**UNIVERSIDAD TECONOLÓGICA NACIONAL FACULTAD REGIONAL CÓRDOBA**

06/11/2012

**Proyecto Final Optical Marketing**

**Reporte de Investigación**

**Técnicas de Captura de información de los Objetos**

**Grupo 4 Carlos Kapica 51482 Rodrigo Liberal 51658 Julián Peker 51395**

**Fernández David 53063**

# Índice

# Introducción 3

# Marco Teórico 4

# Prueba de Concepto 6

# Conclusiones 9

# Introducción

En nuestro Proyecto OpticalMarketing en resumen algunos de los alcances para que se cumplan llevan a cabo ciertos procesos como pueden ser el monitoreo de Personas, la interacción de la personas con el aviso publicitario, capturar movimientos de personas, capturar gestos (mover una mano de cierta forma). Como observamos en unas de las investigaciones que no se puede hacer con una simple cámara web, por que toma videos o imágenes, luego hay que procesarlo con algoritmo de Computer visión, eso llevaría tiempo, no sería una interacción en tiempo real. Para ello sabemos que el sensor Kinect es un sensor que puede captar más datos que un video o una imagen, como por ejemplo el seguimiento de una persona, como los movimientos que realiza o reconocimiento de la voz. ¿Para ellos necesitamos conocer de qué forma capta esos datos?

# Marco Teórico

El sensor Kinect para pode rastrear y reconocer objetos en frente de ella, hacer un seguimiento de las personas, determinar la dirección de las señales de sonido, y aislarlos del ruido de fondo, tomar imágenes en la oscuridad, reconocimiento de voz, requiere manipular las cámaras o micrófonos combinándolos en algunos casos, pero el sensor Kinect en el caso que utilices la cámara RGB te crea una imagen digital, si utilizas la cámara infrarrojo te crea una imagen en profundidad(identificas con distintos niveles de gris la distancia y contorno de objetos), si utilizas el array de micrófonos te crea el sonido sin ruido. Para poder manipular esos datos que te captura el sensor kinect y poder realizar cosas más interesantes (seguimiento de una persona), necesitas utilizar Streams que te propone el SDK de Microsoft. Los Stream manipulan el hardware del sensor a bajo nivel, transforman las señales en flujos de Datos (Stream). Existen varios Streams:

* ColorImageStream
* DepthStream
* SkeletonStream
* AudioSource (no es un Stream)

**ColorImageStream**

Este Stream es el menos complejo de los tres por la forma en que se produce y los datos de configuración utilizando la cámara RGB produciendo una imagen digital. Trabajando con un flujo de datos Kinect es un proceso de tres pasos. Lo primero el Stream debe estar habilitado. Una vez habilitado, la aplicación extrae datos del Frame del Stream y, finalmente, la aplicación procesa los datos de Frame, proporcionándole formato a la imagen.

**DepthStream**

Este Stream produce de datos tridimensionales utilizando la cámara Infrarroja. Depende de usted para crear experiencias emocionantes con los datos. Una condición previa a la creación de una aplicación de Kinect está teniendo una comprensión de la salida del hardware. Más allá de simplemente la comprensión, el significado intrínseco del 1 y 0 es una comprensión de su significado existencial. Procesamiento de imágenes existen técnicas para detectar las formas y los contornos de los objetos dentro de una imagen. El SDK de Kinect utiliza la imagen profundidad para rastrear los movimientos del usuario en el SkeletonStream. Las imágenes de profundidad también pueden detectar objetos no humanos, como una silla o una taza de café, o distancia entre el sensor y el objeto que una imagen digital convencional proporciona solo dos dimensiones. Hay numerosos laboratorios comerciales y universidades activamente estudiando técnicas para llevar a cabo este nivel de detección de objetos a partir de imágenes de profundidad.

**SkeletonStream**

Los datos de profundidad producidas por Kinect tiene usos limitados. Para construir verdaderamente interactivo, divertido y memorable experiencias con Kinect, necesitamos más información más allá de la profundidad de cada pixel. Aquí es donde el seguimiento del Skeleton es parte del procesamiento de los datos de la imagen para establecer la profundidad de las posiciones de las articulaciones del Skeleton en una forma humana. Por ejemplo, el seguimiento de esqueleto determina dónde están: la cabeza de un usuario, las manos, y torso. Seguimiento Skeleton proporciona X, Y, y Z para cada uno de estos puntos del Skeleton. Es necesario analizar imágenes de profundidad que emplean complicados algoritmos que utilizan transformaciones matriciales, aprendizaje automático, y otros mecanismos para calcular los puntos de Skeleton. Este Stream detecta hasta 2 personas que puede identificar todos sus puntos y puede detectar 4 personas mas pero sin identifica su puntos, solo informa que existen. Se pueden identificar hasta 21 puntos del cuerpo como describe la siguiente figura.



AudioSource

El conjunto de micrófonos es la joya escondida del sensor Kinect. La matriz se compone de cuatro separado micrófonos linealmente hacia fuera en la parte inferior de la Kinect. Mediante la comparación de cuando cada micrófono captura la misma señal de audio, el conjunto de micrófonos puede ser utilizado para determinar la dirección desde que la señal está por venir. Esta técnica también se puede utilizar para hacer el conjunto de micrófonos prestar más atención al sonido de una dirección particular en lugar de otro. Finalmente, los algoritmos se pueden aplicar a los flujos de audio capturados de la matriz de micrófonos con el fin de realizar sonido complejo efectos de amortiguación para eliminar el ruido de fondo irrelevante. Todo esto sofisticada interacción entre Kinect Hardware y Kinect SDK software permite a los comandos de voz para ser utilizado en una gran sala donde hay más que de una persona hablando.

El Kinect es capaz no sólo de reconocer su voz, sino también para eliminar los ruidos de fondo, el sonido de su voz y la voz de sus amigos que viene hacia nuestro sistema de sonido, así como en el juego de música y explosiones en el juego. Reconoce la voz en inglés, español, italiano, francés, alemán y japonés.

# Prueba de concepto

1. Capturar una imagen utilizando el ColorImageStream



1. Realizar la captura de una imagen de profundidad utilizando DepthStream

La captura de pantalla a continuación. Me Muestra una imagen con distinto niveles de gris, pero cuanto más lejos estaba el objeto era color blanco



Viendo esta falencia en la imagen, decidimos aplicar un filtro de intensidad de grises para poder realizar una prueba de profundidad correcta. Como Resultado mejorando los niveles de profundidad y realizando mejor cálculo de la distancia de los objetos. EL objeto más cerca como las manos identificando su contorno de un color más oscuro y la persona que se encuentra más alejado identificando su contorno de un color más claro



1. Realizar El seguimiento de 2 personas utilizando SkeletonStream



# Conclusiones

EL Sensor Kinect puede realizar infinitas funcionalidades entre ellas reconocer objetos, movimientos de personas, detectar personas, reconocimiento de voz. Para ellos se aplican distintas técnicas de capturas: ColorImageStream, DepthStream, SkeletonStream y AudioSource.

Realizamos distintas pruebas básicas aplicando los Streams para poder determinar para que se aplicara cada Stream, se determino lo siguiente:

ColorImageStream: Para tomar imágenes o videos de un determinado lugar y realizar un monitoreo de las personas en donde se emite la publicidad.

DepthStream: Para detectar la distancia de una persona y en función la aplicación tomara decisiones de que acciones llevar a cabo. Si es posible detectar algún objeto.

SkeletonStream: Detectar la presencia, movimientos, gestos y posturas de las personas

El AudioSource por el momento no está contemplado en el proyecto por eso no se realizo la prueba.