

Aplicación de un navegador 3D con una interfaz de usuario natural utilizando la plataforma .NET, SDK y Kinect de Microsoft

I. Peredo Valderrama¹, K. Anaya Rivera¹, R. Peredo Valderrama²

¹Universidad Politécnica de Querétaro, Carretera Estatal 420 S/N, El Rosario el Marqués Querétaro CP. 76240

² Escuela Superior de Cómputo, Instituto Politécnico Nacional, Unidad Profesional Adolfo López Mateos, Av. Juan de Dios Bátiz S/N Esq. Miguel Othón de Mendizábal, Col. Linda Vista, Ciudad de México, D.F., C.P. 07738, México.
{ivan.peredo;karina.anaya}@upq.mx; rperedo@ipn.mx

Resumen. Kinect está a la vanguardia de los sistemas conocidos como Interfaz de Usuario Natural (NUI), que permiten una interacción más natural entre un usuario y la computadora a través de los movimientos del cuerpo y la voz. Esta es la próxima generación de la interacción hombre-máquina (MMI), desarrollando nuevas aplicaciones para interactuar con las personas y la computadora de una manera más natural. Las aplicaciones que utilizan estas nuevas NUIs están relacionadas con temáticas tales como: negocios, educación, juegos de video, y aún más allá, amplían los horizontes de desarrollo de aplicaciones. En este trabajo se ofrece una propuesta de bajo costo para diseñar, desarrollar e implementar un navegador 3D con una NUI usando Kinect SDK, .NET y el motor de Unity (Unity Engine). Permitiendo aumentar la interacción hombre-máquina, esta propuesta pretende ofrecer una herramienta de apoyo con una innovadora NUI para el aprendizaje a través de prácticas virtuales reduciendo los costes y riesgos inherentes. Un navegador 3D con una NUI ofrece una simple y fácil manera de manipular objetos virtuales.

Palabras clave: Interface natural de usuario, Microsoft Kinect, navegador 3D, Plataforma SDK.

1. Introducción

En la actualidad estamos viviendo cambios muy rápidos con la tecnología en casi todos campos, el desarrollo de las Interfaces de Usuario para los diferentes tipos de aplicaciones no ha sido la excepción, lo que ha provocado su desarrollo a lo largo de los años. Steve Mann desarrollo a lo largo de los 70's y hasta los 90's diferentes tipos de Interfaces de Usuario alternativas a las tradicionales: línea de comando y por medio del ratón a través de una Interfaz de Usuario Gráfica (Graphical User Interface, GUI por sus siglas en inglés), este nuevo tipo de interfaces las denominó Interfaz de Usuario Natural (Natural User Interface, NUI por sus siglas en inglés). Las NUI son la siguiente generación de Interfaces de Usuario [1], las cuales nos permiten interactuar con los usuarios de formas impensables hace algunos años, usando diferentes tipos de entradas como: rastreo de movimiento, voz, etc., las NUI nos posibilitan nue-

vas maneras de interactuar con los diversos tipos de computadoras como: teléfono inteligente, PC, laptop, iPad, etc., a través de una interfaz natural que explote las habilidades que hemos adquirido a lo largo de nuestra vida.

Una de los éxitos más sonados de Nintendo en los últimos años ha sido la consola Wii [2], causando un boom en el segmento de los videojuegos, vendiendo hasta el momento la consola Wii cerca de 100 millones de unidades [3]. Una de las causas de su éxito se debió a su innovadora NUI manipulada a través de un sensor de movimiento denominado Wii remote, Nintendo desarrollo una NUI que permitía interactuar con la consola en un espacio tridimensional real de una manera radicalmente novedosa. Ante el enorme éxito de Nintendo y su consola Wii, la competencia implementó NUIs similares a las de Nintendo, la consola Playstation 3 desarrollo el sensor de movimiento denominado Move [4], y la consola Xbox 360 desarrollo el sensor de movimiento libre denominando Kinect [5]. Esto ha renovado el interés de desarrollo de NUIs avanzadas e interactivas para los usuarios para los diferentes tipos de aplicaciones.

El éxito del Kinect colocando más de 10 millones de unidades vendidas en sus primeras semanas de venta, lo convirtieron en el accesorio de mayor venta en la historia de las consolas de videojuegos, conformando una amplia base instalada, posibilitando que fuera visto por muchos desarrolladores como una nueva herramienta para desarrollar innovadoras NUIs debido a su bajo costo y a la liberación del Kit de Desarrollo de Software (Software Development Kit, SDK por sus siglas en inglés) [6], posibilitándonos crear NUIs innovadoras que respondan a los movimientos y a la voz humana, posibilitándonos comunicarnos de una manera más natural con las computadoras, de una manera muy similar a como nos comunicamos las personas. Debido a los motivos anteriormente mencionados y al gran soporte que nos facilita la empresa Microsoft para el Kinect [6-7], fue que se eligió el dispositivo para el desarrollo de la presente propuesta. También se utilizó en la presente propuesta el motor de renderizado de la empresa Unity Technologies [8], debido al conjunto integrado de herramientas intuitivas de desarrollo, rápidos flujos de trabajo para desarrollar contenido interactivo 3D, su facilidad de publicación multiplataforma, sus activos (asssets) de calidad y su activa comunidad de desarrollo y compartición de conocimiento. Estas dos tecnologías son utilizadas ampliamente en la presente propuesta como herramientas de desarrollo bajo la plataforma .NET.

En el presente trabajo se presentan las diferentes etapas que se requirieron para el correcto desarrollo del Navegador 3D a través de una NUI innovadora utilizando al Kinect como dispositivo de captura de movimiento para la interacción usuario-computadora.

El software Unity nos sirvió para crear contenido 3D interactivo, el cual es programable mediante librerías en el lenguaje de programación C# utilizando el SDK Open Natural Interaction (OpenNI) [9].

Microsoft Research invirtió veinte años de desarrollo en la tecnología del Kinect, la Figura 1 muestra las partes que constituyen el Kinect. Fue anunciado por primera vez el 1 de junio de 2009 en la Electronic Entertainment Expo 2009 (E3 2009) como "Project Natal", creado por Alex Kipman, desarrollado por Microsoft para la consola de videojuegos Xbox 360, y lanzado en Norteamérica el 4 de noviembre de 2010,

desde junio del 2011 se lanzó su versión beta del SDK para PC para el Sistema Operativo Windows 7 [10].



Figura 1. Kinect de Microsoft.

El Kinect es un dispositivo alargado, diseñado para estar en una posición horizontal. El dispositivo contiene, una cámara RGB, sensores de profundidad, un micrófono multi-array y un motor de inclinación [11].

- Cámara RGB, es una cámara de video con una resolución de 640x480 pixeles a 30 Marcos Por Segundo (Frames Per Second, FPS por sus siglas en inglés).
- Sensores de profundidad, es una combinación de un proyector de profundidad (retícula izquierda) con un sensor de profundidad (retícula derecha), se calcula la distancia en función del tiempo que tarda en reflejar la luz.
- Micrófono Multi-array, es un conjunto de cuatro micrófonos que se monta como un solo micrófono.
- Motor de inclinación, nos permite ajustar el Kinect hacia arriba o hacia abajo hasta 27°.

Y aunque no visibles a simple vista, Kinect también posee:

- Memoria RAM de 512 Mb.
- Acelerómetro, para estabilizar la imagen cuando se mueve.
- Ventilador, no está encendido continuamente para no interferir con los micrófonos.

En conjunto nos permite capturar imágenes y movimientos de los cuerpos, además de ofrecer reconocimiento facial y aceptar comandos de voz.

La intención de utilizar el Kinect con respecto a las distintas aplicaciones computacionales tradicionales está enfocada en reemplazar la entrada de teclado y ratón estándar por gestos basados en umbrales angulares y la distancia entre las partes del cuerpo.

La idea de usar el cuerpo como medio de interacción con las aplicaciones computacionales dio pauta a la NUI, como sucesora de la GUI. La NUI se basa en el uso de habilidades del cuerpo, tales como movimientos y gesticulaciones, para ser utilizadas como medio de interacción con las computadoras, eliminando el uso de los dispositivos de entrada estándar. Es decir, el objetivo de las NUI es lograr que el usuario interactúe con la computadora de una más natural a como lo hacen las personas, sin que el usuario esté consciente de utilizar un dispositivo de rastreo de movimiento y voz. El desarrollo del Kinect ha constituido a la consola Xbox 360 de Microsoft como la primer NUI comercial [12].

2. Estado del arte

Las interfaces de usuario fueron creadas con el propósito de facilitar la interacción hombre-máquina. Así, la evolución de la computación ayudó también a las interfaces de usuario haciéndolas cada vez más poderosas, las cuales buscan ser cada vez más naturales y amigables, adaptándose a las necesidades del usuario. En la última década un nuevo tipo de interfaces de usuario, denominadas NUI, han comenzado a ganar terreno sobre el paradigma GUI, que ha sido el más exitoso y popular para las interfaces de usuario.

En la realización del presente trabajo se consultaron diversos artículos y empresas que están relacionadas con aplicaciones que utilizan NUI mediante el Kinect con la finalidad de eliminar el uso del ratón y teclado para llevar a cabo una forma de interacción más natural con la computadora. Algunos de los artículos y empresas analizadas, además de controlar una aplicación, se enfocan en el análisis de los datos obtenidos con Kinect, o su integración con otros dispositivos referentes al área de investigación.

2.1 Controlar una aplicación

KinEmote [13]: Controla el ratón con la mano. En este trabajo se muestra una aplicación para Windows que permite controlar el ratón con la mano. Está basada en OpenNI/NITE y el reconocimiento de gestos con la mano: juntar el dedo índice y pulgar para hacer clic izquierdo, cerrar la mano para arrastrar y la palma de la mano para deslizarse por las barras de desplazamiento de las aplicaciones.

Kinect-Education [14]: Kinect-Education, es una empresa ya dedicada al desarrollo de software de aprendizaje para niños implementando el uso del Kinect.

Análisis de datos y otros dispositivos: A Kinect-based system for physical rehabilitation. A pilot study for young adults with motor disabilities [15]: En este artículo se muestra una aplicación que ayuda a los pacientes con necesidad de realizar terapias

para recuperación, además de ser una aplicación muy llamativa da la posibilidad de realizar los ejercicios desde casa ofreciendo también una retroalimentación a quien los practican.

Skaneet [16]: Sirve para modelar en 3D, el entorno donde se encuentra el usuario con solo girar poco a poco el Kinect para qué escanee el lugar, también puede escanear el cuerpo de una persona y así tener la posibilidad de reconstruir e implementar un personaje en alguna aplicación como un videojuego con un avatar igual al usuario.

Como se puede ver en los distintos trabajos del estado del arte presentados anteriormente, todos utilizan una NUI, que hacen uso de los movimientos y gesticulaciones del cuerpo humano como datos de entrada, así el usuario puede emplear movimientos que le parecen familiares para crear una interacción más natural con la computadora, como si estuviera manejando objetos reales.

3. Metodología usada

El presente trabajo presenta una propuesta para la Manipulación de un Navegador 3D por medio del software Unity que sirve para modelar en 3D, mediante gestos y reconocimiento de voz utilizando el SDK OpenNI y Kinect de Microsoft.

La inicialización se realiza con la conexión del Kinect y el Unity, posteriormente se inicia una fase de funcionamiento que se detiene hasta que se indique el cierre del sistema. La fase comienza por obtener las imágenes de video, profundidad y el esqueleto del usuario.

Cámara de video: El Kinect adquiere imágenes de video con un sensor CMOS de colores a una frecuencia de 30 Hz, en colores RGB de 32-bits y resolución VGA de 640×480 pixeles. El canal de video monocromo CMOS es de 16-bit, resolución QVGA de 320×240 pixeles con hasta 65,536 niveles de sensibilidad. En un amplio campo visual con objetos, la cámara de video trata de reconocer a qué distancia están los distintos objetos, distinguiendo movimientos en tiempo real.

Sensores de profundidad: Para calcular distancias entre un cuerpo y el sensor, el sensor emite un haz láser infrarrojo que proyecta un patrón de puntos sobre los cuerpos cuya distancia se determina. Una cámara infrarroja capta este patrón y por hardware calcula la profundidad de cada punto. El rango de profundidad del sensor de Kinect está entre 0.4 y 4 metros. Existen 2 modos (Default y Near) para determinar distancias. Se ha elegido el modo “Default” ya que permite medir hasta 4 metros de distancia con respecto al sensor. Los sensores pueden llegar a distinguir la profundidad de cada objeto con una resolución de 1 centímetro y las estimaciones de la altura y anchura con una exactitud de aproximadamente 3 milímetros.

3.1 A partir del esqueleto obtenido se realiza la detección de acciones

La aplicación funciona a través del Kinect, detectando los movimientos de la persona que interactúe con el dispositivo y a su vez la aplicación transforma estos movimientos en instrucciones. La aplicación detecta la posición de las manos para poder mover o acercar un objeto. También es posible realizar un recorrido a través de un mapa tridimensional mediante la detección del movimiento corporal del usuario.

3.2 Interfaz gráfica

La interfaz gráfica de nuestro navegador 3D propuesto se muestra en la Figura 2. El diseño de la interfaz del usuario es minimalista para que el usuario pueda interactuar con los objetos 3D de una manera sencilla por medio de nuestra NUI utilizando el Kinect.

En la Figura 2, se puede ver un prototipo del menú de la interfaz gráfica del navegador 3D, las opciones de los objetos Corazón y Oreja tendrán una animación de carrousel, el usuario puede elegir la opción deseada cuando se posicione sobre ella con la pelota, que es quien sigue la posición de la mano.



Figura 2. Interfaz gráfica del Navegador 3D.

Cuando el usuario posicione la pelota sobre la opción “Corazón”, la aplicación nos redirecciará a la escena Corazón mostrada en la Figura 3. El cual es una vista de un corazón, donde el usuario podrá manipular el objeto con el Kinect, y las funcionalidades son: acercar, alejar y girar.

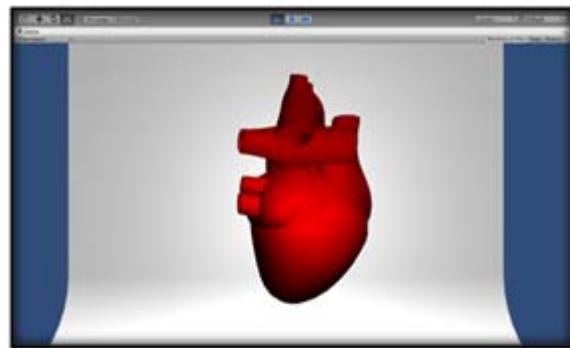


Figura 3. Vista de la escena: Corazón.

De la misma forma cuando el usuario posicione la pelota sobre la opción “Oreja”, la aplicación nos redireccionará a la escena Oreja, mostrada en la Figura 4. La cual es la vista de una oreja, donde el usuario podrá manipular el objeto con el Kinect, y las funcionalidades son: acercar, alejar y girar.



Figura 4. Vista de la escena: Oreja.

3.3 Patrón Modelo-Vista-Controlador

El patrón Modelo-Vista-Controlador (Model-View-Controller, MVC por sus siglas en inglés) es un patrón compuesto para desarrollar aplicaciones complejas, que fue utilizado en la arquitectura de nuestra propuesta. El patrón tiene tres bloques: Modelo, Vista y Controlador. El Modelo tiene los datos de la aplicación, y la lógica de ne-

gocios de nuestra aplicación, la Vista representa la GUI hacia el usuario y muestra el estado de la aplicación, y el Controlador maneja las entradas del usuario y cambia los estados de la aplicación. El punto fundamental del patrón es la alta flexibilidad del patrón claramente aplicable a la separación de los tres bloques, sin superponer sus responsabilidades, posibilitando a cada bloque llevarlas a cabo, pero estos bloques colaboran de forma vinculada, comunicándose entre ellos. La Figura 3 muestra los patrones: MVC, Observador y Singleton, del núcleo fundamental de nuestra arquitectura. El patrón Observador nos posibilita mantener actualizadas las Vistas con el Modelo. El patrón Singleton se centra en asegurar que sólo haya una instancia de una clase en la memoria, y sólo se tenga un acceso global al objeto, con el objetivo de optimizar el uso de los recursos, en nuestra propuesta el patrón Singleton optimiza el acceso a la capa del Modelo. Estos patrones ya han sido utilizados en el desarrollo de propuestas previas [17].

3.4 Rastreo de esqueleto y técnicas de definición algorítmica de los gestos

El Rastreo del Esqueleto (Skeleton Tracking, ST por sus siglas en inglés) lo basamos en un algoritmo de reconocimiento, el algoritmo tiene un entrenamiento basado en muchas imágenes, con la finalidad de lograr una buena precisión en la identificación de los esqueletos [18]. Nuestra propuesta implementa el ST en varias fases. La primera fase captura datos del mapa de profundidad, en la segunda fase se hace una clasificación de las partes del cuerpo con la finalidad de conseguir sus articulaciones, la tercera fase construye el esqueleto en función de las articulaciones.

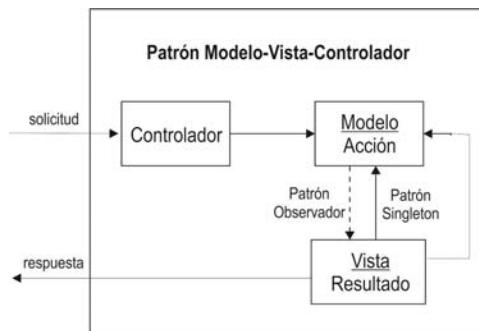


Figura 5. Patrón Modelo-Vista-Controlador.

En nuestra propuesta implementamos diferentes técnicas para definir algorítmicamente los gestos, contrastamos un conjunto de plantillas preestablecidas, comparando cada captura con las plantillas con cierto margen de error [18].

4. Conclusiones y trabajo futuro

Nuestra propuesta de una aplicación de un navegador 3D con una NUI innovadora, para manipular objetos 3D puede ser de gran utilidad en el área de la educación al estimular el aprendizaje por medio de simulaciones interactivas eliminando el factor miedo al momento de aplicar la metodología de prueba y error cuando se utiliza instrumental de alto riesgo o costoso. Este tipo de desarrollos posibilitan agregar herramientas innovadoras en los salones de clase, con la finalidad de estimular la investigación de los estudiantes, y mejorar su entendimiento de los temas tratados.

Las simulaciones computacionales hacen los conceptos más atrayentes y menos genéricos, posibilitando a los estudiantes una mejor compresión y desarrollar nuevas habilidades, mejorando un entendimiento realizado por medio de la tecnología. Las computadoras mejoran las actitudes e intereses de los estudiantes, al posibilitar un aprendizaje personalizable en un ambiente más interactivo por medio de la NUI propuesta y agradable.

Por lo tanto, nuestra propuesta del navegador 3D puede apoyar en la solución del problema de las prácticas inaccesibles para los alumnos o para aquellas personas que quieren manejar o experimentar con materiales muy caros antes de implementar físicamente. Esta aplicación tiene la ventaja de que las personas interesadas pueden trabajar con ella desde casa con solo contar con un Kinect sin necesidad de gastar en el material para realizar una práctica, así los estudiantes podrán tener un mayor control de responsabilidad sobre sus habilidades, sobre el ¿Qué quiero aprender hoy?, y así comenzar a realizar actos de proactividad.

La propuesta pretende fortalecer el área del conocimiento en los alumnos, aprendiendo teoría pero con la opción de aterrizarlo con la práctica.

La propuesta presentada puede llegar a ser una gran herramienta de apoyo para las instituciones y sus alumnos, fomentando el gusto por sus estudios a través de las prácticas virtuales realizadas con sus propios movimientos y a su propio ritmo, sin temor a equivocarse puesto que se trata de un ambiente virtual, donde no se corre el riesgo de causar una perdida fuerte, teniendo la opción de repetir los ejercicios un sinfín de veces, sin los costos generados por el uso de los materiales involucrados en las prácticas.

En el futuro cercano se crearan NUI en la mayoría de las aplicaciones computacionales cuyo objetivo será lograr que el usuario interactúe con la computadora, como si esta fuese otra persona y sin que esté consciente de utilizar un dispositivo.

Al realizar la investigación de esta propuesta se observó que el manejo de objetos 3D con el dispositivo Kinect, es un gran avance en la tecnología, que pueden implementar diferentes conceptos para: negocios, artes, educación, videos juegos, y más allá.

Como trabajo futuro se realizará un sistema que utilice una NUI para modelar o deformar objetos virtuales y experimentar con la interacción en un ambiente de realidad aumentada. En la investigación de esta propuesta, se mostró que hay muchas posibilidades en distintos campos de aplicación de utilizar el Kinect para NUIs de manera innovadora.

Agradecimientos

Los autores de este artículo agradecen a la Universidad Politécnica de Querétaro y al Instituto Politécnico Nacional (IPN) y a la Escuela superior de Cómputo (ESCOM) por su apoyo para este trabajo dentro del proyecto SIP: 20131560. Los autores desean reconocer a todos sus colegas y estudiantes que participaron en el diseño y desarrollo del software descritos en este artículo.

Referencias

1. NUI: Interfaz de Usuario Natural, Sitio Web, Disponible en URL: <http://www.microsoft.com/en-us/kinectforwindows/develop/>.
2. Wii Official Site at Nintendo, Sitio Web, Disponible en URL: <http://www.nintendo.com/wii>.
3. Global Weekly Video Game Charts, Games Sales, Top Sellers, Game Data - VGChartz, Disponible en URL: <http://www.vgchartz.com/>.
4. PlayStation®Move Motion Controller - PlayStation®3 Move Info, Games & Updates, Disponible en URL: <http://us.playstation.com/ps3/playstation-move/>.
5. Kinect - Xbox.com, Disponible en URL: <http://www.xbox.com/es-MX/Kinect>.
6. Kinect for Windows | Voice, Movement & Gesture Recognition Technology, Disponible en URL: <http://www.microsoft.com/en-us/kinectforwindows/>.
7. Microsoft, Sitio Web, Disponible en URL: <http://www.microsoft.com/es-mx/default.aspx>.
8. Unity Technologies, Sitio Web, Disponible en URL: <http://unity3d.com/>.
9. OpenNI | The standard framework for 3D sensing, Disponible en URL: <http://www.openni.org/>.
10. J. Webb, J. Ashley, Beginning Kinect Programming with the Microsoft Kinect SDK (Apress, 2012).
11. R. Miles, Start Here! Learn the Kinect API (Microsoft, 2012).
12. S. Peralta. Interfaz de lenguaje natural usando Kinect. Tesis de Maestría en Ciencias de la computación, CINVESTAV-IPN, México, 2012.
13. KinEmote, Sitio Web, Disponible en URL: <http://www.kinemote.net/>.
14. Kinect-Education, Sitio Web, Disponible en URL: <http://www.kinecteducation.com/>.
15. Y. Changa, S. Chenb, J. Huangc. A Kinect-based system for physical rehabilitation, a pilot study for young adults with motor disabilities. Elsevier, Vol. 32, No. 6, November–December 2011, pp. 2566–2570.
16. Skanect, Sitio Web, Disponible en URL: <http://skanect.manctl.com/>.
17. Rubén Peredo, Alejandro Canales, Alain Menchaca, Iván Peredo, Intelligent Web-based education system for adaptive learning, Expert Systems with Applications, 38(12): pp. 14690–14702, Pergamon Press, 2011.
18. Reto SDK Kinect: Reconocer gestos con Skeletal tracking – MSDN España, Sitio Web, Disponible en URL: <http://blogs.msdn.com/b/esmsdn/archive/2011/08/22/reto-sdk-kinect-reconocer-gestos-con-skeletal-tracking.aspx>.