

Ezequiel Leonardo Aceto

# Manipulación de objetos reales y virtuales a través de las manos

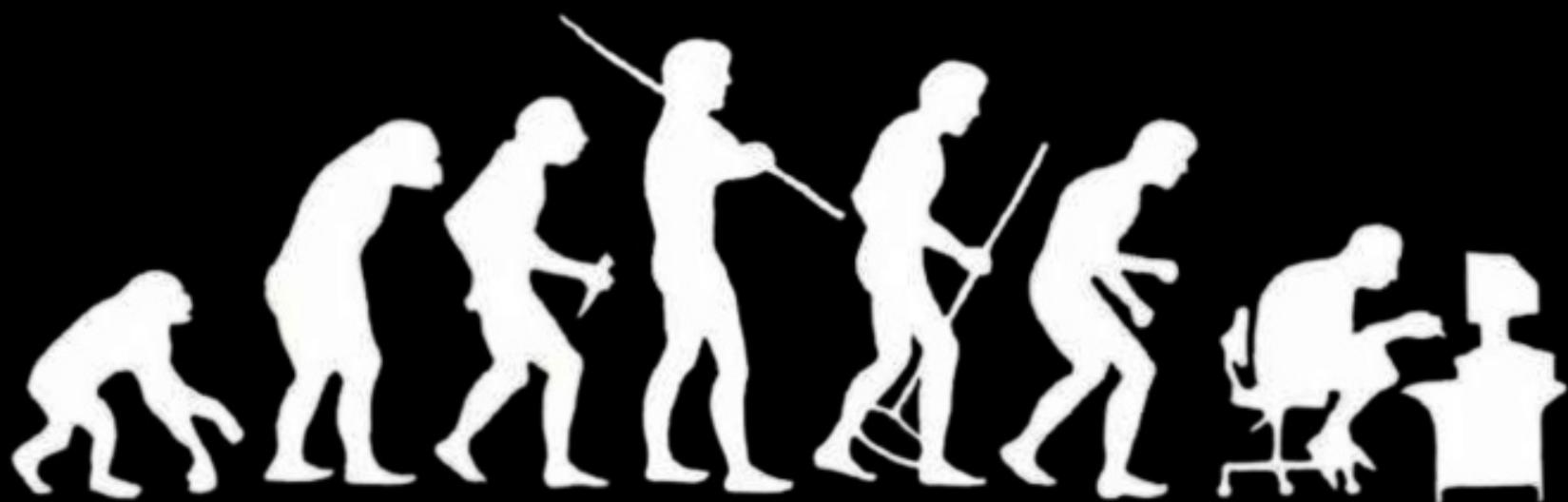


# 1

## Introducción

**“Computers in the future  
may weigh no more than  
1.5 tons.”**

– *Popular Mechanics, 1949*



El ser humano utiliza herramientas a diario, lo hace desde que el hombre descubrió el fuego. Desde la creación misma de las herramientas hasta el día de hoy, no solo evolucionaron estas y sus usos, sino también cómo las utilizamos.

Lo mismo ocurre en el ámbito de los sistemas de información computarizado. Las computadoras se vuelven más rápidas, los sistemas se vuelven más complejos, los negocios evolucionan, y el volumen de datos crece constantemente a un ritmo muy acelerado.

La vertiginosa evolución de estos sistemas tiene un punto débil, las interfaces de usuario. El primer teclado de computadora vio la luz en 1930, y a grandes rasgos su **layout** no ha sufrido modificaciones hasta el día de hoy.

Incluso sabiendo que el layout QWERTY fue diseñado para sortear las trabas mecánicas que tenían las maquinas de escribir convencionales, no hemos sido capaces de adoptar layouts que mejoren la velocidad de escritura, o al menos provean mayor comodidad al usuario.

Por otra parte el mouse fue diseñado en 1960 y se ha convertido en un elemento característico de toda computadora persona. Aún persiste sin mayores cambios, aunque poco a poco esta siendo desplazado por los **trackpads** presentes en las computadoras portátiles.

En el presente trabajo se estudiarán y analizarán nuevas tecnologías que proporcionan formas alternativas de interacción con los sistemas de información. Algunas de ellas interpretan los movimientos naturales del usuario, en lugar de forzarlo a aprender una nueva forma de trabajo, y otras son una versión mejorada de las tecnologías ya existentes.

Esta inspirado en la necesidad de encontrar nuevas formas de interacción con los, cada vez más complejos, sistemas informatizados. La propuesta no se limita a aquellos en los que se manipulan **objetos virtuales**, sino que pretende extenderse a los sistemas vinculados con el hardware. Existe la tecnología necesaria para innovar en este segmento y poner al servicio de los usuarios un interface más natural, desde la cual las acciones se

realicen con fluidez, sin que el usuario pierda el contexto en el que esta trabajando y de una forma cómoda y amigable.

Es entonces el problema sobre el cual gira este trabajo, el relacionado con la investigación de las diferentes tecnologías que existen en el mercado, su comparación, y posterior conclusión de la tecnología, o el conjunto de ellas, que mejor se adapta a la manipulación de objetos tanto virtuales como reales.

2

## Estado del Arte

“La simplicidad es la máxima sofisticación”

– Leonardo da Vinci



La literatura y el cine, considerado el Septimo Arte, han sido pioneros en la concepción de interfaces de usuario no solo futuristas sino también innovadoras.

En 2002 Steven Spielberg dio vida en la pelicula Minority Report, a un sistema de interacción basado en gestos y hologramas. Tiempo después, en 2009, Microsoft se encontraba anunciando el sistema Kinect, una interfaz que permitiría interactuar con la consola XBox 360, y sus video juegos, utilizando movimientos naturales para un ser humano. Es decir, la acción de patear una pelota de football ya no sería a través de una combinación de botones en un teclado o joypad, sino realizando el movimiento con los pies, como lo hacemos naturalmente.

Antes del lanzamiento de Kinect, en 2008, Marvel Studios presentaba el film Iron Man en donde el personaje principal era asistido en todas sus tareas a través de un sistema de realidad virtual comandado principalmente por ordenes de voz y movimiento de manos (J.A.R.V.I.S).

Fue Kinect, también conocido como “Project Natal”, uno de los primeros sistemas que intenta interpretar las acciones naturales del ser humano, en lugar de que sea el ser humano el que se adapte a la interface, como ocurre por ejemplo, con el mouse y el teclado.

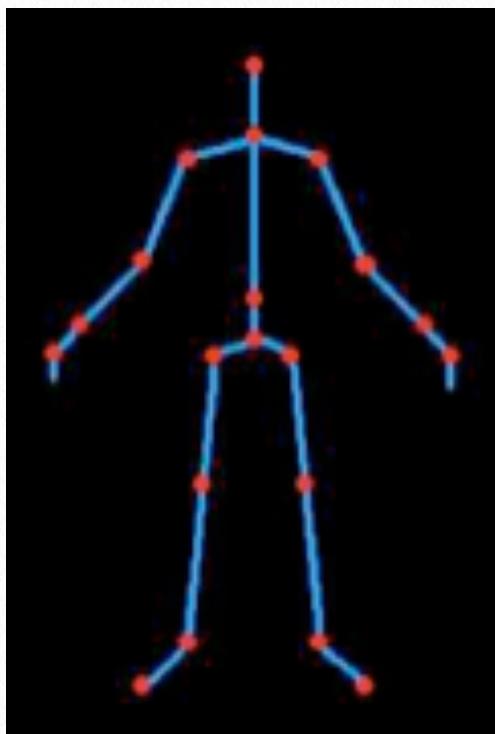
En el siguiente apartado se analizaran cuatro de las más innovadoras tecnologías de interface hombre maquina, y se cerrará el capítulo con un cuadro comparativo

## Microsoft Kinect

Este dispositivo permite interactuar con la computadora (o consola de videojuegos) sin contacto físico. Reconoce gestos, comandos de voz, objetos e imágenes incluso a distancias mayores de 1.5 metros. El dispositivo es capaz de detectar cualquier movimiento físico, aunque la detección y seguimiento de movimiento en los dedos esta aún en desarrollo, y aunque algunas pruebas fueron exitosas, no se ha liberado al SDK oficial. Respecto al kit de desarrollo, no hay SDK oficial para Mac OS X, ni para

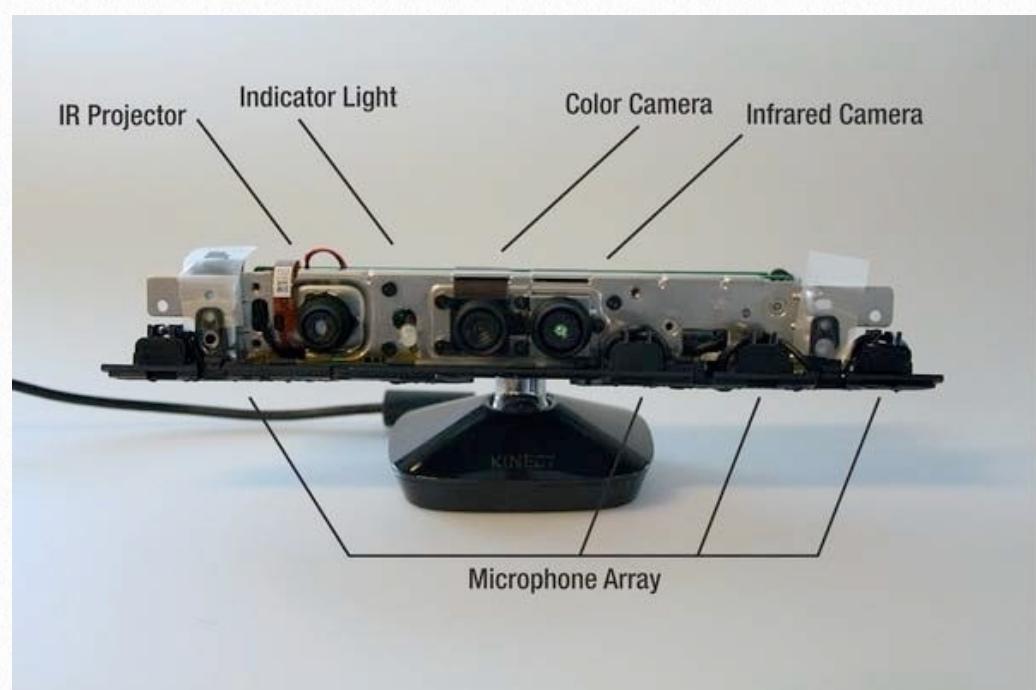
distribuciones Linux, solo para Windows OS. Aunque hay un proyecto open source ([openkinect.org/](http://openkinect.org/)) para que sea posible la utilización de este dispositivo en otras plataformas.

## ¿Cómo funciona la Kinect?



*Algunos puntos detectados por el sistema Kinect*

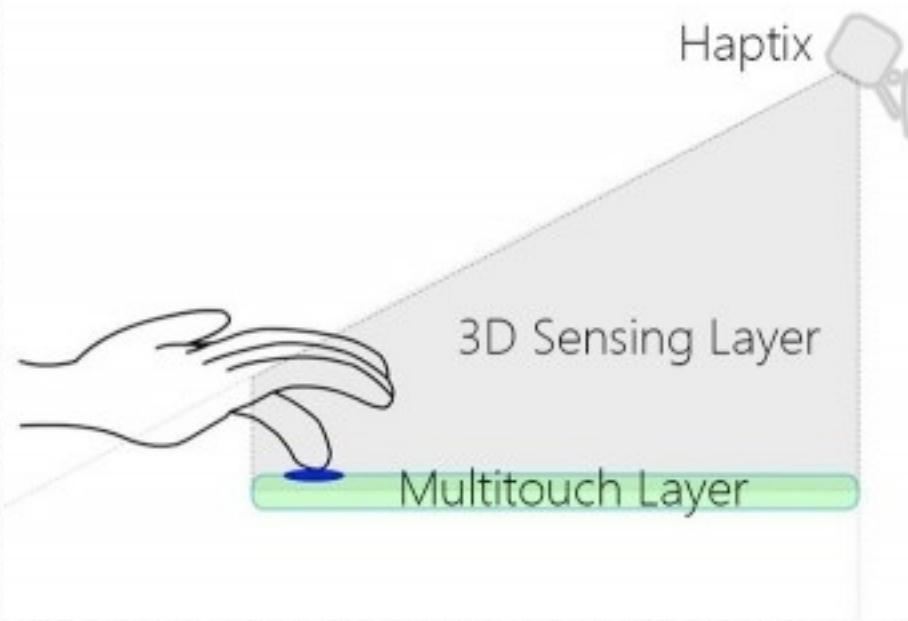
Inferir la posición del cuerpo es un proceso de dos etapas. En primer lugar se crea un mapa de profundidad, y luego se infieren las partes del cuerpo a través de técnicas de “Machine Learning”. El mapa de profundidad se construye mediante el análisis de un patrón generado por un laser de luz infrarroja. Esta técnica de análisis es conocida como “Luz estructurada”. Las partes del cuerpo son inferidas utilizando patrones de decisión que se basan en más de un millón de ejemplos de entrenamiento. Este entrenamiento, integrado en el hardware de la Kinect, requiere de veinticuatro mil horas de CPU para completarse, pero se lo reduce a un día utilizando un cluster de mil núcleos.



*Kinect por dentro*

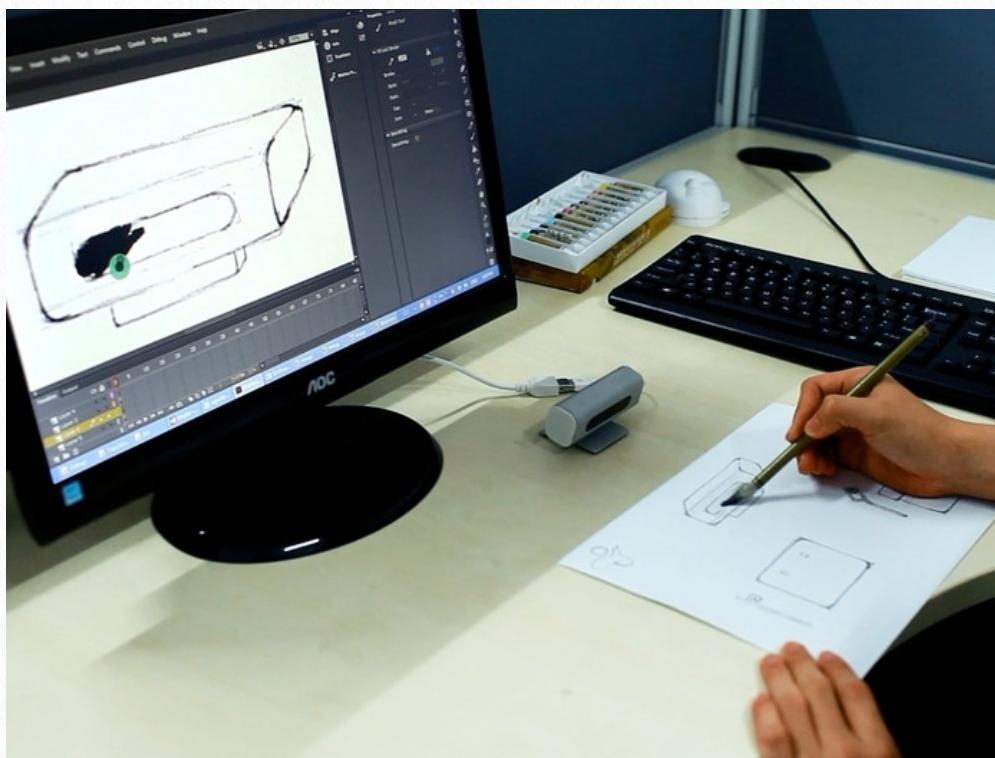
# Haptix

Este dispositivo pretende convertir cualquier superficie en una superficie táctil, sin la necesidad de modificar esa superficie. Tiene como principal ventaja que las manos pueden descansar sobre la superficie, por lo que reducimos la fatiga muscular producida por dispositivos como Kinect o LeapMotion.



Posee una amplia zona de contacto físico, llamada “Multitouch Layer” y una zona no tan amplia para detectar movimientos en el espacio. La excelente respuesta de este dispositivo en la zona de contacto multitouch lo hace perfecto para reemplazar dispositivos como el teclado o el mouse.

más acotada que la del LeapMotion. Ademas requiere de una superficie para interacturar con las manos, mientras el LeapMotion tiene como área de trabajo una esfera (en el espacio) de hasta 1 metro de diametro



## ¿Cómo funciona Haptix?

Haptix está compuesto por dos **sensores** de imágenes CMOS los que capturan la posición de las manos con una resolución de 640x360. Estos sensores están conectados a un microcontrolador que hace la mayor parte del procesamiento. Utiliza luz visible, pero emplea luz infrarroja cuando las condiciones de luz ambiental son muy bajas. Haptix se conecta a la computadora utilizando un puerto USB 2.0.



*Haptix por dentro*

## The Eye Tribe

Este dispositivo realiza un tracking preciso de los movimientos del ojo humano, y en base a estos movimientos desarrolla una interfaz entre la computadora y el usuario.

Este tipo de sistemas genera dudas sobre su implementación a gran escala debido a la cantidad de falsos positivos que el detector puede interpretar. Surgen preguntas como ¿Qué pasa si muevo el ojo sin intención de generar una interacción? ¿Cómo le indico al dispositivo que es un gesto y que no?

Su área de detección es de 40cm x 30cm a una distancia de 65m del sensor. Y el dispositivo posee SDKs en C++, C# y Java, solo para la plataforma Windows OS (XP SP1 o superior)



## Leap Motion

El dispositivo Leap Motion reconoce manos, dedos y hasta herramientas en un espacio de tres dimensiones esférico con centro en el dispositivo y radio de hasta 1 metro (en condiciones optimas).

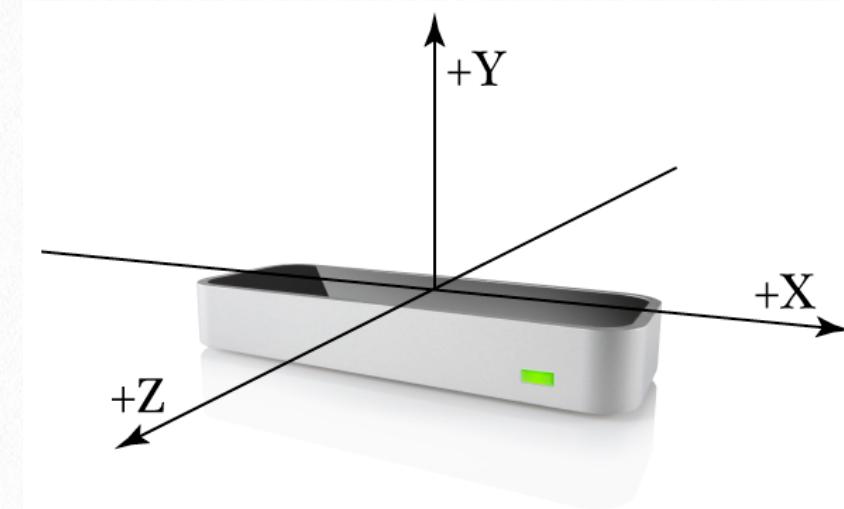
En la mayoría de los escenarios en los que se los utiliza las manos están suspendidas en el aire, lo cual genera fatiga muscular con el uso prolongado. Sin embargo, es posible utilizarlo por debajo de un cristal sobre el cual se pueden apoyar las manos.

Su pequeño rango de detección y tan alta sensibilidad lo diferencian de productos como el Kinetic, el cual esta diseñado para detección del cuerpo entero en amplios espacios.



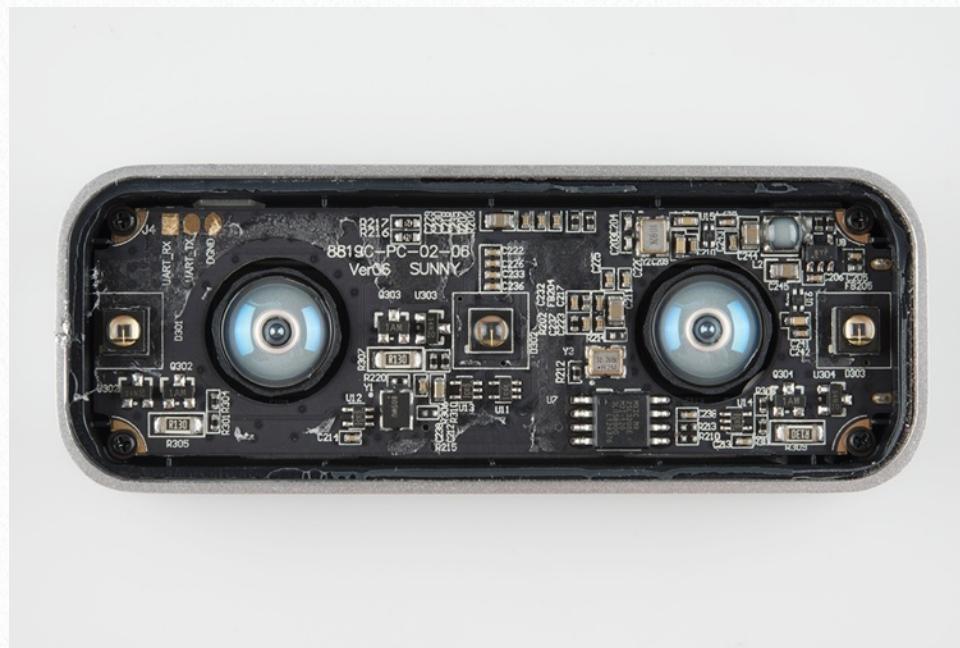
## ¿Cómo funciona Leap Motion?

El sistema de detección esta compuesto por dos cámaras monocromáticas sensibles a la luz infrarroja, y tres LEDs infrarrojos. Los LEDs generan un patrón de puntos en el espacio, y las cámaras capturan hasta 300 frames por segundo para procesar la información reflejada.



## El SDK de Leap Motion

El procesamiento de la información generada por el dispositivo es realizado por un software propietario, y solo se puede acceder a la información ya procesada a través de una API disponible para Linux, Windows y OS X en diferentes lenguajes (Objective C, C / C++, Python, Java, Javascript, C#, etc...). Lo cual lo hace una de las soluciones con mayor soporte multi plataforma y multi lenguaje.



Leap Motion por dentro

# Tabla comparativa

	Microsoft Kinect	Haptix	The Eye Tribe	Leap Motion
Detección de cuerpo entero	SI	NO	NO	NO
Detección de manos	SI	SI	NO	SI
Detección de dedos	NO	SI	NO	SI
Detección de movimiento ocular	NO	NO	SI	NO
Detección de herramientas	NO	NO	NO	SI
Distancia de detección	~ 1.5 metros	8cm (Min)	de 45 a 60 cm	1 metro (Max)
Soporte para Linux	NO	SI	NO	SI
Soporte para Windows	SI	SI	SI	SI
Soporte para Mac OS X	NO	NO AÚN	NO	SI
API en C / C++	SI	SI	SI	SI
API en Java	NO	SI	SI	SI
API en C#	SI	SI	SI	SI
API en Javascript	NO	NO	NO	SI
API en Obj C	NO	NO AÚN	NO	SI
API en Python	NO	NO	NO	SI

3

# El problema a resolver

**“Los ordenadores son  
inútiles. Sólo pueden darte  
respuestas.”**

*– Pablo Picasso*



Como se puede observar en el análisis realizado en el capítulo anterior (Estado del Arte) no existe una única tecnología capaz de cubrir todos los aspectos que se han analizado.

De existir, debería ser poder detectar los movimientos a gran escala del usuario, como por ejemplo: desplazamientos del usuario dentro de la habitación en la que este se encuentra, movimientos de las manos y de las otras partes del cuerpo. A este tipo de movimientos los llamo “movimientos macroscópicos” (desde el punto de vista del sistema) porque involucran traslaciones largas y partes del cuerpo fácilmente identificables.

También tiene que tener la habilidad de percibir los “movimientos microscópicos”, aquellos en los que las partes del cuerpo que utiliza el usuario hacen desplazamientos pequeños, que requieren un alto grado de precisión por parte del sistema para poder ser interpretar.

Y no debemos olvidar que ademas de los movimientos que utiliza el ser humano para expresarse, existe un sentido que es utilizado para focalizar y contextualizar las expresiones que naturalmente realizamos. Ese sentido, que es la visión, debe estar contemplado en un sistema de estas características dado que el sistema debe poder diferenciar cuando el usuario esta enfocado en la tarea, cuando cambia de contexto, y cuando pierde el foco en el trabajo que esta realizando.

Por lo que el desarrollo de un sistema que contemple las características antes mencionadas, para brindar una interfaz más dinámica, flexible y adaptada al comportamiento y movimiento del ser humano, no puede ser menos que una combinación de diferentes sensores. Que conectados a un sistema de procesamiento e interpretación de datos, permitan interpretar los comandos, que a través de gestos y expresiones, el usuario quiere ejecutar.

La aplicación de un sistema de estas características no esta limitado al ámbito de lo lúdico, como originalmente fueron concebidos distintos sensores que aquí se mencionan (Kinect, Leap Motion). Sino que tiene como idea original la utilización en ámbitos profesionales que van desde la robotica hasta la medicina. Considerandose de mucha

importancia en esta última área, por los beneficios que puede aportar en un diagnóstico o intervención.

En las siguientes imágenes se pueden observar situaciones en la que médicos, con el paciente en su consultorio o sala, realizan estudios u operaciones utilizando un instrumento tridimensional cuyo soporte informático es visualmente un sistema holográfico, con una interfaz de usuario basada en la interacción a través de las manos. La practicidad de esta implementación radica en el hecho de que para el médico es natural utilizar las manos para inspeccionar, como también para manipular los instrumentos.



# 4

## Solución propuesta

**“Yo nunca pienso en el futuro. Viene bastante rápido”**

*– Albert Einstein*



Se propone como solución a adoptar un sistema capaz de obtener datos a nivel macro y microscópico del usuario, que ademas debe ser capaz de detectar cuando el usuario esta enfocado en la tarea, y cuando pierde el foco o cambia de contexto.

Para lograr la detección de gestos y movimientos a gran escala se propone la tecnología Kinect de Microsoft. La misma es capaz de percibir los movimientos de todo el cuerpo a distancias mayores a 1.5 metros, de múltiples usuarios al mismo tiempo y con un nivel de precisión aceptable para la tarea que se quiere realizar.



*Dos usuarios interactúando con Kinect en un probar virtual de lentes de sol*

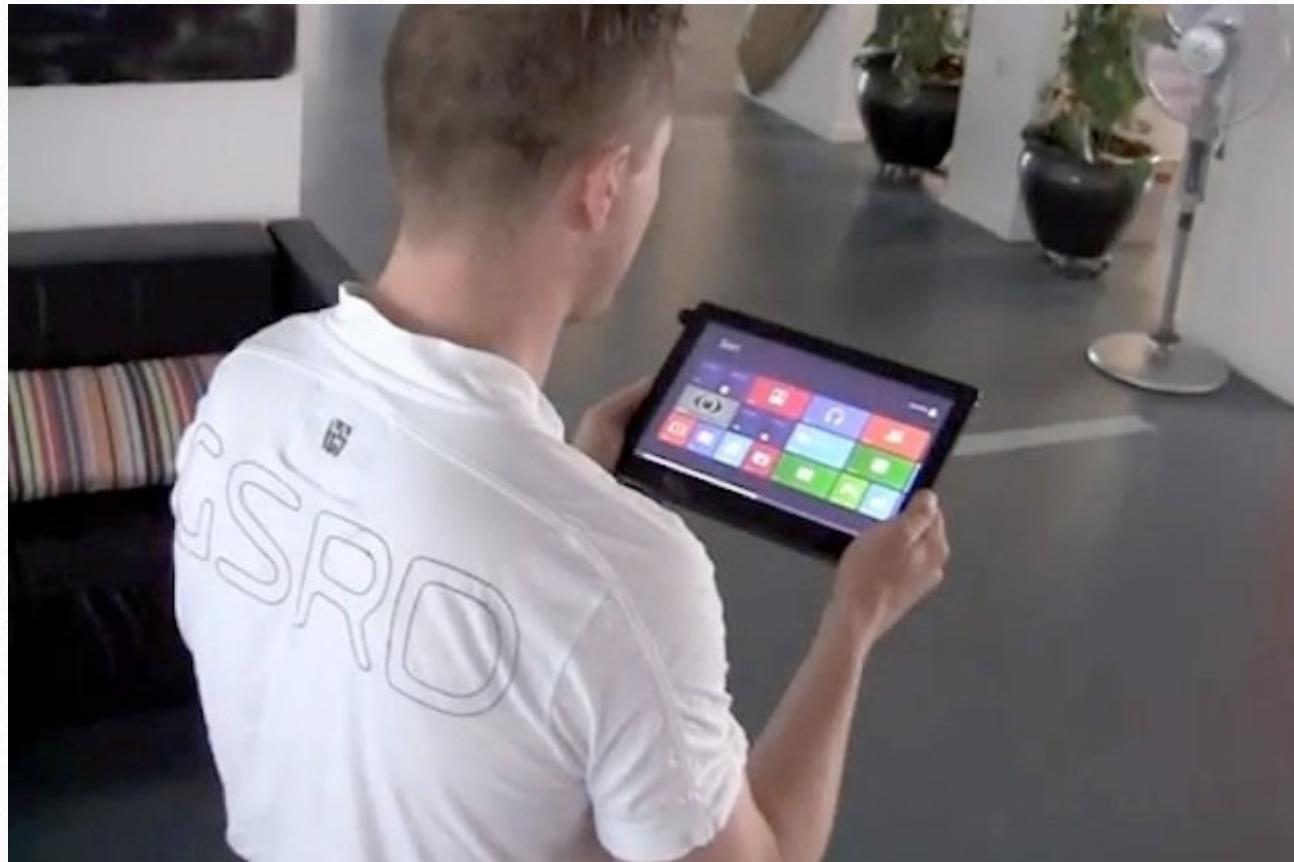
La detección de los movimientos de las manos del o los usuarios del sistema, se utilizará la tecnología Leap Motion. Ya que provee en un rango de acción bastante amplio (una esfera de aproximadamente 1m de diámetro) la capacidad de detectar con mucha precisión los movimientos de manos y dedos de múltiples usuarios al mismo tiempo. Ademas reconoce herramientas, como lápices, lo cual la hace una solución más que atractiva.



*Dos niñas utilizando Leap Motion para crear figuras con impresoras 3D*

Por ultimo, y dado que los seres humanos somos naturalmente multitareas, es importante que este sistema detecte cuando el usuario esta interactuando con el sistema y cuando no. El movimiento del rostro, quitandolo del área de trabajo, o incluso de los ojos al dirigir la mirada hacia otro punto, deben ser contemplados. Es por esto que se integrará a este sistema la tecnología The Eye Tribe que permite la detección del movimiento ocular.

Es importante reconocer que el sistema que el reconocimiento de voz sería un complemento ideal para un sistema como el que se describe en este trabajo. Sin embargo, se intenta circunscribir el mismo a las interacciones que los usuarios realizan con sus manos y su cuerpo. Aunque la tecnología The Eye Tribe esta limitada a campo de lo visual, reconozco la necesidad de utilizarla para poder contextualizar los movimientos del usuario.



*Un usuario utilizando una tablet con Windows 8 y The Eye Tribe*

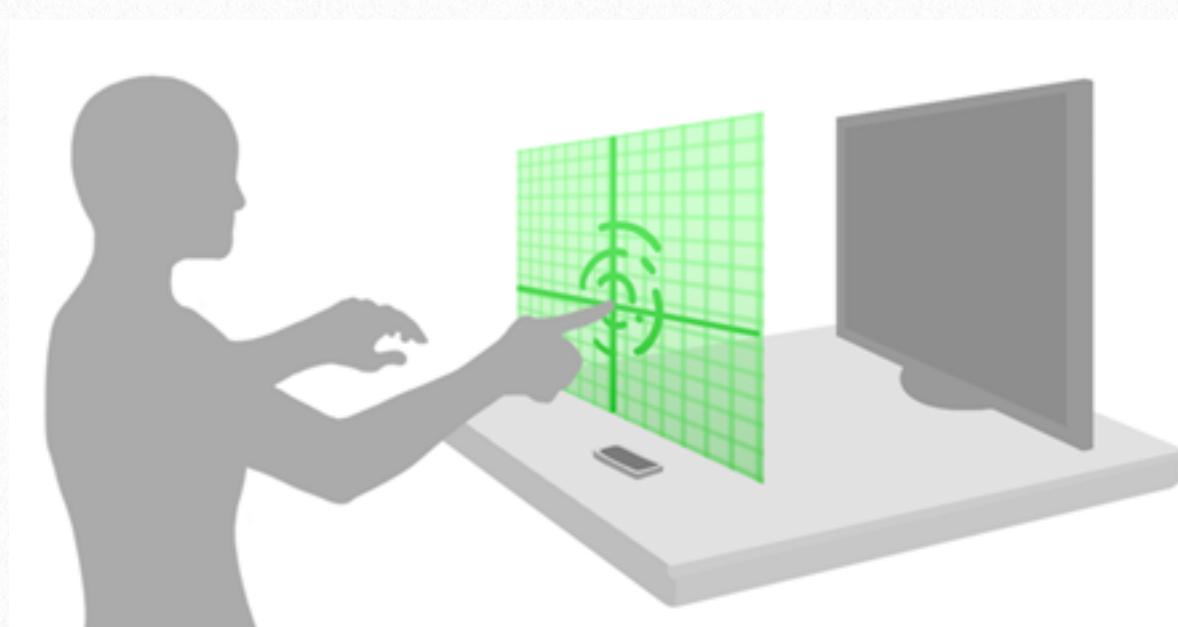
No quiero dejar de mencionar otras tecnologías que pueden servir de soporte para este sistema, como por ejemplo Google Glass, o cualquier otro tipo de gafas de **realidad aumentada**. Como así también la **holografía**.



# Layout de sistema

La disposición que estos sensores pueden adoptar no debe ser estática sino funcional al sistema con el que el usuario interactúa. Tampoco debe existir una restricción sobre el uso o no de alguno de ellos. En algunos sistemas ambientes puede ser útil la sola utilización de el sensor microscópico y el sensor ocular, y en otros el sensor ocular puede no tener mucha utilidad.

Por ejemplo, en el siguiente dibujo puede apreciarse un usuario que utiliza este sistema de manipulación sobre un escritorio con un monitor, ambiente en el que no es necesario utilizar el sensor de movimientos macros, sino que es suficiente con el sensor de manos y el ocular.



# Especificación de Requerimientos del Sistema

## Definiciones

<b>Usuario</b>	Persona que interactúa con el sistema informático
----------------	---

**Interfaz de Usuario** Es el medio con el cual el usuario puede comunicarse con el sistema y comprende todos los puntos de contacto entre el usuario y la maquina.

## Acrónimos

**ERS** Especificación de Requisitos de Software

**RFXXX** El estándar seguido para la especificación del identificador de cada requisito funcional será de la siguiente manera:

R = Requisito

F = Funcional

XXX = secuencia para la numeración de cada requisito

**RNFXXX** El estándar a seguir para la especificación del identificador de cada requisito no funcional será de la siguiente manera:

R = Requisito

F = Funcional

XXX = secuencia para la numeración de cada requisito



# 5

## Prototipo

**“La simplicidad es la máxima sofisticación”**

*– Leonardo da Vinci*



# 6

## Prueba de la solución propuesta

**“La simplicidad es la máxima sofisticación”**

*– Leonardo da Vinci*



7

# Futuras líneas de trabajo

**“La simplicidad es la máxima sofisticación”**

*– Leonardo da Vinci*



# Material Bibliografico

“Siempre imaginé que el Paraíso sería algún tipo de biblioteca.”

– Jorge Luis Borges



- IEEE Recomendad Practices for Sofware Requierements especification ANSI/IEEE 830 1998.

-

# Glosario

“Más vale una palabra a tiempo que cien a destiempo.”

– Miguel de Cervantes

## Glossary

**Agartha:** Secret underground kingdom beneath Reich; it is ruled by the King of the World.

**Ahnenerbe:** SS research organization for study of “heritage,” especially its occult, racial, or patriotic.

**Ariosophy:** Specifically Aryan interpretation of Teutonic paganism, rune lore, and the racial German people.

**Armanenschaft:** According to Guido von List, priest caste of the ancient German people.

**Artamanen:** Back-to-the-land movement current combining nature appreciation, agronomy, and resistance against Slavic incursions.

**Einsatzstab Reichsleiter Rosenberg:** Reichsleiterstab (ERR); organization officially tasked with collecting and cultural material from the occupied territories.

**Externsteine:** Natural stone formation revealed

## Holografía

La holografía es una técnica avanzada de fotografía que consiste en crear imágenes tridimensionales. Para esto se utiliza un rayo láser que graba microscópicamente una película fotosensible. Ésta, al recibir la luz desde la perspectiva adecuada, proyecta una imagen en tres dimensiones (holograma).

## Layout

La noción de layout suele utilizarse para nombrar el esquema de distribución de los elementos dentro de un diseño. En un teclado de computadoras es la disposición de teclas.



## Objetos virtuales

Son objetos intangibles que pueden representar entidades y/o eventos de la vida real, y están modelados en un software de computadora.

## Realidad Aumentada

Es el término utilizado para definir una visión directa o indirecta de un entorno físico del mundo real, cuyos elementos se combinan con elementos virtuales para la creación de una realidad mixta en tiempo real.

## Sensor

Un sensor es un dispositivo capaz de detectar magnitudes físicas o químicas, llamadas variables de instrumentación, y transformarlas en variables eléctricas. Las variables de instrumentación pueden ser por ejemplo: temperatura, intensidad lumínica, distancia, aceleración, inclinación, desplazamiento, presión, fuerza, torsión, humedad, movimiento,, etc

Las características típicas de un sensor son:

- Rango de medida: dominio en la magnitud medida en el que puede aplicarse el sensor.
- Precisión: es el error de medida máximo esperado.
- Resolución: mínima variación de la magnitud de entrada que puede apreciarse

## Trackpad

Panel táctil que permite controlar un cursor o puntero, utilizado generalmente en computadoras portátiles.

