



CÁMARAS DE PROFUNDIDAD

Tecnologías multimedia e interacción

Mario Alonso Núñez
Universidad Complutense de Madrid

Tabla de contenido

1. ¿Qué son las cámaras de profundidad?	4
2. El sensor Kinect	5
2.1 Componentes físicos del sensor Kinect	5
2.2 Funcionamiento software del sensor Kinect	6
2.3 Como desarrollar con Kinect	8
2.4 Aplicaciones de Kinect	8
3. La tecnología Intel RealSense	10
3.1 Productos que implementan la tecnología RealSense	10
3.1.1 Cámaras de profundidad D455 y D515	10
3.1.2 Cámara lidar L515	11
3.1.3 Cámara de reconocimiento facial F455	12
3.2 Cómo desarrollar con RealSense	12
4. Bibliografía	13
Video tutorial cómo instalar SDK kinect:	13

1. ¿Qué son las cámaras de profundidad?

Las cámaras de profundidad son un tipo de cámaras especializadas en medir la distancia a la que se encuentran los objetos captados por estas. Para llevarlo a cabo, la cámara necesita que estos se encuentren en frente de ella, sin embargo, una vez calculada se podrá obtener la distancia entre los mismos mediante triangulación.

Este tipo de cámaras comenzaron a desarrollarse en la primera década del presente siglo y han demostrado ser una tecnología con un gran potencial, esto es debido a la diversidad de aplicaciones que son capaces de llevar a cabo, siendo las más comunes aquellas que tienen que ver con la identificación de objetos tridimensionales en el ambiente.

Algunas de las empresas más importantes dentro del sector informático ya han desarrollado sus propias cámaras de profundidad. Los casos más comunes pueden ser *Microsoft* con su sensor óptico *Kinect* e *Intel* con su amplia gama de cámaras y detectores que implementan la tecnología *RealSense*.



Una de las principales características que tienen actualmente este tipo de cámaras es su predisposición y fácil prestación al ámbito de desarrollo público, adoptando facilidades como la publicación de bibliotecas SDK de desarrollo, la versatilidad de sus funciones o su poco mantenimiento.

Pese a que ya existen cámaras pensadas para el cálculo de distancias, actualmente se trata de una tecnología que también se está implementando dentro de los dispositivos móviles bajo las siglas ToF (Time of Flight).

El funcionamiento general de esta tecnología se basa en la emisión de un haz de rayos infrarrojos en una gran cantidad de direcciones y el cálculo del tiempo que tardan estos en rebotar y volver de nuevo a la cámara, siendo captados por un sensor específico.

2. El sensor Kinect

El sensor *Kinect* es un dispositivo concebido inicialmente para la industria del entretenimiento, cuyo principal objetivo era modificar drásticamente la forma de interacción del usuario pasando de tener que sujetar un mando a emplear gestos con su propio cuerpo a modo de comando.

Kinect fue presentado por primera vez en el evento *Electronic Entertainment Expo 2009* de mano de la propia *Microsoft* y bajo la responsabilidad del jefe de proyecto Alex Kipman. El objetivo de *Microsoft* era brindar una nueva dimensión de jugabilidad al usuario de su consola *Xbox 360*.

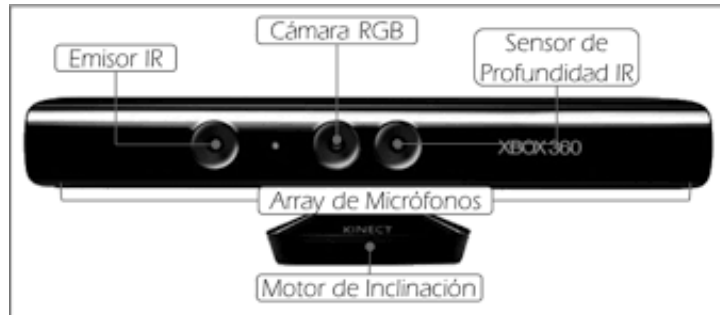


Para el correcto funcionamiento del sensor *Kinect* es necesario implementar un sistema de detección de profundidad, mediante el cual sea posible diferenciar la distancia entre los distintos elementos captados por la cámara. Llevar a cabo la detección de profundidad necesitará tanto un hardware específico como un software dedicado a dicha acción.

2.1 Componentes físicos del sensor Kinect

1. El sensor Kinect consta de los siguientes componentes físicos importantes para su funcionamiento:
 - **Micrófonos:** Gracias a una batería de diferentes micrófonos muy bien situados, el sensor es capaz de identificar la dirección de la cual llega el sonido y operar con el mismo.
 - **Cámara de vídeo:** Constituye una de las principales fuentes de recepción de información externa del sensor *Kinect*. Se trata de una cámara RGB que es capaz de capturar las distintas imágenes de vídeo, las cuales serán procesadas posteriormente para llevar a cabo las diferentes tareas de identificación de elementos.
 - **Emisor IR:** Emisor de infrarrojos capaz de proyectar una gran cantidad de luz con el objetivo de que incidan en todas las partes del escenario donde se encuentra el *Kinect*. A priori puede parecer un elemento secundario, pero se trata de un componente esencial para llevar a cabo las tareas de identificación de patrones en las imágenes obtenidas mediante la cámara.
 - **Cámara de profundidad:** Cámara con el objetivo de recoger todos los rayos de luz emitidos previamente por el sensor IR, pese a que no captura vídeo, se trata del sensor más importante para la recepción de información. Mediante la información obtenida de los rayos infrarrojos y de las imágenes captadas por la cámara principal, es capaz de medir las distancias a los distintos elementos y construir un mapa tridimensional del entorno.

- **Motores de inclinación:** Pequeños servomotores cuyo único objetivo es poder variar el ángulo de inclinación de la cámara y lograr así un mejor ajuste sobre el elemento que el sensor *kinect* necesita detectar. La funcionalidad inicial se basaba en facilitar que los usuarios entraran dentro del ángulo de visión de la cámara con el fin de minimizar los errores producidos al captar sus movimientos.



Podemos ver como el sensor *Kinect* está compuesto por receptores muy similares a los que tiene un lidar, lo cual da una gran versatilidad a la hora de desarrollar proyectos con el mismo y a su vez una gran alternativa muy económica con respecto a los lidar de la competencia con prestaciones similares. ¿Seguro que aún no logramos ver el gran potencial de este pequeño dispositivo?

2.2 Funcionamiento software del sensor Kinect

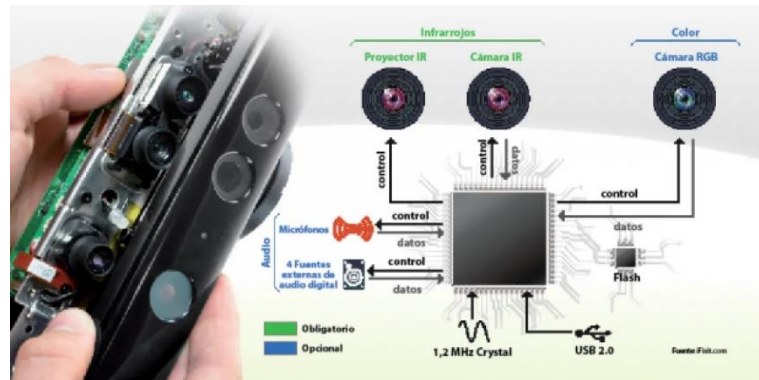
Un producto puede tener una gran batería de componentes hardware de calidad, sin embargo, esto no lo convertirá en un buen producto si no dispone de un software a la altura que pueda exprimir el potencial de dichos componentes.

En el caso del sensor *Kinect* hemos visto como disponemos de una gran cantidad de receptores, pero la clave no se encuentra en estos, sino en el gran potencial de su software, siendo capaz de realizar una gran cantidad de operaciones para el procesamiento de imágenes por segundo.

El proceso que lleva a cabo *Kinect* desde que recibe la información mediante sus sensores hasta que logra diferenciar las diferentes personas que se encuentran en frente al mismo es largo y difícil, pero podemos dividirlo en las siguientes partes:

1. El sensor *Kinect* capta la información en bruto sobre la que aplicará posteriormente el análisis correspondiente y obtendrá información útil para su objetivo. Esta información es recibida a través de los distintos sensores, aunque no todos ellos son imprescindibles para llevar a cabo dicho proceso:
2. A través del sensor IR y la cámara de profundidad el *Kinect* es capaz de detectar las distancias a las que se encuentran los diferentes elementos. Se trata de una información vital para el correcto funcionamiento de este y sin el cual no se puede llevar a cabo la detección de personas y sus gestos.
3. Mediante la cámara de vídeo RGB se recibe información en imágenes sobre todo aquello que se encuentra en frente del sensor *Kinect*. Esto no es vital, pero proporciona información extra sobre el entorno.
4. La batería de micrófonos proporciona información sobre dónde se sitúa la fuente de emisión de dichos sonidos, lo cual ayuda al sensor tanto a detectar a las personas como

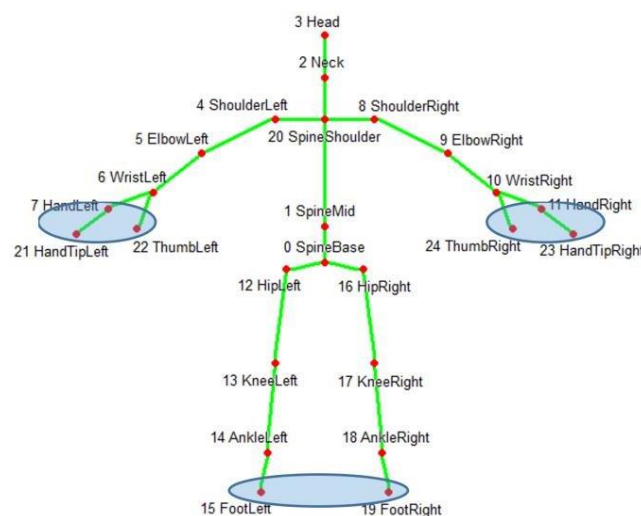
a saber diferenciar cual de ellas ha realizado un comando. Tampoco se trata de una información primordial, pero facilita la correcta interacción del usuario.



5. Tratamiento de la información obtenida en el paso anterior para la detección tanto de la profundidad como de las distancias entre las que se encuentran los elementos detectados entre sí. Este paso finaliza con la realización de un mapa en 3D de todo el ángulo de visión sobre el que se encuentra el sensor *Kinect*, el cual será muy importante para la identificación posterior de los cuerpos.
6. Por último, el sensor *Kinect* realiza un análisis conjunto de la información recibida en el paso 1 como del mapa tridimensional recibido en el paso 2 para la correcta detección de los cuerpos y la posición de estos. A través de esto, el sensor es capaz de diferenciar los comandos llevados a cabo por los usuarios y quien lo ha realizado en concreto.

La identificación de cada una de las diferentes personas que se encuentran delante del sensor *Kinect* se realiza mediante un algoritmo de inteligencia artificial, concretamente del conjunto de algoritmos de *Deep Learning*.

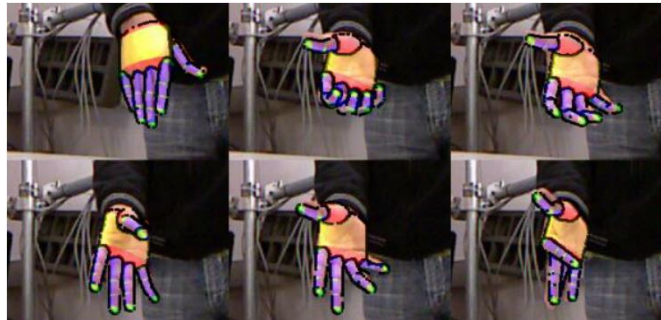
Este algoritmo se basa en identificar estructuras que pueden encajar en el prototipo de un cuerpo humano mediante la identificación de diferentes secciones o “huesos” que formarían el esqueleto de una persona estándar. En la siguiente imagen podemos ver un esquema de cada una de las secciones del cuerpo humano que son reconocidas por el *Kinect*.



Microsoft se dio cuenta del gran potencial de este dispositivo, lo que lo llevó a desarrollar y lanzar al mercado en 2014, una segunda versión, llamada *Kinect V2*. La diferencia principal entre ambas versiones, sin tener en cuenta su aspecto estético renovado, se trata exclusivamente de una

implementación software mucho más potente, de modo que ahora no se diferencian únicamente los cuerpos humanos y sus posiciones, sino también los gestos llevados a cabo con diferentes partes del cuerpo, como pueden ser las manos.

Para poder detectar las manos de los usuarios y saber leer los gestos que están realizando las mismas, se lleva a cabo la detección de cada uno de los cinco dedos de forma independiente. Una vez llevado esto a cabo, además de poder detectar las manos también se pueden detectar los gestos realizados con las mismas.



2.3 Como desarrollar con Kinect

Gracias a que *Microsoft* se percató del potencial de *Kinect* dentro del ámbito de desarrollo, decidió publicar la nueva versión del dispositivo compatible tanto para su nueva consola *Xbox One* como para los ordenadores que tuvieran instalados las versiones de *Windows 7* y *Windows 8*.

Posteriormente decidió liberar tanto las librerías SDK de desarrollo como hacer compatible el uso de ambas versiones con el sistema operativo *Windows 10*. De esta manera podremos tanto modificar los algoritmos que utiliza el sistema, como operar directamente con el hardware del mismo.

Para acceder a la página de descarga del SDK 2.0 de *Kinect* puedes hacer click [aquí](#). En el caso de que quieras ver un vídeo tutorial de instalación del SDK necesario para desarrollar con *Kinect*, haz click [aquí](#).

2.4 Aplicaciones de Kinect

El desarrollo de *Kinect* y la gran utilización de su tecnología por parte de la comunidad ha llevado a extender su uso a campos muy diversos, desde la tecnología, pasando por el entretenimiento, hasta llegar a la medicina. Vamos a resaltar su aplicación en alguno de estos campos:

- **Aplicaciones de rehabilitación:** El uso de *Kinect* ha sido desarrollado muy fructíferamente en el campo de la rehabilitación terapéutica. Esto es debido a que se puede controlar con facilidad la correcta realización de los ejercicios de rehabilitación llevados a cabo por un paciente para su correcta mejora en medicina.
- **Aplicaciones terapéuticas:** El uso del sensor unido a una pantalla para la realización de ejercicios corporales de baja intensidad ayudan a personas de edad avanzada a tener un mantenimiento adecuado del cuerpo sin la necesidad del empleo de máquinas de gimnasia.
- **Traducción del lenguaje de signos:** También se ha demostrado la efectividad del uso de *Kinect* en algunas áreas sociales, como puede ser la interpretación del lenguaje de signos.

- **Aplicaciones médicas para uso en quirófanos:** Se han desarrollado aplicaciones que basadas en la detección de gestos de *Kinect* para facilitar a los médicos trabajadores en quirófanos la interacción con el equipo mientras desarrollan su trabajo que requiere de alta concentración.
- **Aplicaciones para la producción de modelos 3D:** Gracias a la capacidad de *Kinect* para la detección de profundidad se han desarrollado aplicaciones que escanean modelos tridimensionales permitiendo así una impresión de alta calidad de los mismos.

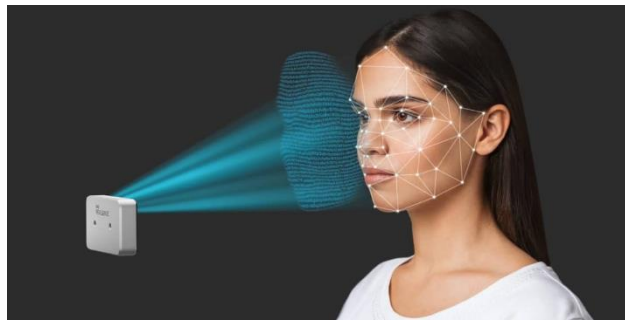
3. La tecnología Intel RealSense

A diferencia de lo que hemos podido ver con el sensor *Kinect*, *RealSense* no se trata de un dispositivo concreto, sino de una tecnología desarrollada por *Intel* cuyo principal objetivo es detectar la profundidad del entorno captado por las cámaras.

El funcionamiento de la tecnología *RealSense* no es muy diferente de la que hemos podido ver implementada en *Kinect*. Aunque podemos apreciar diferentes sensores y modos de ejecución dependiendo del producto concreto que estemos utilizando.

La base de la funcionalidad es común a todos los productos y se asienta en el cálculo de distancias mediante el uso de dos sensores de profundidad que miden el espacio en línea recta a todos los objetos que se encuentran en su ángulo de visión.

Posteriormente se puede calcular la distancia entre los dos elementos en la escena mediante triangulación, lo que permite llevar a cabo una gran cantidad de tareas que requieran la identificación de diferentes figuras, como pueden ser la utilización de gestos por parte del usuario o el reconocimiento facial.



3.1 Productos que implementan la tecnología RealSense

Esta tecnología se encuentra aplicada a una gran cantidad de productos que buscan adaptarse a las diversas necesidades de los usuarios. Podemos encontrar tanto cámaras de profundidad con un diverso rango de precios como lidars más orientados a un uso profesional concreto.

3.1.1 Cámaras de profundidad D455 y D515

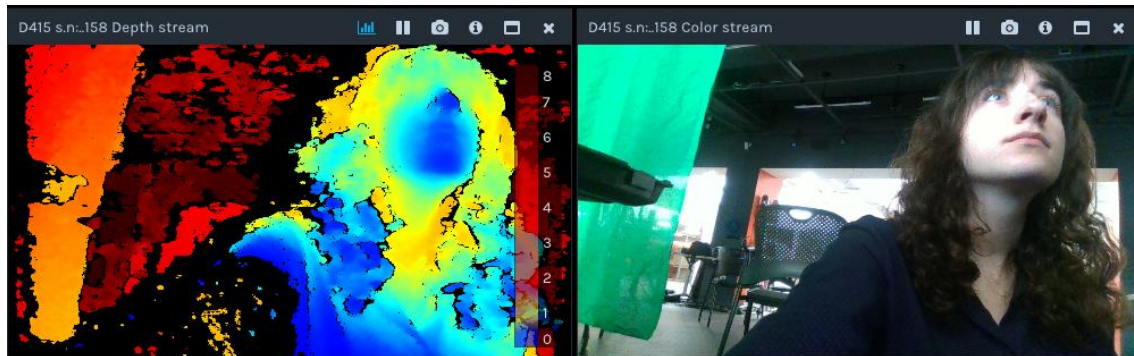
No es sorpresa que los desarrolladores de *Intel* son gente muy activa, lo que los lleva a desarrollar continuas actualizaciones de sus tecnologías, y *RealSense* no es una excepción. Actualmente podemos diferenciar diferentes categorías de los productos que implementan esta tecnología, pero los más utilizados e importantes son las cámaras de profundidad D455 (mejora de su predecesora D435) y D415.



Las principales diferencias entre ambos modelos son:

- El alcance de detección de estas, siendo de 0,6 a 6 metros en el modelo D455 y de 0,5 a 3 metros en el modelo D415.
- El tipo de visor utilizado en las cámaras, teniendo un obturador global el modelo D455 y un obturador de persiana el D415.

A parte de los dos sensores de profundidad necesarios para medir las distancias, también cuentan con una cámara RGB utilizada para dotar de color a los modelos tridimensionales construidos en base al entorno en el que se encuentran. También están dotadas de un emisor de infrarrojos que sirve para emitir rayos de luz con los que iluminar el entorno y mejorar los resultados a la hora de medir la profundidad.



Puede que el punto débil de estas cámaras sea la resolución de imagen captada por sus sensores, siendo apenas de 1280 x 720 píxeles. Sin embargo, esto es contrarrestado por una gran fluidez en la misma, pudiendo detectar hasta un máximo de 90 fotogramas por segundo.

Cabe resaltar que a diferencia de otros productos, las cámaras de la serie D están pensadas para su uso tanto en interiores como en exteriores y cuentan con una gran versatilidad frente a las variaciones del clima. Esto unido a un módico precio de desarrollo (desde 200 a 400 euros) las convierte en una opción competitiva dentro del mercado profesional.

3.1.2 Cámara lidar L515

En vista al éxito recaudado con las primeras cámaras de profundidad, *Intel* ha implementado la tecnología *RealSense* en su cámara L515, la cual se trata de un lidar enfocado a un uso profesional y de desarrollo. Se trata de un producto compacto y de gran potencia, pero con un precio superior a las cámaras de profundidad (aproximadamente 600 euros).

Al contrario que sus hermanas anteriormente vistas, este lidar está pensado únicamente para su uso en interiores debido a su menor versatilidad ante los cambios ambientales. Sin embargo, tiene la gran ventaja de necesitar un bajo consumo de únicamente 3,5W, lo que la convierte en un excelente dispositivo para su implementación en robots autónomos



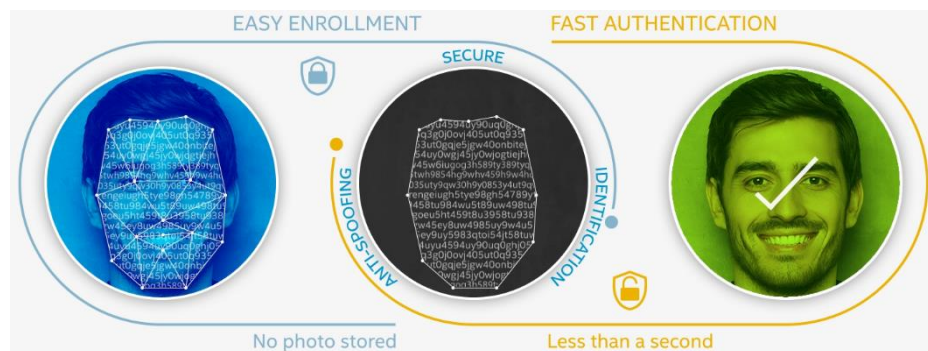
3.1.3 Cámara de reconocimiento facial F455

Se trata de otra de las interesantes apuestas de *Intel* que dan uso a su tecnología *RealSense* y cuyo objetivo se basa en proporcionar un sistema de autenticación facial que sea lo suficientemente seguro y fiable para poder utilizarlo de manera profesional.



El producto en sí se trata de una cámara pequeña y compacta que mediante el uso de medición de profundidad es capaz de determinar si el rostro escaneado de una persona coincide con los de los usuarios verificados o por el contrario se trata de un rostro desconocido.

Además de este modelo, se puede adquirir la versión F450, la cual carece de carcasa externa y está diseñada para poder incluirse como un sensor adicional en cualquier proyecto.



3.2 Cómo desarrollar con RealSense

La tecnología *RealSense* ha sido diseñada con fines de desarrollo, lo ha hecho que Intel ponga a disposición de usuario usuarios una biblioteca SDK con la que poder interaccionar fácilmente con esta tecnología. Se trata de un recurso de código abierto que nos permite programar utilizando C++, Phytón y JavaScript gracias a su interacción con node.js.

Esta herramienta permite la interacción con todos los elementos de los que están dotados los productos vistos anteriormente y facilita la elaboración de aplicaciones y proyectos que utilicen la tecnología *RealSense*.

Todos los productos utilizan la misma SDK *RealSense 2.0*, la cual se encuentra disponible tanto para Windows y MacOS como para Linux Y Android. Podemos acceder a la página de descarga oficial de la biblioteca haciendo click [aquí](#).

4. Bibliografía

Enlaces RealSense:

- Página oficial RealSense:
 - <https://www.intel.es/content/www/es/es/architecture-and-technology/realsense-overview.html>
- Artículo sobre las cámaras de la serie D:
 - <https://www.xataka.com/realidad-virtual-aumentada/intel-lanza-dos-camaras-realsense-con-sensor-de-profundidad-vision-en-3d-para-cualquier-dispositivo>
- Descarga oficial SDK 2.0 RealSense:
 - <https://www.intelrealsense.com/sdk-2/>

Enlaces Kinect:

- Tutorial cómo empezar a utilizar Kinect:
 - <https://programarfacil.com/blog/vision-artificial/kinect-y-processing/>
- ***Vídeo tutorial cómo instalar SDK kinect:***
 - <https://www.youtube.com/watch?v=dmumO8GBRbY>
- Descarga oficial SDK 2.0 Kinect:
 - <https://www.microsoft.com/en-us/download/details.aspx?id=44561>
- Artículo funcionamiento Kinect:
 - <https://programarfacil.com/podcast/86-sensor-kinect-inteligencia-artificial/#:~:text=Funciona%20a%20modo%20de%20webcam,luz%20infrarroja%20en%20una%20habitaci%C3%B3n.>
- Artículos sobre las diversas aplicaciones de Kinect:
 - <https://computerhoy.com/noticias/hardware/como-funciona-kinect-aplicada-ciencia-medicina-7596>
 - <https://sites.google.com/site/kinecteduka/home/ahora-futuro/hh>