of the latency incurred, which indicates how long ago an optical data sample was captured by motion capture cameras, with reference to the system clock of the machine performing the time-matching calculation. Thus, latency between the two readings is accounted for.

An offset between the coordinate spaces of the two systems is then calculated for the time-matched samples. Further, because inertial data is subject to drift but the optical data is not, the offset must be continuously updated to correct for the HMD-based imagery from going out of sync with the spatially consistent environment imagery. After a drift correction is calculated, it is
5 used to modify the imagery displayed in the HMD only while the user is moving their head. This mitigates the user noticing any discontinuities in the imagery presented on the HMD display resulting from the drift correction.

The descriptions of the various embodiments of the present invention have been
10 presented for purposes of illustration, but are not intended to be exhaustive or limited to the embodiments disclosed. Many modifications and variations will be apparent to those of ordinary skill in the art without departing from the scope and spirit of the described embodiments. The terminology used herein was chosen to best explain the principles of the embodiments, the practical application or technical improvement over technologies found in the marketplace, or to
15 enable others or ordinary skill in the art to understand the embodiments disclosed herein.

Although embodiments of the invention have been described with a certain degree of particularity, it is to be understood that the present disclosure has been made only by way of illustration and that numerous changes in the details of construction and arrangement of parts may be resorted to without departing from the spirit and the scope of the invention as determined
20 by the claims.

Claims

What is claimed is:

1. A motion capture system for use in a virtual reality (VR) and/or augmented reality
5 (AR) environment, the system comprising:
an optical tracking system that tracks position and orientation, that does not suffer from drift or accumulated error, but does incur latency, including:
at least one infrared (IR) light emitting diode (LED);
a head-mounted display (HMD) coupled to the LED;
10 a plurality of stationary cameras individually configured to detect the LED,
determine range and direction information from the camera to the LED, and transmit the range and direction information, wherein the stationary cameras are disposed at distributed locations within a mapped process plant environment;
an optical tracking information receiver configured to receive the range and
15 direction information transmitted by the cameras as optical tracking information;
a processor operatively coupled to the receiver, configured to resolve, using the received range and direction information, a location of the LED within a mapped process plant environment; and
a transmitter operatively coupled to the processor that transmits the resolved
20 location of the LED to the HMD; and
an inertial tracking system that tracks a rotation (yaw) from full rotational data of the HMD within the mapped process plant environment, that suffers from drift, but has a lower latency than the optical tracking system, including:
an inertial sensor coupled to the HMD that detects rotation (yaw) information of
25 the sensor and transmits the yaw information; and
an inertial tracking information receiver, operatively coupled to the processor and configured to receive the yaw information transmitted by the inertial sensor as inertial tracking information;
wherein the processor is configured to fuse the optical tracking information with the
30 inertial tracking information such that a user wearing the HMD obtains an improved immersive

experience over the experience obtained from the optical tracking system or the inertial tracking system alone.

2. The motion capture system of claim 1, wherein the optical tracking system is an outside-
5 in optical tracking system, and wherein space matching occurs with mathematical corrections that consider both the incurred latency of the optical tracking system and the yaw information from the HUD compared to equivalent information from the optical tracking system, wherein the full rotational data, of which the yaw information is a component, has minimal latency compared to the incurred latency.

10 3. A method of capturing motion information of a target within a generated virtual reality (VR) and/or augmented reality (AR) environment, the method comprising:

receiving, by a plurality of stationary cameras, infrared (IR) light emitted by at least one light emitting diode (LED) coupled to a head-mounted display (HMD);

15 determining, by one of the stationary cameras, range and direction information from the camera to the LED, and orientation information of the HMD;

transmitting the range, direction, and orientation information as optical tracking information (OTI);

receiving, by an optical tracking information (OTI) receiver, the transmitted OTI;

20 resolving, by a processor operatively coupled to the OTI receiver, a location and orientation of the HMD within a mapped process plant environment using the received OTI; and

transmitting, by a transmitter operatively coupled to the processor, the resolved location and orientation to the HMD; and

determining, via information from a sensor coupled to the HMD, rotation (yaw)

25 information of the sensor;

transmitting the yaw information as inertial tracking information (ITI);

receiving the ITI by an ITI receiver operatively coupled to the processor;

30 combining, by the processor, the OTI and the ITI such that a user wearing the HMD obtains an improved immersive experience compared to an analogous immersive experience obtained using information from the optical tracking system or the inertial tracking system alone.

4. The method of claim 3, wherein the stationary cameras are disposed at distributed locations within the mapped process plant environment.
5. The method of claim 3, further comprising:
- 5 obtaining the position and orientation data from the optical tracking system as optical data samples with corresponding latency information;
- obtaining yaw data in an arbitrary coordinate system from the inertial tracking system as inertial data samples without latency information; and
- 10 matching ones of the inertial data samples to corresponding ones of the optical data samples.
6. The method of claim 5, further comprising:
- storing the ones of the inertial data samples at least until the corresponding ones of the optical data samples are received;
- 15 responsive to receiving ones of the optical data samples, determining which one of the stored inertial data samples the optical data sample corresponds to.
7. The method of claim 6, further comprising:
- performing a time-matching calculation to match the ones of the optical data samples' latency information with the system clock of the machine having the processor performing the time-matching calculation;
- 20 calculating an offset between the coordinate spaces of the OTI and the ITI systems.
8. The method of claim 7, wherein the offset is continuously calculated as drift corrections to images derived from the ITI presented in the HMD, to keep the ITI-derived images in sync with the spatially consistent environment images derived from the OTI, wherein spatial consistency uses space matching based on the yaw component of the full rotational data obtained from the HMD, wherein portions of the full rotational data other than the yaw component are not utilized for the space matching.
- 30

9. The method of claim 8, wherein the drift corrections are presented in the HMD only while the user wearing the HMD is moving their head, so as to mitigate the user noticing discontinuities in the imagery presented on the HMD display.