



UNIVERSIDAD VERACRUZANA

FACULTAD DE INGENIERÍA

PROYECCIÓN ADAPTATIVA
DE UN ESCENARIO 3D
BASADO EN LA PERSPECTIVA
DE UNA PERSONA

QUE PARA OBTENER EL GRADO DE:
INGENIERO EN INFORMÁTICA

PRESENTA:
YADIRA FLEITAS TORANZO

ASESOR DE TESIS:
DR. LUIS FELIPE MARÍN URÍAS

BOCA DEL RÍO, VERACRUZ, JUNIO DE 2017

*Dedicado a mis padres,
por sacrificar gran parte de su vida para darme la mejor educación.*

Agradecimientos

gracias (:

Resumen

Una bonita historia jejeje

Índice general

Agradecimientos	II
Resumen	III
Lista de figuras	VI
Lista de tablas	VII
1. Introducción	1
1.1. Planteamiento del problema	3
1.2. Justificación	4
1.3. Objetivos	4
1.3.1. General	4
1.3.2. Específicos	4
1.4. Hipótesis	5
2. Marco Teórico	6
2.1. Antecedentes	6
2.1.1. Realidad Aumentada	6
2.1.1.1. Técnicas de visualización	6
2.2. Estado del arte	8
2.2.1. Projection mapping	9
2.2.2. Interacción en SAR	9
2.2.2.1. Proyecciones de Sobremesa	10
2.2.2.2. Proyecciones en superficies irregulares	10
2.2.2.3. Proyección en torno a la perspectiva	11
3. Diseño	12
3.1. Análisis de Requerimientos	12
3.1.1. Hardware del sistema	12
3.1.1.1. Kinect	12
3.1.1.2. Proyector	14

3.1.1.3.	Computadora	15
3.1.2.	Software del sistema	16
3.1.2.1.	Sistema Operativo	16
3.1.2.2.	Python	16
3.1.2.3.	LibFreenect	16
3.1.2.4.	OpenCV	17
3.1.2.5.	Blender	17
3.2.	Diseño General del Sistema	18
3.2.1.	Prototipo 1	19
3.2.1.1.	Detección manual	20
3.2.1.2.	Posición de la Persona	21
3.2.1.3.	Proyección con base en la Posición	22
3.2.2.	Prototipo 2	22
3.2.2.1.	Detección automática	23
3.2.2.2.	Altura de la Persona	25
3.2.2.3.	Proyección en base a la Altura y Posición	25
3.2.3.	Diseño preliminar de escenario	26
4.	Implementación y resultados	27
5.	Conclusiones	28
5.1.	Trabajo Futuro	28
	Bibliografía	29

Índice de figuras

2.1. Funcionamiento de los HMD's	7
2.2. Ejemplos de AR en smartphones y tablets	8
3.1. Componentes internos del Kinect	13
3.2. Esquema general del sistema	18
3.3. Esquema general de prototipo 1	19
3.4. Esquema general de detección del prototipo 1	20
3.5. Esquema general de posición del prototipo 1	21
3.6. Esquema general de proyección del prototipo 1	22
3.7. Esquema general de prototipo 2	23
3.8. Esquema general de detección automática	24
3.9. Esquema general de proyección de prototipo 2	25
3.10. Diseño preliminar del escenario de pruebas	26

Índice de tablas

3.1. Especificaciones técnicas del Kinect para Xbox 360.	14
--	----

Capítulo 1

Introducción

Vivimos en una época en la cual la tecnología forma parte esencial de nuestras vidas. En donde el ser humano, es capaz de mejorar y optimizar su entorno con el objetivo de satisfacer cada una de sus necesidades. El alcance que ha tenido la tecnología digital, nos abre camino para el surgimiento de nuevas técnicas capaces de colaborar en el desarrollo de múltiples áreas o disciplinas, ayudando a la creación de distintas formas para facilitar su aprendizaje, optimización y avance de las mismas.

Así, la tecnología digital ha servido de apoyo incluso en situaciones en donde el ser humano requiere de entrenamiento para realizar ciertas actividades peligrosas, costosas o difíciles de ejecutar y que son proporcionadas a través de entornos virtuales, para mejorar sus habilidades y evitar a todo costo el riesgo de perder incluso vidas humanas. Una aplicación que ejemplifica claramente lo mencionado, lo tenemos en los simuladores de vuelo, que lo que se intenta hacer es replicar la experiencia de pilotar una aeronave de la manera más realista posible. No obstante, el desarrollo e implementación de simuladores de este tipo, implica la fabricación de cabinas en tamaño real con accionadores hidráulicos o electromecánicos, el cual conlleva un gran costo de los mismos.

Es por ello que con el objetivo de minimizar costos y de no requerir de tantos equipos para la inmersión a estos escenarios artificiales, existe una rama de la tecnología conocida como Realidad Virtual (VR, Virtual Reality), que sustituye nuestro mundo real a través de dispositivos que nos permitan encontrarnos en otro lugar, es decir, sumergirnos en una realidad que no existe. Tomando el ejemplo anterior, se han estado desarrollando simuladores a través de este tipo de tecnología, con el objetivo de ahorrar recursos y que se puedan disponer de una manera más rápida y frecuente [20].

Los dispositivos utilizados para visualizar este tipo de Realidad Virtual, son conocidos como HMD (del inglés head-mounted display), similares a un cas-

co, que reproducen imágenes sobre una pantalla muy cercana a nuestros ojos, permitiendo abarcar todo el campo de visión del usuario y así brinden una inmersión de éste en un mundo ficticio.

Igualmente, existen algunos HMD's que son utilizados por otro tipo de tecnología basada en entornos virtuales y que adopta el nombre de Realidad Aumentada.

AR (del inglés Augmented Reality), es una tecnología que intenta lograr perfeccionar nuestra propia realidad, combinando nuestro mundo real con el virtual que, a diferencia de la Realidad Virtual, que crea todo un entorno artificial desde cero, éste tipo de tecnología agrega elementos virtuales a una realidad que ya existe.

Hoy en día, la Realidad Aumentada ha adquirido mayor popularidad y expansión debido a las aplicaciones en donde se encuentra. Por ejemplo, en la industria del entretenimiento, específicamente en el área de los videojuegos, existe una aplicación muy popular, conocida como Pokémon GO [10], el cual consiste en la búsqueda y captura de personajes de la saga Pokémon, que se encuentran escondidos en el mundo real, visualizados a través de un smartphone, en donde claramente se puede ver el concepto de Realidad Aumentada debido a que los personajes están superpuestos en nuestro mundo real. Así mismo, encontramos que la AR abarca áreas como la medicina [18, 21], turismo [16], manufactura [11], entre muchas otras.

Todas estas aplicaciones en donde se encuentra la Realidad Aumentada, son visualizadas a través de dispositivos tales como HMD (Hololens de Microsoft, GoogleGlass de Google), computadoras, tablets o smartphones, y aunque gradualmente se han ido optimizando tanto el hardware como el software de cada uno de ellos, su uso impide al observador tener una conexión más real con su mundo, debido a factores tales como el campo de visión limitado por los mismos dispositivos, latencia en los gráficos que en ocasiones puede provocar mareos, y la necesidad obligatoria de llevar un equipo encima para poder visualizar este tipo de realidad. Debido a lo anterior, existe una variante de ésta tecnología, conocida como Realidad Aumentada Espacial (SAR, Spatial Augmented Reality), que mantiene el mismo concepto de la AR, pero con la diferencia que ésta es mostrada mediante proyectores digitales, es decir, proyecta información u objetos virtuales directamente sobre objetos físicos, por lo tanto, ya no necesitaríamos llevar con nosotros algún dispositivo para visualizarla.

En un inicio, SAR, simplemente se utilizaba en la proyección de imágenes o videos que se adaptaban a la forma de fachadas de edificios ofreciendo espectáculos visuales. Pero más adelante se integró la fase de interacción con estos elementos virtuales, lo cual aunque ayudó a involucrarse en varias industrias y expandir su aplicación, aumentó su nivel de complejidad, encontrando

varias problemáticas que actualmente siguen en proceso de desarrollo.

1.1. Planteamiento del problema

La Realidad Aumentada Espacial (SAR), siendo entonces un nuevo paradigma que se deriva de la Realidad Aumentada, es una nueva forma de visualizar elementos virtuales sin la necesidad de verlos a través de dispositivos sujetos a estar en nuestro cuerpo.

Estos objetos virtuales proyectados en nuestro entorno, a pesar de jugar en algunos casos con nuestra mente, para dar un efecto 3D, utilizando sombras en lugares específicos y aplicando técnicas para dar profundidad, se puede lograr distinguir de lo real con sólo situarnos en otro lugar en torno a la proyección y verlo desde otra perspectiva. Esto ocasiona por lo tanto que exista aún una gran barrera entre lo que es real y lo que es virtual en términos de SAR.

El estudio de proyectar objetos virtuales 3D, que se asemejen tanto a los objetos físicos, pudiendo en algunas ocasiones sustituirlos, es una de las problemáticas actuales sin resolver y que muy pocos han intentado abordar, debido al grado de complejidad que pudiese tener el desarrollo de ésta. Considerando que, para que se pueda implementar esta mejora, se debe tener en cuenta varios aspectos tales como:

- La deformación de los objetos virtuales al proyectarse directamente sobre los objetos físicos.
- La detección de una o más personas.
- El seguimiento de la persona a la cual se le tomará en cuenta la perspectiva, debido a que únicamente se podría proyectar la perspectiva de una sola persona.
- La calidad de la proyección de los elementos virtuales.

Se propone, por lo tanto, a través de un esquema simple inicial y en un futuro, aplicable a cualquier escenario virtual, la proyección de un escenario con elementos 3D, adaptable a la perspectiva de la persona, perfeccionando y cumpliendo aún más con el objetivo del término de Realidad Aumentada Espacial.

1.2. Justificación

Aunque varios proyectos en donde se aplica SAR, se encuentran dirigidos a la industria de los videojuegos (lo cual, a pesar de lo que muchos opinan ser una pérdida de tiempo, éstos son capaces de desarrollar múltiples habilidades en los usuarios y vivir experiencias únicas), el uso de SAR, se expande a distintas áreas proporcionando enriquecer la experiencia visual, y por lo tanto apoyando el desarrollo y entendimiento de las mismas.

Es por ello que proyectar escenarios u objetos virtuales cambiando su ángulo de vista cuando una persona se mueve alrededor de éste, nos brinda la oportunidad de darle más realidad a un objeto virtual y vivir una experiencia mejorada de SAR.

En la educación el uso de esta técnica nos permitiría, agilizar el proceso de aprendizaje, debido al impacto que causaría la proyección de objetos virtuales 3D, en temas como figuras o lugares históricos, planetas, el cuerpo humano, entre otros, que dieran la posibilidad de mostrar más información de la proporcionada en una simple imagen, y así motivar su estudio.

En la rama de la arquitectura, la visualización de una estructura, ya sea un edificio o casa, proyectada en 3D sin la necesidad de crear una maqueta, ayudaría a ahorrar recursos y vivir la misma experiencia, incluyendo la ventaja de poder modificar su modelado y ver instantáneamente el resultado que obtendrían de manera virtual.

Por lo tanto, el desarrollo de mostrar un escenario virtual 3D, según la perspectiva de una persona, sería un paso más hacia la adaptabilidad de un mundo virtual en el mundo real, siendo una potente herramienta de visualización capaz de contribuir en el desarrollo de múltiples áreas.

1.3. Objetivos

1.3.1. General

Desarrollar un sistema capaz de determinar el ángulo correspondiente de un entorno virtual basado en la perspectiva de una persona y proyectarlo en nuestro mundo para proporcionar un efecto de realidad.

1.3.2. Específicos

- Determinación de los módulos que intervendrán en el sistema mediante el análisis de los requerimientos.
- Detección del número de personas en la escena y selección del individuo

de interacción, por medio del desarrollo y la implementación de un módulo.

- Determinar la posición en donde se encuentre la persona y la altura proporcionada por la imagen de profundidad de un sensor RGB - D, por medio del desarrollo e implementación de un módulo.
- Optimización del sistema por medio de una comparativa de diversas técnicas en cada uno de los módulos del sistema.

1.4. Hipótesis

Basado en la posición y altura de una persona, obtenidas mediante un sensor RGB – D es posible obtener su perspectiva respecto a un entorno virtual que será proyectado y creará el efecto 3D que tanto se intenta conseguir.

Capítulo 2

Marco Teórico

Dado que el presente trabajo de tesis, se centrará en el área de la Realidad Aumentada Espacial (SAR), específicamente en la optimización de la percepción de objetos virtuales 3D, será necesario establecer algunas terminologías y trabajos relacionados que sirvan de apoyo para el correcto entendimiento de éste.

2.1. Antecedentes

2.1.1. Realidad Aumentada

La Realidad Aumentada (AR), ha sido definida por varios autores desde distintos puntos de vista [3,19]. Algunos, la describen como una variante de la Realidad Virtual (VR); otros la considera como un concepto general, en donde VR se encuentra dentro de ésta tecnología. Ramesh Raskar, en su libro [6] menciona que AR, a diferencia de VR (que crea un entorno virtual desde cero), incorpora en nuestro entorno elementos sintéticos; es decir, superpone objetos o información virtual en una secuencia de imágenes capturadas en tiempo real de nuestro propio entorno, creando la ilusión de que ambos mundos (virtual y real) se fusionan en uno, es decir, creando una realidad mixta (MR).

2.1.1.1. Técnicas de visualización

Según Azuma [2], existen 3 principales formas en las que podemos visualizar la AR. Tales son a través de: dispositivos montados en la cabeza (head-worn displays, HWD), dispositivos de mano (handheld displays) o dispositivos de proyección (projective displays). La manera tradicional de per-

cibir esta realidad es a través de las dos primeras clasificaciones mencionadas anteriormente (en inglés See-through Augmented Reality, STAR [25]), mientras que la última es mayormente conocida como Spatial Augmented Reality (SAR).

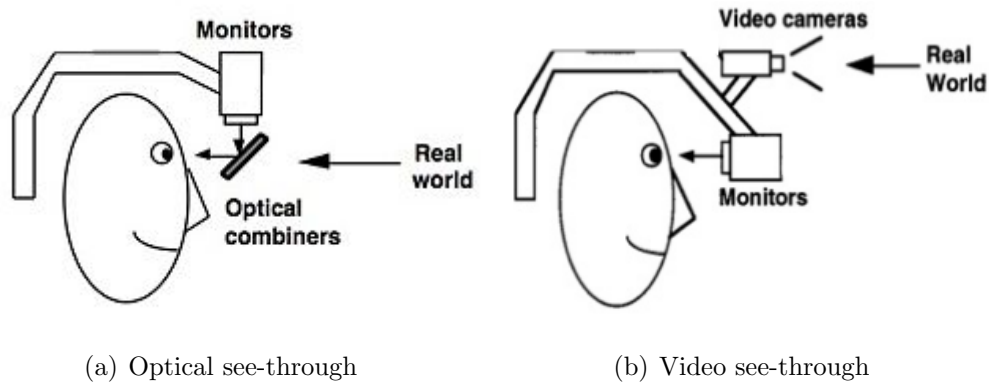


Figura 2.1: En (a) la superposición AR es proporcionada a través de una pantalla transparente, mientras que (b) las imágenes son capturadas por la cámara de video y usadas como fondo para la superposición AR.

See-through Augmented Reality. En el caso de los HWD's, también conocido como Head Mounted Displays (HMD's), estos son clasificados por dos tipos de técnicas de visualización: optical see-through y video see-through (Figura 2.1). Comercialmente hablando y los mas populares que podemos encontrar hoy en día, tenemos los Hololens de Microsoft y los Google glass, que a pesar de ser llamados como lentes de AR, se encuentran dentro del concepto de dispositivos en la cabeza.

En cuanto a los handhelds displays, traducido al español como dispositivos de mano, encontramos una amplia diversidad de dispositivos de varias marcas por parte de smartphones y tablets, que debido al bajo costo (en comparación con los HWD's) y al uso común de ellos, la AR se ha expandido y dado a conocer por medio del desarrollo de aplicaciones, que así como se ha utilizado como medio de entretenimiento, también ha servido de apoyo en el campo de la enseñanza, a través de aplicaciones tales como se presenta en (Figura 2.2). Sin embargo, aunque cada vez surgen mas aplicaciones entorno a estos artefactos, el uso de ellos impide al observador tener una conexión mas real con su mundo, debido principalmente al campo de visión limitado por los

mismos.

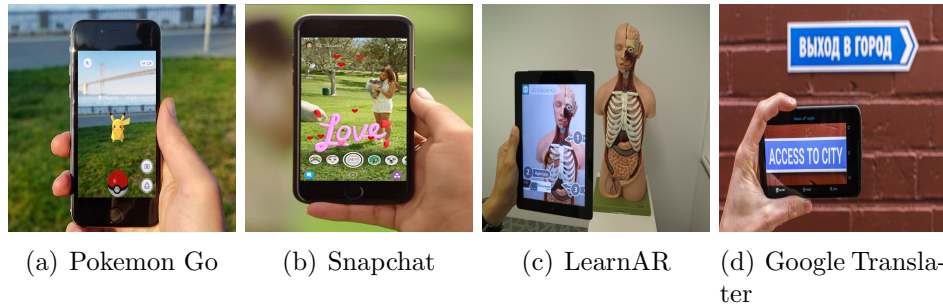


Figura 2.2: Los ejemplos (a) y (b) están enfocados al entretenimiento; (c) y (d) están enfocados a la educación

2.2. Estado del arte

La Realidad Aumentada Espacial, es una variante de la AR, que muestra mediante proyectores digitales, información u objetos virtuales directamente sobre objetos físicos, lo cual una de las ventajas que trae consigo con respecto a STAR, es el hecho de no tener que llevar un equipo encima para poder visualizar este tipo de realidad e involucrar la capacidad de poder interactuar directamente (sentido del tacto) de tal modo que nos ayude a estar mas en contacto con los objetos virtuales.

Raskar fue uno de los primeros en dar a conocer este nuevo paradigma a través de varios prototipos [22–24], los cuales se basaban principalmente en resolver problemas en cuanto a la distorsión de la imagen en la irregularidad de la superficie. Mas adelante, se implementó la fase de interacción entre el usuario y los elementos virtuales, lo cual atrajo la atención a grandes compañías tales como Microsoft, Sony, Disney, pero a la vez, surgiendo problemas mas complejos que aún siguen en proceso de desarrollo.

Dependiendo del uso o la aplicación en donde se encuentre ésta tecnología, se le atribuyen otros nombres que ayuden a entender mas su propósito en donde se esté implementando, ésto debido a que cuando hablamos de “Espacial”, no solo nos estamos refiriendo a la percepción de información adicional a través del sentido visual, sino que también podemos involucrar el tacto y el sentido auditivo.

2.2.1. Projection mapping

Comercialmente SAR, también es conocida como “projection mapping” o “video mapping”, refiriéndose a la proyección de información virtual en objetos reales, con el objetivo de enriquecerlos visualmente, mostrando una nueva forma de percibir el objeto. De igual manera se pueden proyectar imágenes o elementos virtuales en superficies tales como paredes de edificios para publicidad, como medio artístico en escenarios o incluso en el rostro o cuerpo de las personas [1].

Su nivel de complejidad, por lo tanto, varía dependiendo de la superficie en donde pueda ser proyectada la información. Por ejemplo, inicialmente, en proyectos como Luminous Room [27] se proyectaba la información en una superficie plana, donde únicamente se ajustaba la deformación que ocasionaba la proyección dependiendo de donde se encontrara el proyector. Sin embargo, años antes, ya se tomaba en cuenta la estructura del escenario [9] (en éste caso pantallas curvas) para proyectar información, en donde generaban las imágenes pre-distorsionadas para que luego se pudiese visualizar de la forma correcta. En cambio, cuando la superficie en donde se proyecta la información posee objetos y por lo tanto se muestra de forma irregular, se requiere un análisis previo de su forma para así capturar la información 3D de los objetos físicos, pre-distorsionar la imagen conforme a la irregularidad del escenario, tomar en cuenta la posición en donde se encuentra la persona y seguido a esto, la proyección correcta de los elementos virtuales [22, 24, 26, 29]. IllumiRoom [15], es un claro ejemplo en donde la información virtual se adapta en nuestro entorno, tomando en cuenta la geometría y apariencia de éste, con el objetivo de aumentar la zona que rodea una pantalla de televisión y mostrar un escenario extendido de éste.

2.2.2. Interacción en SAR

En sus primeros años de desarrollo, no se contaba con una fase de interacción entre el usuario y los elementos virtuales, debido a que su propósito en ese momento, era mostrar información adicional, aumentando la experiencia visual. Fue entonces, que gracias a grandes compañías como Microsoft Research, desarrollaron proyectos y propusieron nuevas técnicas que incluían la interacción a través de las manos o algún material o dispositivo de apoyo, lo cual propició la expansión de SAR en distintas áreas, principalmente en el aprendizaje y entretenimiento.

2.2.2.1. Proyecciones de Sobremesa

Mucho de éste desarrollo y avance se inició a través de proyectos que mostraban la información en sobremesa (tabletop). Éstas eran visualizadas a través de proyectores que generalmente estaban ubicados en el techo, proyectando la información encima de la mesa (front-projected displays) o debajo de ésta (técnica conocida como retroproyección).

Augmented Reality SandBox [17], es un proyecto, basado en front-projection, que actualmente se encuentra en algunos museos, desarrollado por investigadores de la Universidad de California en Davis, en el cual proponen una interactiva forma de aprender topografía, esto, utilizando arena en una caja especial, creando un modelo topográfico en relación a la forma en la que se encuentre la arena, es decir, dependiendo de la altura del relieve detectado por un sensor de profundidad, es proyectada la información real de como se comportaría el terreno.

La proyección frontal, en la mayoría de los proyectos ocasiona que las manos u otras partes del cuerpo de las personas, obstruyan la proyección, ocasionando poca realidad de inmersión de elementos virtuales en nuestro entorno [7, 17, 28]. Como una propuesta de solución a éste problema, Mirage-Table [4] proyecto desarrollado por Microsoft Research, propone una técnica, que es capaz de manipular los elementos virtuales como si fuesen objetos reales, adoptando las características físicas que pudiesen tener, tomando en cuenta la interacción y superposición de estos en las manos, sin la distorsión que causaría ésta.

En lo que respecta a la retroproyección, tomando el ejemplo anterior, trabajo previo a MirageTable, fue propuesto una robusta técnica de manipulación de objetos virtuales para este tipo de proyección, simulando la manera en la que nosotros utilizamos los objetos físicos [13]. También, encontramos que el uso de la retroproyección, ha servido para mostrar información virtual con múltiples usuarios, optimizando las técnicas de seguimiento de manos y reconocimiento de gestos, para un correcto uso multi-touch, apoyando el trabajo colaborativo [8, 12].

2.2.2.2. Proyecciones en superficies irregulares

Debido a la necesidad de llevar la tecnología SAR a nuestros hogares, principalmente como medio de entretenimiento, se han propuesto y desarrollado varias técnicas, en donde una de las principales cuestiones en las que se ha enfocado el estudio, es en la detección y modelado de la geometría de los objetos físicos, e incluso en la detección y ubicación de elementos en nuestro

entorno, para involucrar a los elementos virtuales en nuestro medio, con el objetivo de que se puedan desenvolverse de una forma mas real.

RoomAlive, es un complejo proyecto propuesto por Microsoft Research, que es capaz de transformar cualquier habitación en un área de entretenimiento de realidad aumentada espacial [14]. Poseen un sistema de seguimiento de personas que es utilizado para que el escenario cambie dinámicamente alrededor de ellos. El sistema, se basa principalmente en identificar superficies planas como paredes y piso, esto a través de una combinación de mapas de profundidades de 6 sensores RGB-D distribuidos por toda la habitación, de manera que se puedan clasificar las superficies, y así brindar la oportunidad de crear situaciones y escenarios virtuales mas reales.

2.2.2.3. Proyección en torno a la perspectiva

A pesar de que se ha logrado un buen avance en la proyección de elementos virtuales en nuestro entorno, de tal forma que no se distorsionen al ser mostrados en tiempo real, no se ha abarcado mucho en el estudio que toma en cuenta al objeto virtual 3D, mostrando sus respectivos ángulos o lados en torno a la perspectiva de una persona.

Un ejemplo que toma en cuenta éste estudio, lo podemos encontrar con otro proyecto por parte de Microsoft, llamado Mano-a-mano [5], el cual involucra las técnicas anteriormente mencionadas en RoomAlive y adicionalmente se realiza una adaptación del elemento virtual con respecto a la posición de la persona. En ésta propuesta, es tomada la perspectiva de dos personas que se encuentran una frente a la otra, lo cual involucra la interacción multi-usuario, y ser capaces de visualizar los objetos virtuales tal como si fuesen reales. Sin embargo, el ángulo en el que se pueden desplazar es limitado (menos de 90 grados) debido a la posición de los proyectores (parte de la imagen es proyectada directamente sobre la persona contraria), por lo tanto la interacción y visualización de los objetos virtuales 3D sigue siendo tema de estudio.

Capítulo 3

Diseño

El correcto diseño de un sistema así como el análisis de los elementos que lo conformarán, es de suma importancia, para su futura escalabilidad y optimización. Éste trabajo de tesis fue realizado a través de módulos los cuales permitan ser mejorados de forma independiente evitando a todo costo que la modificación de alguno de ellos afecte el funcionamiento de todo el sistema en general.

3.1. Análisis de Requerimientos

Antes de iniciar el desarrollo de cualquier sistema, se debe realizar un análisis de los elementos que estarán involucrados en éste para su correcto funcionamiento. Es por ello que a continuación se describen los componentes necesarios que se involucraron en el diseño y desarrollo del sistema.

3.1.1. Hardware del sistema

La parte física encargada de adquirir la información y mostrarla, consta de los siguientes dispositivos: computadora, sensor RGB-D (debido a que se utilizará un Kinect®) y un proyector.

3.1.1.1. Kinect

Para la detección y obtención de la información de la persona, se requiere de un dispositivo el cual nos proporcione todos los datos necesarios para facilitar el desarrollo del sistema. Para ello, se propuso el uso de un sensor

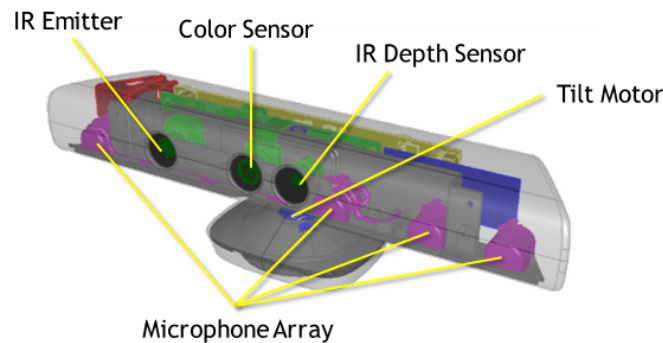


Figura 3.1: Componentes internos del Kinect

de profundidad el cual nos brindaría todo lo necesario para la fase de programación. Es por eso que se decidió utilizar un Kinect® versión 1, debido a su bajo costo en el mercado y a que posee dicho sensor de profundidad, además de una cámara RGB la cual se podría utilizar en trabajo futuro para conseguir un modelo de detección mas robusto. Dicho esto, procedamos a conocerlo a detalle.

Kinect es un dispositivo originalmente dirigido a los juegos de la consola Xbox 360®, desarrollado por Microsoft, el cual es capaz de a través de gestos, controlar e interactuar con la interfaz de la consola y los juegos en donde se requiera el uso de éste. Está conformado por los elementos mostrados en la Figura (3.1), en donde el sensor de profundidad, está formado por dos componentes: un proyector de luz infrarroja (IR Emitter) y un sensor CMOS monocromo estándar (IR Depth Sensor), los cuales se basan en un principio parecido al de triangulación entre los rayos visuales de los puntos proyectados sus correspondientes proyecciones en la imagen; además posee una cámara RGB (Color Sensor), cuyo funcionamiento es igual a la de una cámara digital estándar. También tiene 4 micrófonos (Microphone Array) y un motor (Tilt Motor) que es utilizado para orientar el campo de visión de las cámaras. Cabe mencionar, que existen dos modelos para la versión 1 del kinect, uno conocida como kinect para Xbox 360, que para su uso en una computadora se requiere de un adaptador USB, mientras que existe un kinect para Windows, que aunque posee mejoras, su costo hoy en día es mucho mas elevado que la versión para la consola y es por ello que en ése trabajo de tesis se optó por el primero. Teniendo en cuenta ello, en la tabla 3.1, se presentan las características principales del kinect para xbox 360.

Componentes	Características
Cámaras: Profundidad y Color	<ul style="list-style-type: none"> ■ Resolución: 320x240 ó 640x480 ■ Campo de Visión: 43° vertical, 57° horizontal ■ Velocidad de frames: 30 Frames por segundo (30 FPS) ■ Distancia recomendada: entre 0.8m y 3.5 m
Micrófonos	<ul style="list-style-type: none"> ■ Posee 4 micrófonos (ADC), con cancelación de eco acústico y supresión de ruido ■ Frecuencia: 16-khz ■ 24-bit mono modulación de código de pulso (PCM)
Motor	<ul style="list-style-type: none"> ■ Rango de inclinación: $\pm 27^\circ$

Tabla 3.1: Especificaciones técnicas del Kinect para Xbox 360.

3.1.1.2. Proyector

Debido a que éste trabajo de tesis es basado en la tecnología SAR, el uso de un proyector digital es indispensable para mostrar la información virtual. La calidad de éstos varía principalmente en el tipo de tecnología utilizada, modelo y distancia focal con el objetivo. A partir de ello, los elementos a tener en cuenta al momento de elegir un proyector son: tamaño de la imagen, resolución, luminosidad y contraste, tiempo de vida de la lámpara y el nivel sonoro. Enfocándonos en la luminosidad de un proyector, ésta es expresada en lúmenes y se puede definir como la medición de la emisión de color de un proyector. Mientras mayor sea su número, más vivo y real será la imagen, que es uno de los objetivos que se intenta conseguir en éste trabajo de tesis. Para la implementación se utilizará un proyector por parte de la Universidad Veracruzana con las siguientes características:

- x lúmenes
- -x- píxeles

Lugar de proyección Para obtener buenos resultados de la imagen, es necesario tomar en cuenta el color y el material en donde será proyectado. En un inicio se había considerado proyectar la imagen encima de una mesa de color negro (éste tipo de proyección es considerada como front-projection que fue explicado en el capítulo anterior). Sin embargo, esta disposición del proyector, no permite que el observador se acerque mucho a la proyección, debido a que puede obstruir la imagen proyectada con su cabeza o manos. Como solución a ello, se decidió utilizar la técnica de retroproyección, la cual permitirá acercarse a la proyección sin la limitante anterior debido a que el proyector se encontrará dentro de una mesa y la imagen será proyectada de adentro hacia afuera. El material ideal para este tipo de proyección es conocido como DNP Holoscreen, pero el precio es muy elevado, rondando los 500 dls, por lo que se tiene pensado utilizar materiales de papel, en donde se pueda retener parte de la luz proyectada.

3.1.1.3. Computadora

Para determinar las características mínimas que se necesitan para que nuestro sistema funcione correctamente, se debe realizar un análisis previo del hardware y software que se utilizará. Después de dicho análisis, se determinó que una computadora con los siguientes requisitos mínimos no debería tener problemas en la ejecución del programa:

- Procesador de 32 bits (x86) o 64 bits (x64)
- Procesador de doble núcleo, de 2.66-GHz o más rápido
- Bus USB 2.0 dedicado
- 2 GB de RAM
- Puerto VGA para proyector

La computadora en la cual se realizará el desarrollo y ejecución del sistema, es una laptop dell, modelo 1550 Precision, que cuenta con las siguientes características:

- información de la workstation dell

3.1.2. Software del sistema

En esta sección se describen los elementos de software que se utilizarán en el desarrollo del sistema, desde el lenguaje de programación, hasta los programas auxiliares y librerías, para una correcta comprensión del código del programa.

3.1.2.1. Sistema Operativo

El sistema operativo en donde se desarrollará el proyecto, será Ubuntu, específicamente la versión 16.04. Es una de las distribuciones Linux, que a diferencia de otros sistemas operativos es capaz de aprovechar mejor los recursos hardware, y que debido a su bajo consumo de éstos la velocidad de procesamiento es mayor. Una de las ventajas y una de las razones por la cual se utiliza éste OS, es debido a que cuenta con el paquete de python 2.7 (por defecto) que será el lenguaje de programación utilizado, lo cual facilita el uso del mismo a través de terminal, además de que la instalación de las librerías y conexión con el kinect se obtienen de una manera sencilla a través de pocos comandos.

3.1.2.2. Python

Python es un lenguaje de programación interpretado multiparadigma y multiplataforma y será el lenguaje de programación que se utilizará para el desarrollo del sistema. Entre muchas de las ventajas que nos ofrece son su facilidad de uso, portabilidad, legibilidad y simplicidad. Fue diseñado principalmente para expresar en forma clara y directa instrucciones que debe seguir un programa, sin la necesidad de estar indicando los detalles de bajo nivel los cuales involucran los tipos de variables, sus tamaño y manejo de memoria; es por ello que se le denomina intérprete, debido a que es capaz de inferir esos detalles. Lamentablemente esas ventajas tienen su costo, debido a que se ejecuta mas lento que otros lenguajes tradicionales (por ejemplo los compilados, que son ejecutados directamente por el procesador). A pesar de ello, cada vez existen mas librerías (tales como numpy, math las cuales fueron utilizadas en el desarrollo) y herramientas de apoyo que agilizan algunos procesos que son mas lentos sin son desarrollados directamente en este lenguaje, lo cual convierte a python en un lenguaje potente.

3.1.2.3. LibFreenect

Casi inmediatamente con la aparición del kinect en el mercado, y debido a su facilidad de adquisición (precio muy bajo en comparación con otros

sensores), muchas fuentes alternas a través de ingeniería inversa lograron desarrollar controladores que permitían acceder a la información tanto de la cámara RGB como la de profundidad, convirtiéndose entonces en un sistema de visión artificial muy utilizado por investigadores, que incluso Microsoft un tiempo después decidió lanzar un SDK (Software Development Kit) para el uso del kinect, pero dirigido a su sistema operativo y utilizando csharp como lenguaje de programación.

OpenKinect es una comunidad abierta, que distribuyen utilidades y aplicaciones para el kinect, y que ha sido de gran ayuda para el desarrollo de éste sistema. La librería en donde se encuentran todos los drivers, funciones y ejemplos es libfreenect, y está disponible para muchos sistemas operativos (entre ellos Ubuntu), es fácil su instalación y uso.

3.1.2.4. OpenCV

OpenCV (Open Source Computer Vision Library) es un librería que ha sido de gran utilidad en el área de la visión por computadora. Debido a que python es el lenguaje utilizado, el uso de ésta optimiza y agiliza el proceso de análisis y procesamiento de imágenes, lo cual es una gran ventaja debido a que el proceso se realiza en tiempo real. La versión utilizada es la 2.4.9.1, y se utilizó en gran medida para el tratamiento de las imágenes y detección, en conjunto con otra librería muy utilizada, numpy, que se utilizó mayormente para operaciones con matrices.

3.1.2.5. Blender

La información virtual que se mostrará en nuestro entorno, será diseñada a través de Blender. Éste es un programa multiplataforma que se puede utilizar como modelado, animación y creación de gráficos en 3D. De igual manera, posee un motor de juegos internos, lo cual abre la posibilidad de que en futuro se pueda agregar una fase de interacción entre los elementos virtuales y el usuario. Es distribuido de forma gratuita y es compatible con Ubuntu. Además, el uso de la cámara virtual en blender nos proporciona el campo de vista que el observador debería tener, habiendo entonces una equivalencia entre la cámara y la persona en el mundo real. El lenguaje utilizado es python (pero la versión 3, lo cual tiene algunas diferencias con la 2.7). La versión de blender utilizada es la 2.76

3.2. Diseño General del Sistema

Después de un análisis detallado de los elementos que podrían estar involucrados en el desarrollo del sistema, se determinó como diseño general el siguiente esquema (Figura 3.2), el cual se realizó para definir las etapas por las cuales tendrá que seguir el sistema cada vez que se capture una imagen (frame) a través de un sensor de profundidad, en éste caso el kinect de Microsoft©, descrito en la sección anterior. Este análisis fue basado en la premisa de que para obtener la perspectiva de un persona correspondiente a un escenario 3D, se debe determinar primero si se encuentra una persona en el entorno, para así conocer la información respecto a ésta (es decir su posición y altura), y así poder proyectar los elementos virtuales en relación con el observador.

Se siguió el patrón de desarrollo de software conocido como modelo incre-

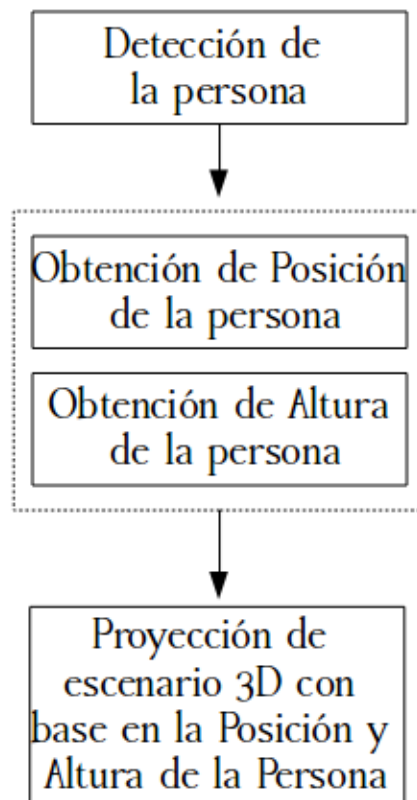


Figura 3.2: Esquema general del sistema

mental (o iterativo), que es una mejora del modelo tradicional de cascada,

el cual consiste en la secuencia de las etapas del tradicional (análisis, diseño, implementación y pruebas) pero de una manera superficial, el cual se utiliza mayormente para obtener resultados a corto plazo del software (a veces se pueden obtener prototipos funcionales por varios incrementos), con la posibilidad de agregar o quitar funcionalidades conforme vaya avanzando el sistema, para obtener mejores resultados en cada uno de los módulos. Dicho esto, como primera fase o prototipo 1, se diseñó el sistema detectando a la persona de forma manual (mas adelante se explicará ésto), basándose únicamente en la posición de la persona (sin tomar en cuenta su altura), lo cual traería como resultado, que el escenario se adaptara a la posición de la persona rotando en el eje vertical, sin variar el ángulo de inclinación entre la proyección y el observador. Una vez implementado esta fase, pasaríamos a la siguiente o prototipo 2, el cual incluirá la detección automática y la obtención de la altura de la persona basado en la imagen de profundidad para la proyección correcta del escenario.

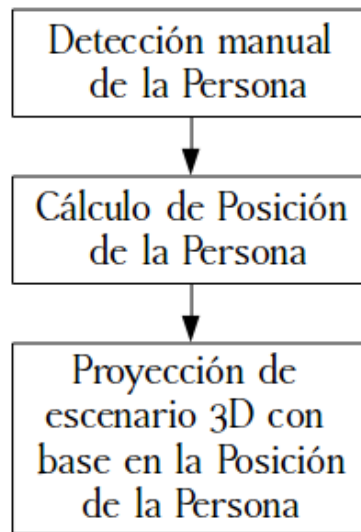


Figura 3.3: Esquema general de prototipo 1

3.2.1. Prototipo 1

Como fase inicial, este prototipo será la base para mejoras futuras, por lo que debe tener un diseño claro y bien estructurado, para evitar problemas en su funcionamiento a largo plazo. Constará con un fase inicial de detección la cual denominaremos detección manual debido a que se ajustará el umbral

de detección de forma manual; además se diseñará el módulo para obtener la posición de la persona y el módulo para proyectar el escenario 3D con base en su posición (Figura 3.3)

3.2.1.1. Detección manual

Como fue mencionado anteriormente, el diseño inicial de detección de la persona se basará dependiendo del umbral asignado por el usuario. Es decir, la imagen de profundidad será obtenida por el kinect en la escala de grises, se binarizará a través de un umbral dado por el usuario que delimitará el fondo de la persona para poder obtener su contorno. Debido a que el sensor de profundidad se encontrará a una distancia aproximada de 2.90 metros del suelo, capturando entonces la información desde arriba, el contorno que necesitaremos será el de la cabeza de la persona. El esquema general que muestra los pasos para éste módulo lo podemos encontrar en la Figura 3.4.

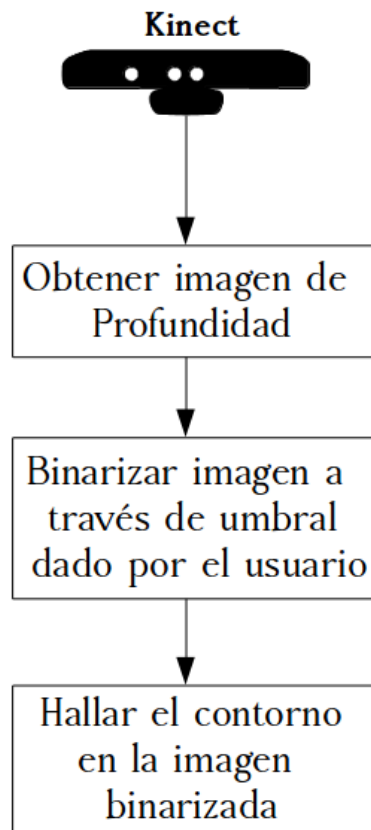


Figura 3.4: Esquema general de detección del prototipo 1

3.2.1.2. Posición de la Persona

Para calcular la posición de la persona, se determinó que a través de la figura generada del contorno de ésta, se hallaría el punto central, lo cual, en otras palabras nos daría la posición del pixel calculado, en un rango de 640x480 (tamaño de la imagen), en donde ésta información se almacenará en un archivo para que se pueda obtener en blender. Para poder llegar a éste módulo, es necesario que se haya detectado la persona en la etapa anterior, de lo contrario, si no se obtiene un contorno, no se puede determinar la posición, y el programa se detendría. En la figura 3.5 se muestran los pasos para determinar la posición de la persona.

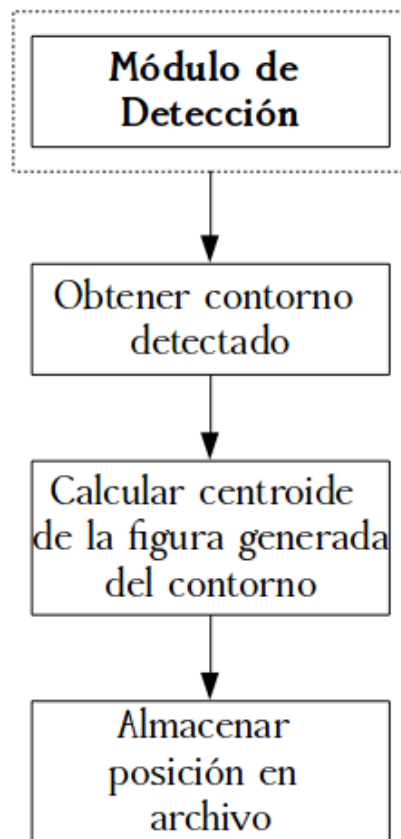


Figura 3.5: Esquema general de posición del prototipo 1

3.2.1.3. Proyección con base en la Posición

Este módulo, se encuentra internamente en el programa Blender, y es el encargado de posicionar la cámara virtual, en la posición obtenida del archivo. El escenario recreado posee las mismas medidas que el de la cámara depth 640x480, pero una unidad menos, es decir 64.0x48.0 debido a la gran expansión que ocasionaría si se utilizara las dimensiones originales. De éste modo, al obtener la posición de la persona, la posición de la cámara será la dimensión original dividida entre diez. Es importante recalcar, que aunque la cámara y la posición de la persona, serán equivalentes, al momento de posicionar la cámara virtual, esta mantendrá el mismo ángulo en el eje vertical, por lo cual será necesario dirigir el campo de vista de ésta hacia el centro del escenario calculando su respectivo ángulo. En la figura 3.6 podemos ver los pasos para llegar al objetivo final.

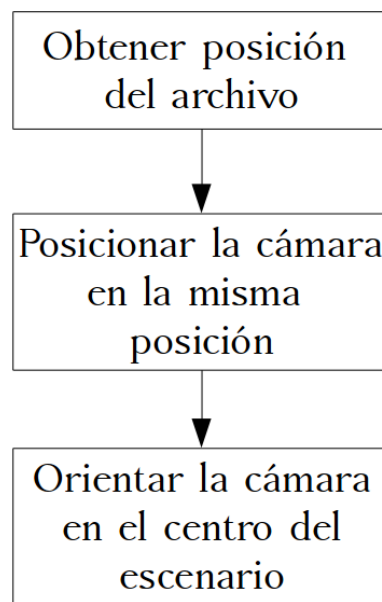


Figura 3.6: Esquema general de proyección del prototipo 1

3.2.2. Prototipo 2

Después de obtener un prototipo funcional, se incluirá a través de varias mejoras un prototipo 2, el cual involucre la optimización de la detección de la persona, logrando que se realice de manera automática y la adición del

módulo de la altura, lo cual implicaría modificar un poco la parte de proyección en blender y sus ángulos en la cámara. Lo anterior se puede resumir en la figura 3.7, en donde se definen cada una de las etapas que se analizarán a continuación.

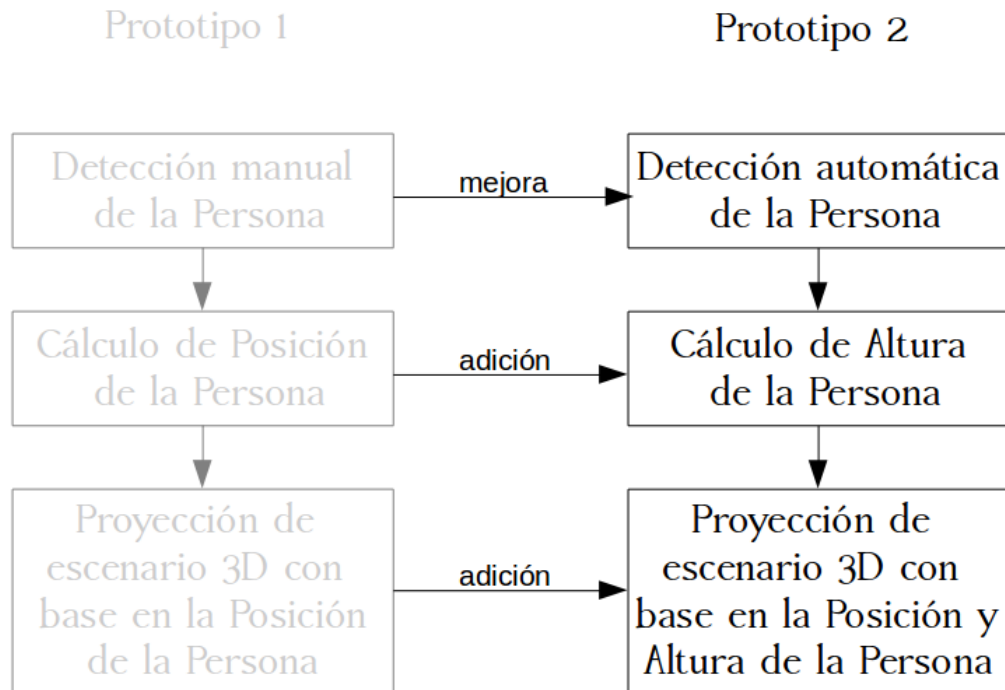


Figura 3.7: Esquema general de prototipo 2

3.2.2.1. Detección automática

Basado en los pasos definidos en el módulo de detección manual (donde para detectar el contorno de la cabeza de la persona se ajusta el umbral por el mismo usuario dependiendo de la altura de éste), se diseña un modelo de detección capaz de ajustar éste umbral de manera automática, basado en un rango de área permitida. Es decir, si el umbral es muy elevado, el contorno detectado estará involucrando más allá del área de la cabeza de la persona (por ejemplo los hombros) dando como resultado que el cálculo de la posición podría verse afectada. Por lo tanto, se define un intervalo de área mínima y máxima permitida, calculada a través de un rectángulo delimitador que

contendrá el contorno (esto para dar un poco mas de libertad en las áreas detectadas), y dependiendo del área calculada del rectángulo delimitador con el umbral actual, se irá ajustando éste umbral hasta que su área se encuentre dentro del rango definido. Al tomar en cuenta esto, el umbral se ajusta de manera automática, de tal modo que si una persona es de mayor o menor altura, detectará la región de la cabeza, y ésta aunque varíe su altura (se incline o agache) siempre se seguirá el área de la cabeza, algo que en el módulo anterior no se podía realizar. En la figura 3.8, se muestra un esquema de cómo se realizaría ésta etapa del sistema.

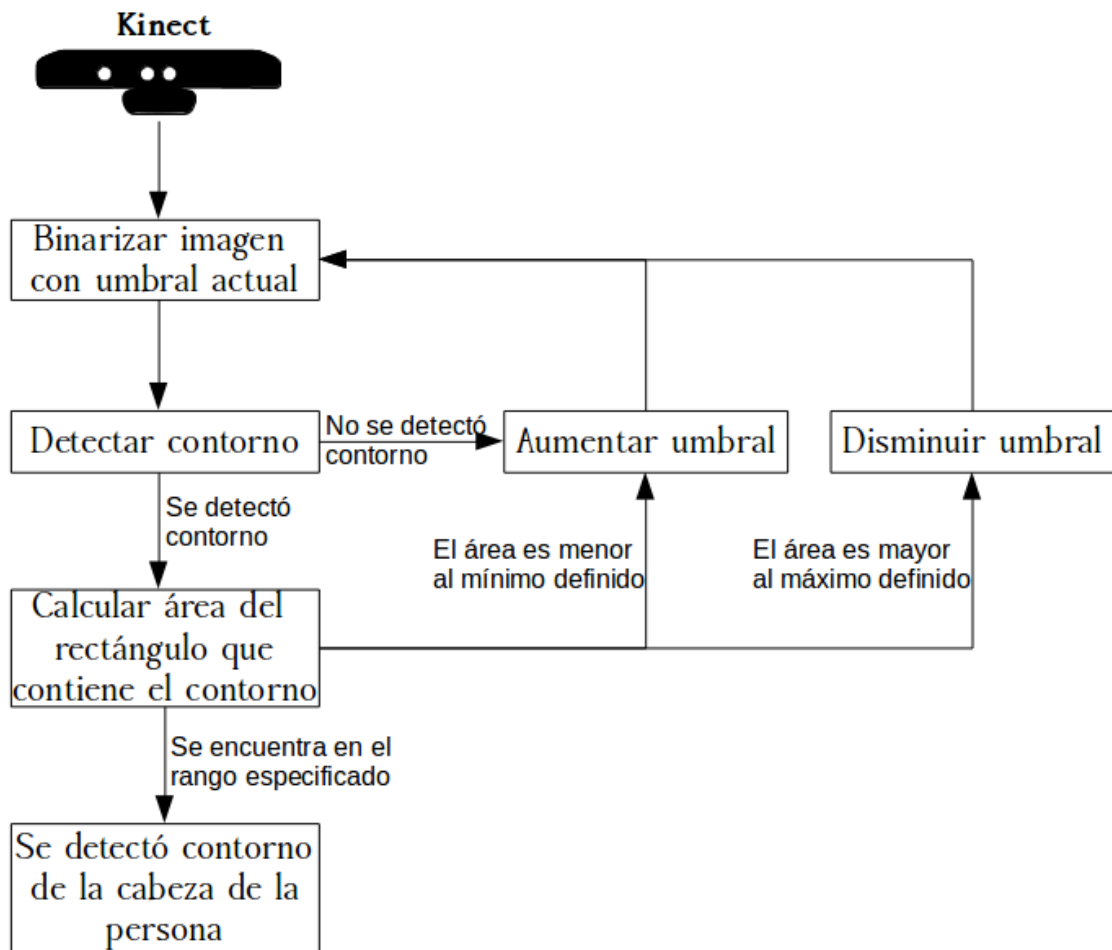


Figura 3.8: Esquema general de detección automática

3.2.2.2. Altura de la Persona

Éste tal vez sea el módulo mas rápido y fácil de implementar. Debido a que la detección automática ajusta el umbral para obtener el contorno de la cabeza de la persona y luego es obtenida la posición a través el centro del contorno, el cual se puede interpretar como la posición en píxeles de la imagen, la altura entonces es determinada por el valor de la imagen de profundidad en el punto proporcionado por el módulo de posición. Por lo tanto, éste módulo podría encontrarse dentro del módulo de posición y almacenar la información en un archivo como coordenadas en un espacio 3D, para su posterior uso en el módulo de proyección.

3.2.2.3. Proyección en base a la Altura y Posición

En lo que respecta a éste módulo, al adicionar la fase de detección de la altura de la persona, se debe hacer una equivalencia en el eje vertical de la cámara en blender con el valor dado por la imagen de profundidad. Una vez realizado el paso anterior, se orienta la cámara, es decir se inclina enfocando al centro del escenario (ésto se verá en detalle en la fase de implementación). En la figura 3.9 se muestra un resumen de lo explicado anteriormente.

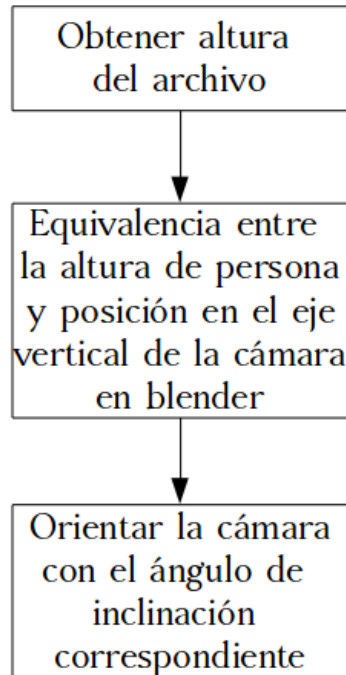


Figura 3.9: Esquema general de proyección de prototipo 2

3.2.3. Diseño preliminar de escenario

El lugar en donde se proyectará el escenario virtual 3D y se realizarán las pruebas del sistema, será a través de una técnica conocida como retroproyección, explicada anteriormente, a través de un material semi-transparente, simulando una mesa, en donde el usuario podrá caminar al rededor de ésta, pudiendo abarcar todo el campo de visión, es decir, podrá darle una vuelta de 360° en torno a ésta. Como fue analizado en la sección anterior, como diseño previo del escenario en donde serán realizadas las pruebas del sistema, constará de un proyector que se localizará en la parte inferior de la mesa. Para que tengamos una idea mas clara de lo explicado, en la figura 3.10 se muestra un diseño previo del escenario de pruebas.

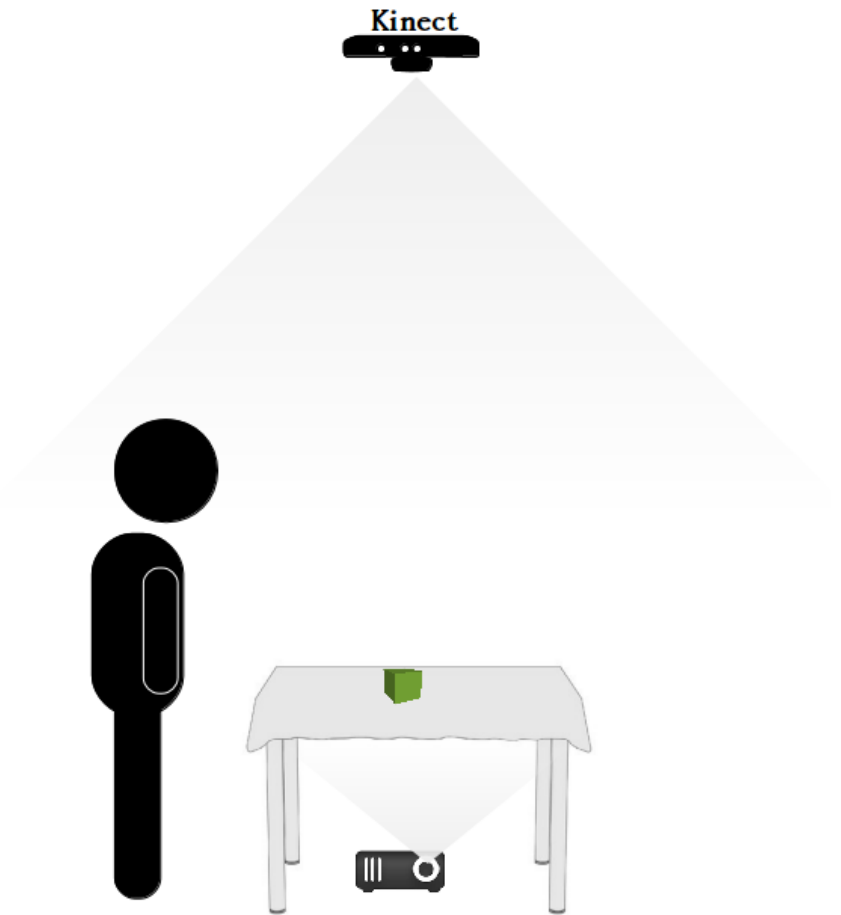


Figura 3.10: Diseño preliminar del escenario de pruebas

Capítulo 4

Implementación y resultados

tampoco nadita):

Capítulo 5

Conclusiones

y con esto concluimos /O/

5.1. Trabajo Futuro

muchos xD

Bibliografía

- [1] Projection Mapping Central. <http://projection-mapping.org>, 2017.
- [2] AZUMA, R., BAILLOT, Y., BEHRINGER, R., FEINER, S., JULIER, S., AND MACINTYRE, B. Recent advances in augmented reality. *IEEE computer graphics and applications* 21, 6 (2001), 34–47.
- [3] AZUMA, R. T. A survey of augmented reality. *Presence: Teleoperators and virtual environments* 6, 4 (1997), 355–385.
- [4] BENKO, H., JOTA, R., AND WILSON, A. Miragetable: freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI conference on human factors in computing systems* (2012), ACM, pp. 199–208.
- [5] BENKO, H., WILSON, A. D., AND ZANNIER, F. Dyadic projected spatial augmented reality. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (2014), ACM, pp. 645–655.
- [6] BIMBER, O., AND RASKAR, R. *Spatial augmented reality: merging real and virtual worlds*. CRC press, 2005.
- [7] DIETZ, P., AND LEIGH, D. Diamondtouch: a multi-user touch technology. In *Proceedings of the 14th annual ACM symposium on User interface software and technology* (2001), ACM, pp. 219–226.
- [8] DOHSE, K., DOHSE, T., STILL, J. D., AND PARKHURST, D. J. Enhancing multi-user interaction with multi-touch tabletop displays using hand tracking. In *Advances in Computer-Human Interaction, 2008 First International Conference on* (2008), IEEE, pp. 297–302.
- [9] DORSEY, J. O., SILLION, F. X., AND GREENBERG, D. P. Design and simulation of opera lighting and projection effects. In *ACM SIGGRAPH computer graphics* (1991), vol. 25, ACM, pp. 41–50.

- [10] DORWARD, L. J., MITTERMEIER, J. C., SANDBROOK, C., AND SPOONER, F. Pokémon go: benefits, costs, and lessons for the conservation movement. *Conservation Letters* 10, 1 (2017), 160–165.
- [11] FRÜND, J., GAUSEMEIER, J., MATYSCZOK, C., AND RADKOWSKI, R. Using augmented reality technology to support the automobile development. In *International Conference on Computer Supported Cooperative Work in Design* (2004), Springer, pp. 289–298.
- [12] GELLER, T. Interactive tabletop exhibits in museums and galleries. *IEEE Computer Graphics and Applications* 26, 5 (2006), 6–11.
- [13] HILLIGES, O., IZADI, S., WILSON, A. D., HODGES, S., GARCIA-MENDOZA, A., AND BUTZ, A. Interactions in the air: adding further depth to interactive tabletops. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology* (2009), ACM, pp. 139–148.
- [14] JONES, B., SODHI, R., MURDOCK, M., MEHRA, R., BENKO, H., WILSON, A., OFEK, E., MACINTYRE, B., RAGHUVANSHI, N., AND SHAPIRA, L. Roomalive: magical experiences enabled by scalable, adaptive projector-camera units. In *Proceedings of the 27th annual ACM symposium on User interface software and technology* (2014), ACM, pp. 637–644.
- [15] JONES, B. R., BENKO, H., OFEK, E., AND WILSON, A. D. Illu-miroom: peripheral projected illusions for interactive experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013), ACM, pp. 869–878.
- [16] KOUNAVIS, C. D., KASIMATI, A. E., AND ZAMANI, E. D. Enhancing the tourism experience through mobile augmented reality: Challenges and prospects. *International Journal of Engineering Business Management* 4 (2012), 10.
- [17] KREYLOS, O. Augmented reality sandbox. *Recuperado de <http://idav.ucdavis.edu/~okreylos/ResDev/SARndbox>* (2016).
- [18] MARESCAUX, J., RUBINO, F., ARENAS, M., MUTTER, D., AND SOLER, L. Augmented-reality-assisted laparoscopic adrenalectomy. *Jama* 292, 18 (2004), 2211–2215.

- [19] MILGRAM, P., AND KISHINO, F. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329.
- [20] PAUSCH, R., CREA, T., AND CONWAY, M. A literature survey for virtual environments: Military flight simulator visual systems and simulator sickness. *Presence: Teleoperators & Virtual Environments* 1, 3 (1992), 344–363.
- [21] PLODER, O., WAGNER, A., ENISLIDIS, G., AND EWERS, R. Computer-assisted intraoperative visualization of dental implants. augmented reality in medicine. *Der Radiologe* 35, 9 (1995), 569–572.
- [22] RASKAR, R., WELCH, G., CUTTS, M., LAKE, A., STESIN, L., AND FUCHS, H. The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques* (1998), ACM, pp. 179–188.
- [23] RASKAR, R., WELCH, G., AND FUCHS, H. Spatially augmented reality. In *First IEEE Workshop on Augmented Reality (IWAR'98)* (1998), Citeseer, pp. 11–20.
- [24] RASKAR, R., WELCH, G., LOW, K.-L., AND BANDYOPADHYAY, D. Shader lamps: Animating real objects with image-based illumination. In *Rendering Techniques 2001*. Springer, 2001, pp. 89–102.
- [25] ROO, J. S., AND HACHET, M. Interacting with spatial augmented reality.
- [26] STARNER, T., LEIBE, B., MINNEN, D., WESTYN, T., HURST, A., AND WEEKS, J. The perceptive workbench: Computer-vision-based gesture tracking, object tracking, and 3d reconstruction for augmented desks. *Machine Vision and Applications* 14, 1 (2003), 59–71.
- [27] UNDERKOFFLER, J. A view from the luminous room. *Personal Technologies* 1, 2 (1997), 49–59.
- [28] WILSON, A. D. Playanywhere: a compact interactive tabletop projection-vision system. In *Proceedings of the 18th annual ACM symposium on User interface software and technology* (2005), ACM, pp. 83–92.
- [29] WILSON, A. D. Depth-sensing video cameras for 3d tangible tabletop interaction. In *Horizontal Interactive Human-Computer Systems*,

2007. *TABLETOP'07. Second Annual IEEE International Workshop on* (2007), IEEE, pp. 201–204.