

UNIVERSIDAD TECNOLÓGICA NACIONAL
FACULTAD REGIONAL CÓRDOBA



Proyecto Final

Optical Marketing

Reporte de Investigación - Técnicas de Captura de
información de los Objetos

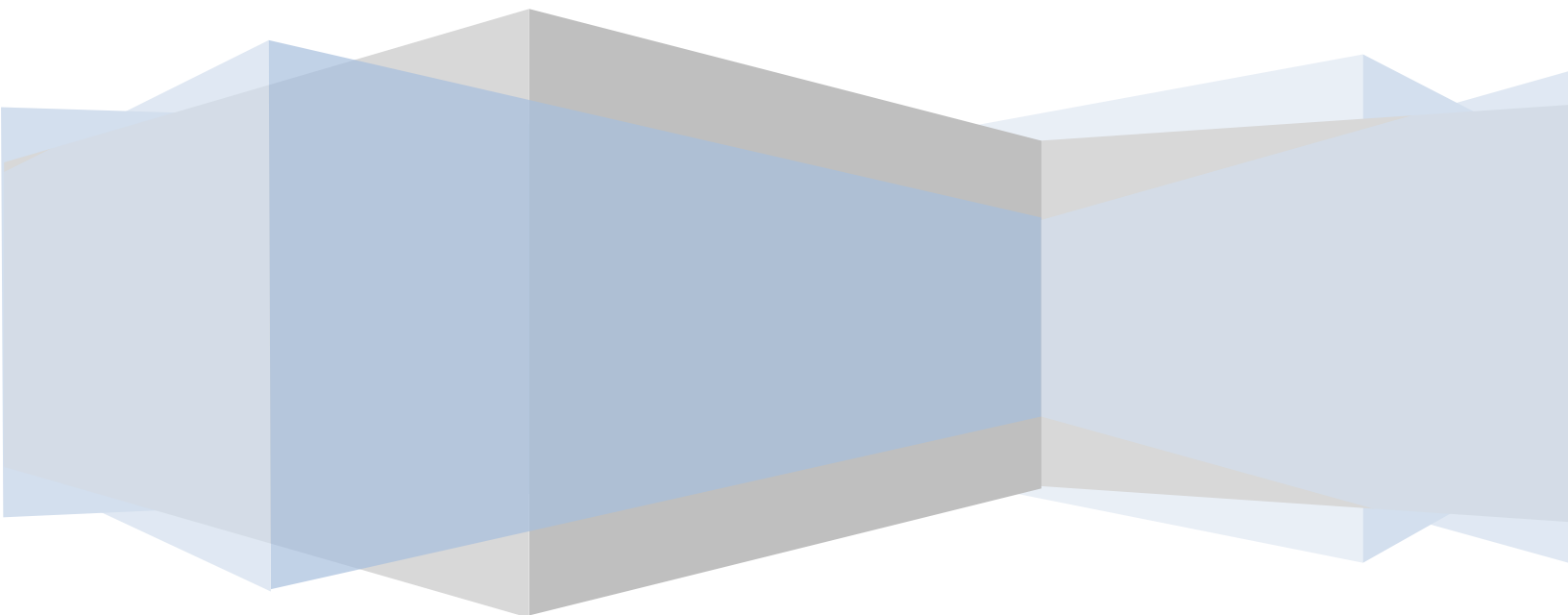
Grupo 4

Carlos Kapica 51482

Rodrigo Liberal 51658

Julián Peker 51395

Fernández David 53063



Índice

• Introducción	3
• Marco Teórico	4
• Prueba de Concepto	8
• Conclusiones	11

Introducción

En nuestro Proyecto Optical Marketing se 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 y capturar gestos.

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 algoritmos de Computer Vision, 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 a fin de poder procesarlos correctamente.

Marco Teórico

El sensor Kinect para poder 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 y combinar los mismos en algunos casos según el propósito de la aplicación a desarrollar.

Kinect en el caso que utilice la cámara RGB crea una imagen digital, si utilizas la cámara infrarrojo crea una imagen en profundidad (identifica con distintos niveles de gris la distancia y contorno de objetos), si utiliza el array de micrófonos crea el sonido sin ruido. Para poder manipular esos datos que captura el sensor Kinect y poder realizar cosas más interesantes (seguimiento de una persona), se necesita utilizar Streams. Los Stream manipulan el hardware del sensor a bajo nivel, transforman las señales en flujos de Datos (Stream). Los stream son abstracción necesaria para poder realizar el procesamiento inicial de la información. Existen varios Streams:

- ColorImageStream
- DepthStream
- SkeletonStream
- AudioSource (no es un Stream completamente definido)

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.

CATEDRA DE PROYECTO FINAL –OPTICAL MARKETING – REPORTE INVESTIGACION TECNICAS DE
CAPTURA DE INFORMACION DE LOS OBJETOS

DepthStream

Este Stream produce de datos tridimensionales utilizando la cámara Infrarroj. 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.

Dentro del 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.

SkeletonStream

Los datos de profundidad producidas por Kinect tiene usos limitados. Para construir algo verdaderamente interactivo, con Kinect, necesitamos más información más allá de la profundidad de cada pixel.

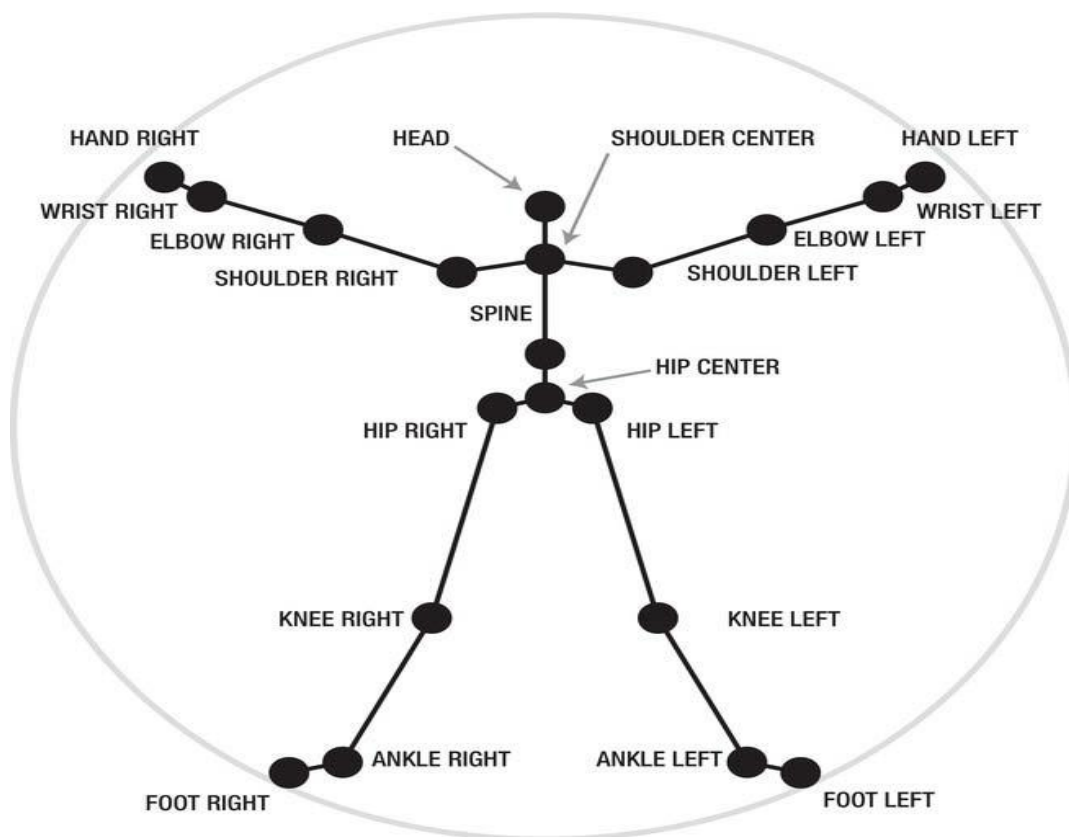
Aquí es donde el seguimiento del Skeleton es parte del procesamiento de los datos de la imagen para establecer la profundidad de las posiciones de las articulaciones del Skeleton en una forma humana.

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.

CATEDRA DE PROYECTO FINAL –OPTICAL MARKETING – REPORTE INVESTIGACION TECNICAS DE CAPTURA DE INFORMACION DE LOS OBJETOS



AudioSource

El conjunto de micrófonos es otro aspecto interesante 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.

CATEDRA DE PROYECTO FINAL –OPTICAL MARKETING – REPORTE INVESTIGACION TECNICAS DE
CAPTURA DE INFORMACION DE LOS OBJETOS

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



- 2) Realizar la captura de una imagen de profundidad utilizando DepthStream

La captura de pantalla a continuación muestra una imagen con distintos niveles de gris, mostrando con niveles de gris la profundidad de los objetos.

CATEDRA DE PROYECTO FINAL –OPTICAL MARKETING – REPORTE INVESTIGACION TECNICAS DE
CAPTURA DE INFORMACION DE LOS OBJETOS

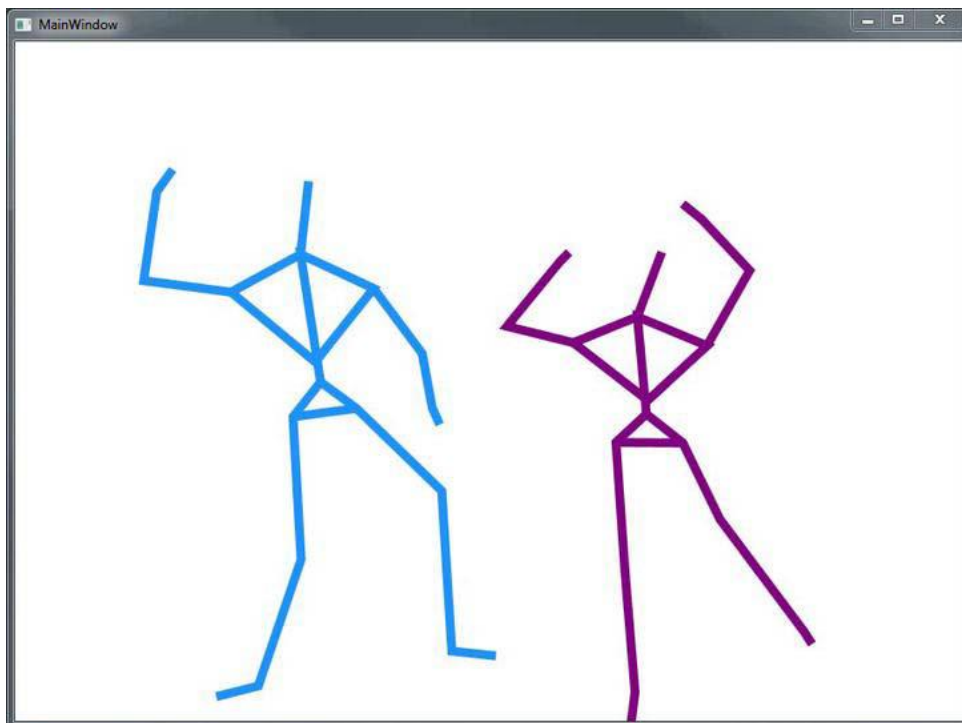
Viendo esta falencia en la imagen, decidimos aplicar un filtro de intensidad de grises para poder realizar una prueba de profundidad correcta. Como Resultado mejorando los



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



- 3) Realizar El seguimiento de 2 personas utilizando SkeletonStream:
Este ejemplo muestra 2 esqueletos trackeados con todas las joints posibles.



Conclusiones

El Sensor Kinect puede realizar infinitas funcionalidades entre ellas reconocer objetos, movimientos de personas, detectar personas, reconocimiento de voz.

Para ello se aplican distintas técnicas de captura empleando streams: ColorImageStream, DepthStream, SkeletonStream y AudioSource.

Realizamos distintas pruebas básicas aplicando los Streams para poder determinar para que se aplicará 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 tomará decisiones de que acciones llevar a cabo. Si es posible detectar algún objeto.

CATEDRA DE PROYECTO FINAL –OPTICAL MARKETING – REPORTE INVESTIGACION TECNICAS DE
CAPTURA DE INFORMACION DE LOS OBJETOS

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.