

Documentación Kinect

Jennifer Cruz Muñoz José Daniel Ku Ek
Israel Navarrete Alvarado

2 de septiembre de 2012

Índice

1. Descripción del Dispositivo	4
1.1. Capacidades técnicas[2].	4
2. Descripción de las API's.	5
2.1. OpenNi	5
2.1.1. NITE.	5
2.2. Open Kinect	6
2.2.1. Libfreenect.	6
2.3. Kinect SDK	6
2.3.1. Tabla comparativa de versiones.	6
2.3.2. XNA.	6
2.3.3. Microsoft Speech Platform Runtime.	7
2.3.4. Kinect Speech Language Pack.	8
2.4. Comparativa entre APIS.	8
Glossary	10

Índice de cuadros

1.	Tabla comparativa de versiones SDK de Microsoft	7
2.	Comparativa de API	9

1. Descripción del Dispositivo

Kinect empezó como una propuesta de Microsoft en Junio del 2009, al principio fue conocido con el nombre de "Project Natal" y lanzado en noviembre del año siguiente ya con el nombre con el que se le conoce actualmente. El dispositivo provee una Natural User Interface (NUI) que hace posible la interacción intuitiva sin ningún dispositivo intermedio, como había sido antes el control remoto. El dispositivo Kinect hace posible que una computadora utilice señales visuales y auditivas de su entorno, lo logra usando una cámara, un micrófono direccional y un sensor de profundidad. Este dispositivo mejora el modo de interacción humano - maquina, permitiendo que aplicaciones reaccionen gestos o movimientos corporales realizados por el usuario.

1.1. Capacidades técnicas[2].

Barra de sensores[2].:

- Lentes sensibles al color y la profundidad.
- Arreglo de micrófonos.
- Motor para ajustar la inclinación del sensor.
- Compatible con Windows y consolas Xbox existentes.

Campo de visión.

- Campo de visión horizontal: 57 grados.
- Campo de visión vertical: 43 grados.
- Rango de inclinación física: +/- 27 grados.
- Rango de sensor de profundidad: 1.2 a 3.5 metros.

Flujo de datos.

- 320x240 profundidad de 16 bits, 30 cuadros por segundo.
- 640x480 color de 32 bits, 30 cuadros por segundo.
- 16 bits de audio, 16 KHz.

Sistema de rastreo del esqueleto.

- Rastrea hasta 6 personas, con solo 2 jugadores activos.
- Rastrea hasta 20 articulaciones por jugador activo.

- Posee la habilidad de reconocer jugadores activos con sus avatares de LIVE.

Sistema de audio.

- Reconocimiento de voz.
- Sistema de cancelación de audio que mejora la recepción de voz.
- Party chat y chat de voz.

2. Descripción de las API's.

En esta sección se dará una breve explicación sobre las diferentes API's que pueden ser usadas para programar aplicaciones que usen los recursos proporcionados por Kinect. Una API es un grupo de rutinas (conformando una interfaz) que provee un sistema operativo, una aplicación o una biblioteca, que definen cómo invocar desde un programa un servicio que éstos prestan. En otras palabras, una API representa un interfaz de comunicación entre componentes software. Una compañía de software libera su API a disposición del público para que otros desarrolladores de software pueden diseñar productos que funcionan con su servicio. Un API a menudo forma parte de SDK (Kit de desarrollo de software)[1].

2.1. OpenNi

Es un **framework**, que funciona con varios lenguajes de programación , **multiplataforma** que define un API para desarrollo de aplicaciones, está compuesta por un grupo de interfaces que permiten el acceso a los sensores de audio y visión, OpenNi es un software Open Source desarrollado por PrimeSense, empresa israelí que desarrolló la tecnología del Kinect.

2.1.1. NITE.

NITE es un **middleware** para el procesamiento de los datos de Kinect que provee las siguientes características:

- Análisis de cuerpo completo.
- Análisis de mano
- Detección de gestos
- Análisis de escena

2.2. Open Kinect

También conocida como Free Kinect es un librería Open Source cuya función es acceder a la cámara usb Kinect, se comenzó a desarrollar en noviembre del 2011; esta librería, fue creada por una comunidad de programadores llamada Open Kinect [6], actualmente ofrece soporte para: cámara RGB y sensor de profundidad, motores de posicionamiento, acelerómetro,¹ led (indicador de estado).

2.2.1. Libfreenect.

Librería de bajo nivel que incluye todo el código necesario para activar, inicializar y establecer la comunicación de datos con el hardware kinect, incluye una API multiplataforma que funciona con Windows, Linux y OS X. El API tiene adaptaciones y extensiones para los siguientes lenguajes de programación: C, C++, .NET (C#/VB.NET), Java (JNA and JNI), Python, C Synchronous Interface,

2.3. Kinect SDK

Fue lanzado por Microsoft en Junio del 2011 para ser usado con Windows 7, el paquete de desarrollo SDK ofrece la posibilidad de explotar las capacidades del sensor Kinect dentro del entorno de trabajo de Windows. Incluye los **drivers** para el funcionamiento del dispositivo Kinect, estos pueden usarse para la programación de aplicaciones con C++, C# o Visual Basic, usando Microsoft Visual Studio 2010. Al descargar el SDK de la pagina oficial se incluyen ejemplos de códigos en los tres lenguajes mencionados.

2.3.1. Tabla comparativa de versiones.

La siguiente **1** tabla fue hecha con base en la información de cada versión del SDK de Microsoft.

2.3.2. XNA.

XNA es un conjunto de tecnologías para vídeo juegos de Microsoft, incluye los siguientes complementos:

- DirectX. Se trata de una librería de C++ para programar gráficos.
- XNA Game Studio. Son complementos de Visual Studio para la canalización del contenido y el despliegue de funciones usadas con el frame-

¹• Utiliza el acelerómetro para obtener los grados de inclinación del dispositivo.

Característica	Beta 1	Beta 2	Versión 1	Versión 1.5
Robust Skeletal Tracking	•	•	•	•
API de Reconocimiento de voz	•	•	•	•
Cancelación de eco	•	•	•	•
Camara XYZ (depth)	•	•	•	•
Windows 7	•	•	•	•
Windows 8		•	•	•
Cámara a colores		•	•	•
Acceso a datos en bruto del sensor depth		•	•	•
Acceso a datos en bruto de la cámara		•	•	•
Acceso a datos en bruto del micrófono		•	•	•
Multithreading al usar Skeletal Tracking		•	•	•
Detección y gestión del estado de Kinect		•	•	•
Construcción de apps de 64 bits		•	•	•
Supresión de ruido de fondo		•	•	•
Near Mode			•	•
Soporte de 4 sensores kinect al mismo PC			•	•
Control de detección de usuario en S.T.			•	•
Microsoft Speech V.11 incluido			•	•
Acustica mejorada con far-talk			•	•
Soporte para idioma Español				•
Kinect Studio				•
Human Interface Guidelines (HIG)				•
Face Tracking SDK				•
Seat Skeletal Tracking				•
Skeletal Tracking Near Mode				•

Cuadro 1: Tabla comparativa de versiones SDK de Microsoft

work. Esta es la herramienta usada al hacer un juego para Xbox y enviar los datos a este.

- Framework de XNA. Es el conjunto de bibliotecas .NET construidas.

2.3.3. Microsoft Speech Platform Runtime.

Microsoft Speech Platform SDK v11 ofrece una nueva funcionalidad en las herramientas de desarrollo de Microsoft para manejo de gramática, que permite entre otras cosas validar, depurar, probar y optimizar el uso de elementos gramaticales de las aplicaciones de voz. Algunas de las características descritas anteriormente se implementan en las siguientes herramientas .

- **CheckPhrase.exe.** Se determina si una gramática usa una frase dada y regresa información sobre la frase recibida.

- **Confusability.exe.** Esta nueva herramienta identifica frases en una gramática que son fonéticamente similares. La herramienta puede ayudar a detectar las frases que a la larga pueden causar a los usuarios tener una mala experiencia, de tal manera que una frase en la gramática es falsamente reconocida como otra frase que esta también incluida en la gramática. La herramienta Confusability acepta varios archivos de gramática como entrada donde realiza su análisis en el conjunto de archivos de entrada especificados.
- **GrammarValidator.exe** Valida la sintaxis del elemento gramatical. Emite una serie de advertencias al validar una gramática, que incluye la detección de la repetición y las reglas inalcanzables.
- **PhraseGenerator.exe** Genera un subconjunto de frases apoyados en las ponderaciones de una gramática, y si lo desea, puede optar por no ampliar las referencias específicas de la regla.
- **Simulator.exe** Se puede elegir si se desea o no volver a utilizar el estado del motor en la estructura de la **EMMA** (Extensible MultiModal Annotation) de entrada, en lugar de una opción en la línea de comandos. Es decir, el estado motor de reconocimiento es ahora automáticamente reutilizado sólo cuando las expresiones están contenidas dentro de un EMMA en un bloque de secuencia.
- **SimulatorResultsAnalyzer.exe** Proporciona información adicional para el ajuste de sus gramáticas, incluyendo detección de frases que no se encuentran en la gramática ya definida, conocidas como fuera de gramática, y nuevos tipos de categorías de errores para darle un mejor análisis y métricas para el reconocimiento simulado. Añade nuevos elementos XML (AudioType, RecoResultSemantics y TranscriptSemantics) a la salida para hacer más fácil analizar los resultados sin necesidad de referirse de nuevo al archivo de salida del simulador.

2.3.4. Kinect Speech Language Pack.

Los paquetes de lenguaje de Kinect para Windows son complementos para el Kinect en tiempo de ejecución que usan el reconocimiento de voz para habilitar la forma en que cierto idioma se habla en diferentes países.

2.4. Comparativa entre APIS.

Como ya se ha mencionado existe mas de una API para para desarrollar aplicaciones que hagan uso del kinect, se elaborado la siguiente tabla 2 para compararlas y facilitar elegir una API

API	OpenNI	OpenKinect	Kinect SDK
Multiplataforma	si	si	no
Multilenguaje	si	si	si
Procesamiento de flujo de datos	si	no	si
Licencia	Open source y propietario	Open source	Propietario
Comentarios	Usa NITE para procesar los datos del kinect	No soporta skeleton tracking.	Solo funciona con Windows 7 o superior.

Cuadro 2: Comparativa de API

Todas estas API's permiten acceso a bajo nivel, esto es manejo del flujo de datos (audio, video etc) que proporciona kinect pero no todos proporcionan procesamiento del flujo de datos (rastreo de esqueleto, mano o rostro, etc). Se ha elegido Kinect SDK sobre el resto porque esta en continuo desarrollo y proporciona documentación mas amplia que el resto

Glossary

driver Es un traductor que permite al sistema operativo comunicarse con un dispositivo de hardware y viceversa [8] . 6

EMMA Herramienta usada para representar información extraída automáticamente de la entrada de un usuario por un componente interprete, dicha entrada es tomada en cualquier modalidad soportada por la plataforma. Los componentes que niegan el marcado de EMMA son: reconocimiento de habla, reconocimiento de escritura, ingenierías de entendimiento de lenguaje natural, otros medios interpretes de entrada (como teclado o joystick) y componentes de integración multimodal. [9] . 8

framework Un framework es un conjunto de bibliotecas, herramientas y normas a seguir que ayudan a desarrollar aplicaciones [3] . 5

middleware Es un modulo intermedio que actúa como conductor entre dos módulos de software [7] . 5

multiplataforma Perteneciente a entornos informáticos heterogéneos. Por ejemplo, una aplicación de plataforma cruzada es uno que tiene un único código base para múltiples sistemas operativos. . 5

Referencias

- [1] ALEGSA. Diccionario de informática. <http://www.alegsa.com.ar/Dic/api.php>.
- [2] Rakuten group. Pay.com. <http://www.play.com/Games/Xbox360/4-/10296372/Kinect/Product.html>, 2012.
- [3] Jérôme Lafosse. *Expert IT Struts 2 - El framework de desarrollo de aplicaciones Java EE*. Ediciones ENI, 2010.
- [4] Microsoft. Msdn. <http://msdn.microsoft.com/en-us/library/hh362855>, 2012.
- [5] Sun Microsystems. Desing guidelines. <http://java.sun.com/products/jlf/ed1/dg/higq.htm>, 1999.
- [6] OpenKinect. Openkinect wiki. http://openkinect.org/wiki/Main_Page, 2012.
- [7] M. Wheeldin. *Cuerpo de Gestión Administrativa*. MAD-Eduforma, 2005.
- [8] M. Wheeldin. *The "No Assumptions" Guide to Windows XP*. Chestnut Cove Solutions, 2005.
- [9] Roberto Pieraccini Dave Raggett Wu Chou(Avaya) Deborah A. Dahl, Michael Johnston. Emma: Extensible multimodal annotation markup language. <http://www.w3.org/TR/2004/WD-emma-20041214/>, 2004.