Presentación:

Buenos días estimados miembros del tribunal, soy Juan José Arias y voy a presentarles mi trabajo de fin de grado, un sistema de seguimiento de manos para escenas en realidad extendida.

Primero, se dará una breve introducción al respecto, dando contexto sobre la realidad virtual, técnicas de inmersión en la escena y los objetivos de este proyecto. Posteriormente, se hablará de las tecnologías utilizadas, su implementación en el proyecto y los experimentos realizados para probar los resultados. Finalmente se hablará del desarrollo del proyecto y de las conclusiones.

Introducción:

La realidad virtual es una tecnología que lleva en desarrollo desde hace ya bastantes años, aunque este tomo fuerza con el lanzamiento de las Oculus en 2014, sus orígenes se remontan años atrás, y desde el principio, la idea principal era la inmersión de los usuarios dentro de las escenas, hacerles sentir que están dentro de estas.

Para lograr esa sensación de inmersión, se utilizaron varios métodos, pero el más sencillo y común, incluso hoy en día, es el uso de controladores. Los controladores permiten la interacción dentro del entorno virtual, pero al requerir un dispositivo externo como es el propio mando y que no resulta tan natural como usar las propias manos hacen que la sensación de inmersión no sea tanta como se deseaba. Cuando se empezó a desarrollar el seguimiento de manos, se utilizaban guantes especiales con sensores, dichos guantes, daban la sensación de que se deseaba, pero la necesidad de un dispositivo externo y el alto coste que estos tenían era un problema, haciendo esta tecnología poco accesible para la mayoría. Hoy en día, ya existen sistemas de seguimiento de manos que no requieren de dispositivos externos, pero su desarrollo insuficiente, respecto a otros campos de la realidad virtual, y su uso limitado, limitan esa sensación de inmersión a ratos.

Por eso, el objetivo principal de este proyecto fue desarrollar una serie de herramientas que permitan a los usuarios poder interactuar con los distintos elementos dentro de la escena, mediante el uso de sus propias manos y desde la comodidad de un navegador, sin necesidad de descargar ninguna aplicación o servicio externo, aprovechando los sistemas de seguimiento de manos ya existentes y sin necesidad de dispositivos externos.

Tecnologías:

Para este proyecto se hizo uso de las siguientes tecnologías:

* Aframe: un framework basado en HTML y JavaScript diseñado para crear escenas VR y AR desde navegadores
* HTML: tecnología utilizada para crear las distintas escenas en el navegador
* JavaScript: La tecnología más importante ya que fue la que se utilizo para crear y programar todos los componentes que se crearon a lo largo de este proyecto
* Three.js: una biblioteca de JavaScript que permite la creación de y manipulación de geometrías 3D
* WebXR: Tecnología que permitió acceder a la información necesaria para el seguimiento de manos a través de si API
* WebGL: Tecnología que permite que las escenas creadas sean compatibles con los navegadores permitiendo visualizarlas y probarlas desde estos
* Visual Studio code: Entorno de desarrollo en el cual se realizo la totalidad de este proyecto
* GitHub: Plataforma que permitió ir almacenando el código al igual que permitió controlar las distintas versiones a medida que avanzaba el proyecto
* Meta Quest 3: Dispositivo en el cual se realizaron todas las pruebas y experimentos
* Latex: Tecnología realizada para la creación de la memoria

Implementación:

Como se mencionó anteriormente, para el seguimiento de manos se utilizo webxr ya que esta tecnología ya cuenta con una forma de seguimiento de manos integrada. Utilizar esta tecnología fue un reto, ya que únicamente tiene compatibilidad con los dispositivos Quest desde hace 2 años, con lo cual no hay mucha información o ejemplos. Pero después de explorar las distintas opciones que se podían utilizar con Aframe, esta era la mejor opción, ya que nos permitía acceder de forma individual a la información de cada una de las articulaciones que forman la mano. Como se puede apreciar en la imagen, webxr divide la mano en un total de 25 articulaciones, desde la punta de cada dedo hasta la muñeca incluida,

Toda esta información nos la proporciona la API de webXR. Para acceder a esta información y posteriormente procesarla se creo el componente manos, dicho componente es el encargado de tomar la información necesaria de la api para posteriormente renderizar las manos en la escena y detectar los gestos. A través de la api se puede acceder a la información de todas las articulaciones de manera individual y para asegurarnos que el movimiento sea fluido tomamos esa información a cada frame, es decir, 60 veces por segundo.

Para renderizar las manos dentro de la escena, lo primero era crear una serie de esferas que representarían las articulaciones, habiendo una por cada articulación y por cada mano. Posteriormente, en una función dentro del componente manos, se procedia a acceder a la información de la posición de la articulación en cuestión y a actualizar la posición de la esfera, dando como resultado la imagen de la izquierda donde se aprecian las manos.

Respecto a la detección de poses, el poder acceder a la posición de cada articulación de forma individual facilitaba bastante las cosas, ya que, con eso, a través de distancias entre las articulaciones podíamos saber que dedos se tocaban o incluso que dedos estaban estirados o doblados. Utilizando esta lógica, en las demos finales se integraron dos gestos principales, el gesto de pinch, que se usa para la mayoría de las interacciones y el gesto de point. El gesto de point es algo mas complejo, ya que se puede considerar que tiene dos estados, el estado normal, con el dedo índice y pulgar estirados, en este estado, se crea un puntero que nace desde la punta del dedo índice y sigue el vector generado por la punta del dedo y su nudillo. Y también está el estado de pistol, donde con los dedos se imita la acción de disparar una pistola, haciendo que se ejecute la acción de click. Estas poses son las que posteriormente permiten al usuario interactuar con los elementos de la escena.

Al final del proyecto, se llegaron a implementar 5 acciones diferentes, las acciones de: drag, slide, stretch, hoover y click. En dichas acciones se hace uso de los gestos ya mencionados. Todas las acciones salvo la de click, están controladas por un componente grabable, dicho componente a cada frame comprueba si se está realizando algún gesto y si hay colisión con algún elemento de la escena, posteriormente, dependiendo del gesto y de la mano que lo realice este emite un evento correspondiente a la acción. La acción drag como su nombre indica, permite arrastrar cualquier elemento dentro de la escena y rotarlo, la acción de slide permite lo mismo, pero únicamente el eje o plano que se especifique en el squema del componente al añadirlo a la escena, la acción de stretch requiere de las dos manos y permite agrandar o encoger los elementos y la acción de hoover permite saber de forma visual cuando se esta tocando un elemento y se puede interactuar con él. Estas 4 acciones requieren de grabable ya que requieren de contacto directo con el elemento en la escena a diferencia de la acción de click. Para esta acción, como ya se explico previamente al explicar el gesto de point, utiliza el puntero generado en el dedo y la acción de pistol para lanzar el evento de click.

Para mostrar posibles implementaciones de estos componentes se crearon 2 escenas de demo donde se emplean los componentes de forma más práctica. En la primera, la foto de la izquierda, se trata de un juego donde se utiliza la acción de click, y en el de la derecha, utiliza las acciones de drag y hoover para poder dibujar en la escena. Esta segunda escena será la que posteriormente se mostrara como demo.

Experimentos:

Para probar los resultados, 4 voluntarios ayudaron a probar la demo de componentes, en dicha demo hay una serie de cubos formando un semicírculo alrededor del usuario al entrar. Hay en total 6 cubos, 1 por cada acción implementada y uno extra que contiene los 5 componentes. Cada cubo tiene encima un panel en ingles con una breve explicación de como interactuar con el cubo correspondiente. Para realizar los experimentos se utilizaron las gafas quest 3 cedidas por la universidad en un cuarto amplio para que los usuarios pudieran caminar libremente. Yo preparaba el dispositivo con la escena para introducirles directamente a esta y se les dieron algunas instrucciones como hacer que se miren las manos y realicen los gestos para familiarizarse y también se les pidió que vieran a su alrededor para asegurarse que podían entender las instrucciones de cada panel.

Luego de dejar que exploren la escena, se les pidió que compartieran sus opiniones. Respecto a las manos decían que se sentían bastante naturales y que el movimiento era fluido. Cada gesto se representaba bastante bien, pero resaltaban la poca variedad de gestos implementados al haber únicamente 2. Con las interacciones con los objetos también estaban satisfechos ya que eran fáciles de usar y en general bastante intuitivos, pero resaltaron que se podrían implementar alguno más para que haya mayor variedad. En general las opiniones de los voluntarios fueron bastante positivas, pero sí que resaltaron algunos puntos que requerían más trabajo a futuro, también se plantearon las opciones de implementar la compatibilidad con los controladores para poder utilizar ambos en la misma escena.

Desarrollo:

Este proyecto comenzó en octubre de 2024, y su progreso sin incluir la memoria se puede dividir en 7 sprints. Cada uno de los sprints contaba con sus propios objetivos y resultados, dichos objetivos se definían siempre al final del sprint anterior. El llamado sprint 0 fue la toma de contacto con este proyecto, donde se domino el uso de A-Frame y se formalizo el proyecto. El sprint 1 fue el mas largo, ya que tuvo un periodo de exploración de las posibilidades para el seguimiento de manos antes de decantarme por webxr, y una vez tomada la decisión debido a la falta de ejemplos o información al respecto tomo tiempo hasta conseguir renderizar las manos. Durante los siguientes 3 sprints se añadieron el manejo de eventos, la implementación de gestos y de colisionadores respectivamente. En el 5 sprint se creo el componente grabable mencionado anteriormente y en el ultimo sprint se desarrollaron las acciones implementadas.

Conclusiones:

En conclusión, este proyecto se trata de una caja de herramientas que permite a los usuarios utilizar sus manos dentro del entorno virtual para interactuar con los distintos elementos. Aunque es un proyecto que aun requiere trabajo para perfeccionar su funcionamiento, el poder usar las manos abre un abanico de posibilidades casi infinitas. Y por la forma en la que se ha desarrollado el proyecto, los usuarios pueden implementar en sus escenas únicamente el componente de las manos y ya ellos trabajar con estas o implementar también los componentes de interfaz de usuario.

En la pagina web se encuentra un enlace a todas las demos creadas al igual que al código principal, la memoria y esta presentación. También hay una serie de videos donde se muestra el funcionamiento de las demos expuestas en esta presentación.