Technical Disclosure Commons

Defensive Publications Series

August 2020

Vergence and Gaze Based 3D Mapping of Local Environment

Anonymous

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation

Anonymous, "Vergence and Gaze Based 3D Mapping of Local Environment", Technical Disclosure Commons, (August 03, 2020)
https://www.tdcommons.org/dpubs_series/3482

This work is licensed under a Creative Commons Attribution 4.0 License.
This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

Vergence and Gaze Based 3D Mapping of Local Environment

ABSTRACT

This disclosure describes techniques for utilizing a user's natural perception of their environment to obtain information about the local environment. With user permission, eye positions and eye movements of a user as they interact with their environment are utilized to substitute for and/or augment machine-based perception of the environment. Eye movement information such as vergence, fixation/gaze points, saccade patterns, etc. are obtained from eye tracking sensors. A 3D volumetric map of the environment is created based on the eye movement information. Based on identified user fixations, an interest map of the room is generated. 3D spatial information can be obtained in user environments where an outward facing camera is infeasible, since mapping can be performed based only on vergence information obtained from the eyes.

KEYWORDS

- Virtual Reality (VR)
- Augmented Reality (AR)
- Vergence
- Volumetric mapping
- Eye tracking
- Visual fixation
- Saccades
- Head mounted device (HMD)
- Stereovision

BACKGROUND

Wearable devices such as augmented reality (AR) and virtual reality (VR) headsets utilize on-board sensors for functions such as head tracking, user orientation, spatial localization, etc. to better enable the user (wearer) to interact with the environment. For example, the sensors provide information about the dimensions of a room where the user is situated, location information of objects and/or other people in the room, etc.

DESCRIPTION

This disclosure describes techniques for utilizing a user's natural perception of their environment to obtain information about the user environment. Per techniques of this disclosure, with user permission, eye positions and eye movements of a user as they interact with their environment are utilized to substitute for and/or augment sensor-based perception of their local environment.

Eye positions and eye movements of a user can be tracked by a variety of sensors, e.g., optical tracking using cameras mounted on a wearable device that include the user's eyes in their field of vision. The magnitude of eye movements is directly related to a distance of the object of the user's interest. Based on the tracked eye positions and eye movements, distances of different object(s) of interest are estimated, and a three dimensional (3D) mapping of the user environment is determined.

Human vision is based on light from an object of interest falling on the fovea centralis, a region of high visual acuity in their retina. As a user scans their environment for object(s) of interest, they move their eyes around (vergence) to ensure that light from the object(s) falls on the fovea centralis. In addition, for accurate stereo vision (3D depth perception), the user's eyes move such that images formed from the object fall on corresponding points in each eye. The combination of the images in the user's brain provides accurate 3D depth perception.

Having information from two eyes enables the brain to determine the depth and distance of an object (stereovision) and provides perception of three-dimensionality in vision. Both eyes must point accurately towards the object such that the object of regard falls on corresponding points of corresponding eye retinas to stimulate stereovision.

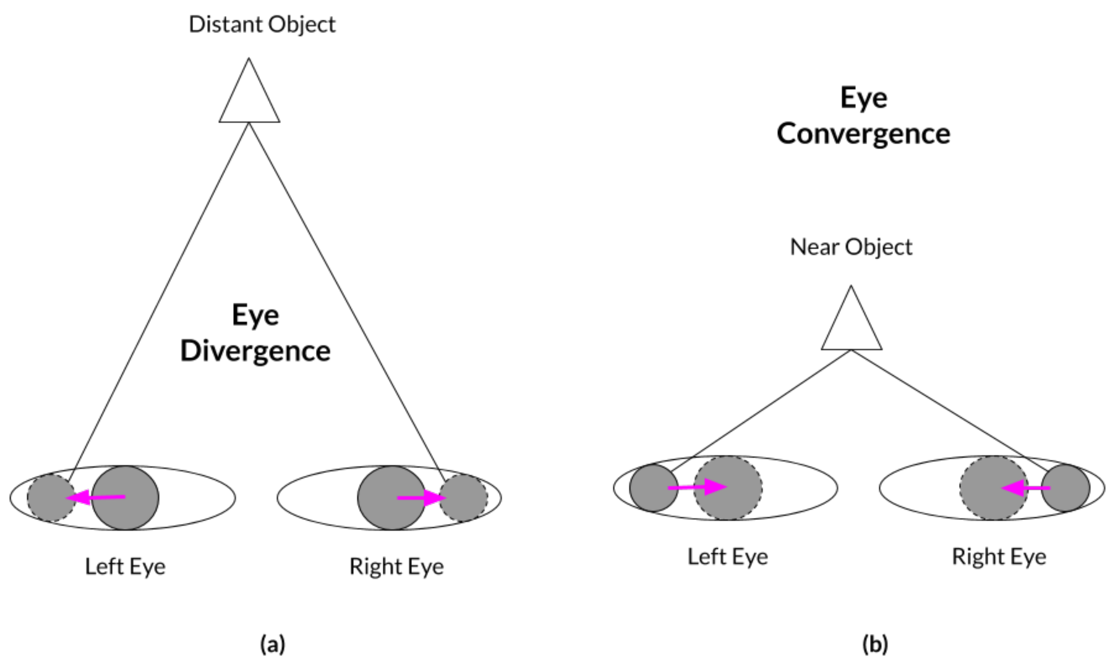

Fig. 1: Vergence eye movements enable stereovision of objects

Fig. 1 depicts examples of eye vergence, which is the movement of eyes to enable 3D perception. Fig. 1(a) depicts eye divergence, which is the movement of the eyes away from each other to focus on a distant object. Fig. 1(b) depicts eye convergence, which is the movement of the eyes towards each other to focus on a near object.

Once the user's gaze falls upon the object of interest, fixation is the phase whereby the user maintains their visual gaze on a single location. This enables the user to accurately perceive the object of interest. During the phase of visual fixation, image from the object of interest falls directly on the fovea.

Other eye movements that aid visual perception include saccades, which are rapid and simultaneous movement of the eyes between two or more phases of visual fixation. Human vision includes various saccade patterns, depending on the environment and user intent.

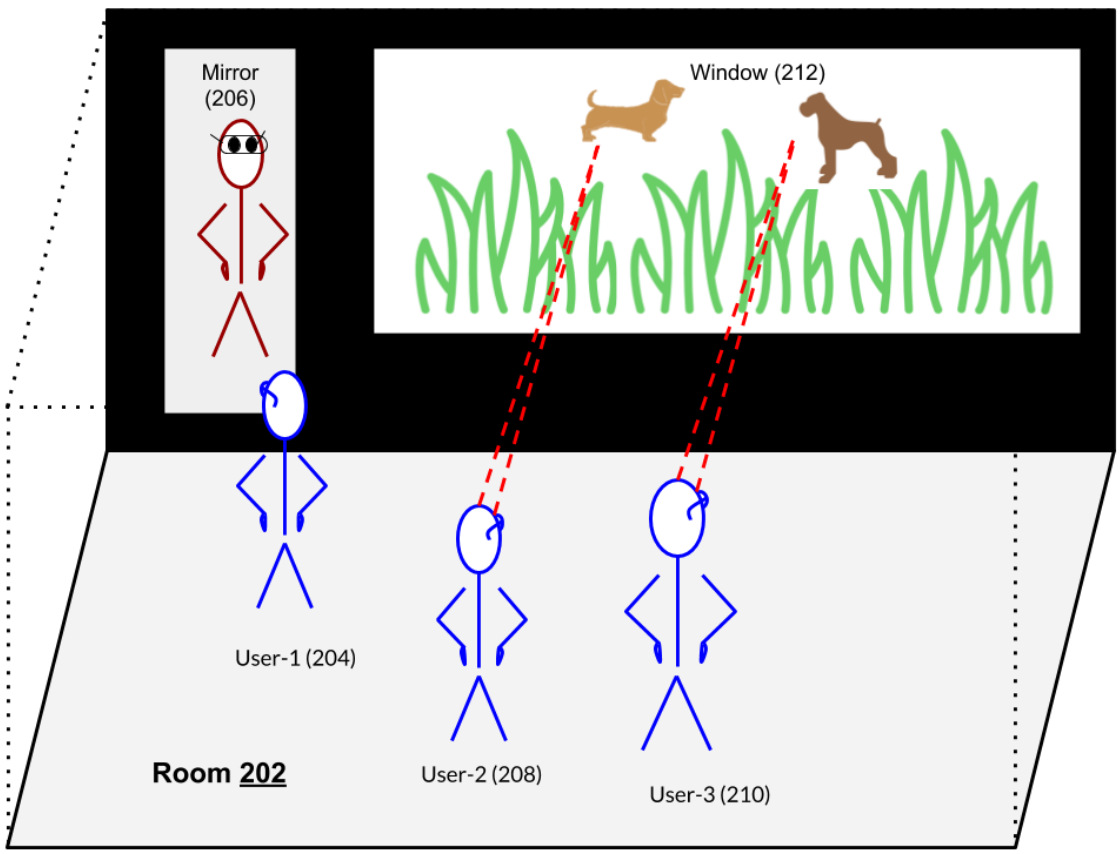

Fig. 2: Vergence and gaze based 3D mapping of a local environment

Fig. 2 illustrates an example of utilizing natural perception of users in their environment to obtain information about the user environment. In this illustrative example, users (204, 208, and 210) wearing AR/VR headsets are depicted inside a room (202). Per techniques of this disclosure, the headsets include sensors, e.g. cameras, that can track eye positions and eye movements of the user(s).

As each user scans the room, with user permission, their respective eye movements are tracked by one or more sensors located in their headsets. Identifying the volumetric space associated with the eye movements provides additional signals about the type of environment, e.g. surfaces, depth, and/or distance of different objects, location of other users, etc.

Information from eye focus depth enables a system to extend the mapping of the local environment (sensed space) beyond a range provided by 3D sensors located on the user's head mounted displays. Further, the mapping of the environment can be performed with less frequent operation of the 3D sensors, by supplementing the mapping with signals based on eye focus depth. If information about the scene, e.g., flat images, is available, 3D depth in the scene can be imputed based on eye vergence with solid objects. Less frequent operation of the 3D sensors can provide advantages such as lower power consumption, less heat generation leading to increased user comfort, etc.

For example, windows in a room can be identified based on eye vergence information indicative of an eye focus depth that is greater than the dimensions of the room, since users do not look through solid surfaces but do look through windows. In the example illustrated in Fig. 2, eye focus depth information obtained from the headsets of user-2 (208) and user-3 (210) indicate that a window (212) is located in the direction of their gaze, since their eye focus depth (which corresponds to the dogs seen outside the window) is greater than the distance between the users and the wall of the room.

Eye vergence information can be used in combination with previously generated maps and/or sensor information for greater accuracy. For example, if a map or sensor indicates that a portion of a user environment is a wall, but the eye vergence information indicates that users have looked through that portion, that portion of the wall can be interpreted to be a window, without any additional information from device sensors.

In a similar manner, a mirror (206) can be identified based on the eye vergence information obtained from user-1 (204). Specific saccade patterns associated with person recognition that are obtained from the headsets of the users and corresponding vergence depth information can be utilized to estimate distance(s) between the users in the room. This can serve as a substitute for and/or complement simultaneous localization and mapping (SLAM) or positional tracking techniques.

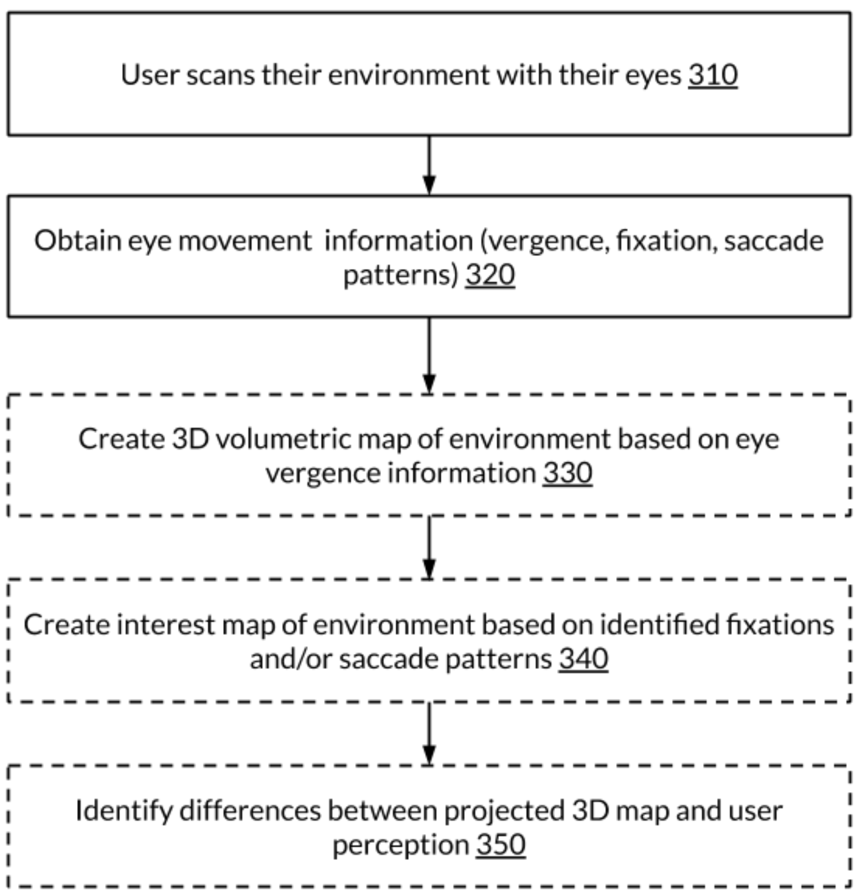

Fig. 3: Vergence and gaze based 3D mapping

Fig. 3 depicts an example method for the use of vergence and gaze point estimation to create a 3D map of a local environment, per techniques of this disclosure. In this example, 3d mapping is initiated when a user wearing the wearable device scans their environment with their eyes (310). Typically, as the user scans their environment, their eyes focus on different objects in the environment. With user permission, eye movement information such as vergence, fixation/gaze points, saccade patterns, etc. are obtained (320) from one or more sensors on-board the wearable device.

Based on the obtained eye vergence information, a (coarse) 3D volumetric map of the environment is created (330). The eye vergence information can be combined with information about the user environment obtained from other sources, e.g., inertial sensors, SLAM sensors, wireless signal levels, GPS location information, etc.

Based on identified user fixations - objects that a user looked at for which their gaze time exceeded a certain threshold time - an interest map of the room can be generated (340) which maps objects of interest in an environment. Fixation information from multiple users can be combined for greater accuracy of the interest map.

In some virtual reality (VR) systems, vergence information is used to identify differences (350) between a projected 3D map (based on design) and actual user perception (based on eye-tracking of users within the virtual environment), which can be indicative of errors in scene design, rendering, etc. The identified differences can be provided as feedback for adjustment of the projected 3D map. Depending on the application, one or more of the generation of volumetric map (330), interest map (340), and difference identification (350) can be performed as necessary, based on eye tracking.

Techniques of this disclosure can be utilized to obtain 3D spatial information for environments where an outward facing camera is not permitted/desired, since the mapping can be performed based on vergence information detected using sensors that track a user’s eyes. Human intent and situational awareness can thus be obtained and utilized for AR/VR applications, with minimal impact on security or privacy. By leveraging the user’s natural perceptive reactions to augment machine perception using sensors, power savings from reduced sensor use can be realized while still obtaining information about the 3D world around a user.

CONCLUSION

This disclosure describes techniques for utilizing a user’s natural perception of their environment to obtain information about the local environment. With user permission, eye positions and eye movements of a user as they interact with their environment are utilized to substitute for and/or augment machine-based perception of the environment. Eye movement information such as vergence, fixation/gaze points, saccade patterns, etc. are obtained from eye tracking sensors. A 3D volumetric map of the environment is created based on the eye movement information. Based on identified user fixations, an interest map of the room is generated. 3D spatial information can be obtained in user environments where an outward facing camera is infeasible, since mapping can be performed based only on vergence information obtained from the eyes.