Capítulo 1 Introducción

1.1 Descripción del problema

En la actualidad las interfaces más comunes de interacción entre humanos y computadoras aún son el teclado y el mouse (Murthy & Jadon, 2009), pero la tendencia a corto plazo ha sido la pantalla táctil y el reconocimiento de gestos con movimientos realizados por el usuario (Bondre & Pimple, n.d.)(Ver figura 1). Los gestos son generados a partir de movimientos del cuerpo como brazo, manos, dedos, cabeza, cara o cuerpo (Mitra & Acharya, 2007).

Karam (Karam, 2006) reporta que las manos son las más usadas en comparación con otras partes del cuerpo para hacer gestos como parte natural del medio de comunicación entre humanos, tanto sentimientos como intensiones, por ello es que podrían ser lo más adecuado para la interacción natural con las computadoras también.

La investigación con referencia al reconocimiento de gestos se dirige principalmente a sistemas que puedan identificar gestos humanos específicos como entrada y procesarlos para control de dispositivos mapeándolos como comandos de salida. Las principales tecnologías son basadas en visión y contacto (Athavale & Deshmukh, 2015). Nosotros nos enfocaremos en los que son basados en visión artificial.

La capacidad de detectar gestos utilizando visión artificial y reconocimiento de patrones, permite explorar toda una gama de técnicas de interacción para controlar un ambiente, por ejemplo cambiar el volumen de la música o manipular la temperatura del termostato sin necesidad de acercarnos a él (Pu, Gupta, Gollakota, & Patel, 2013).

FIGURA 1 ROMPECABEZAS TÁCTIL

En los dispositivos basados en pantallas táctiles, para la detección de los gestos, primero es necesario detectar el comienzo del movimiento llamado, localización del gesto (Elmezain, Al-Hamadi, & Michaelis, 2009), el cual se reconoce al momento de hacer contacto con la superficie o con la parte sensible del dispositivo y termina al dejar de tocar el sensor de dicho dispositivo (ver figura 2).

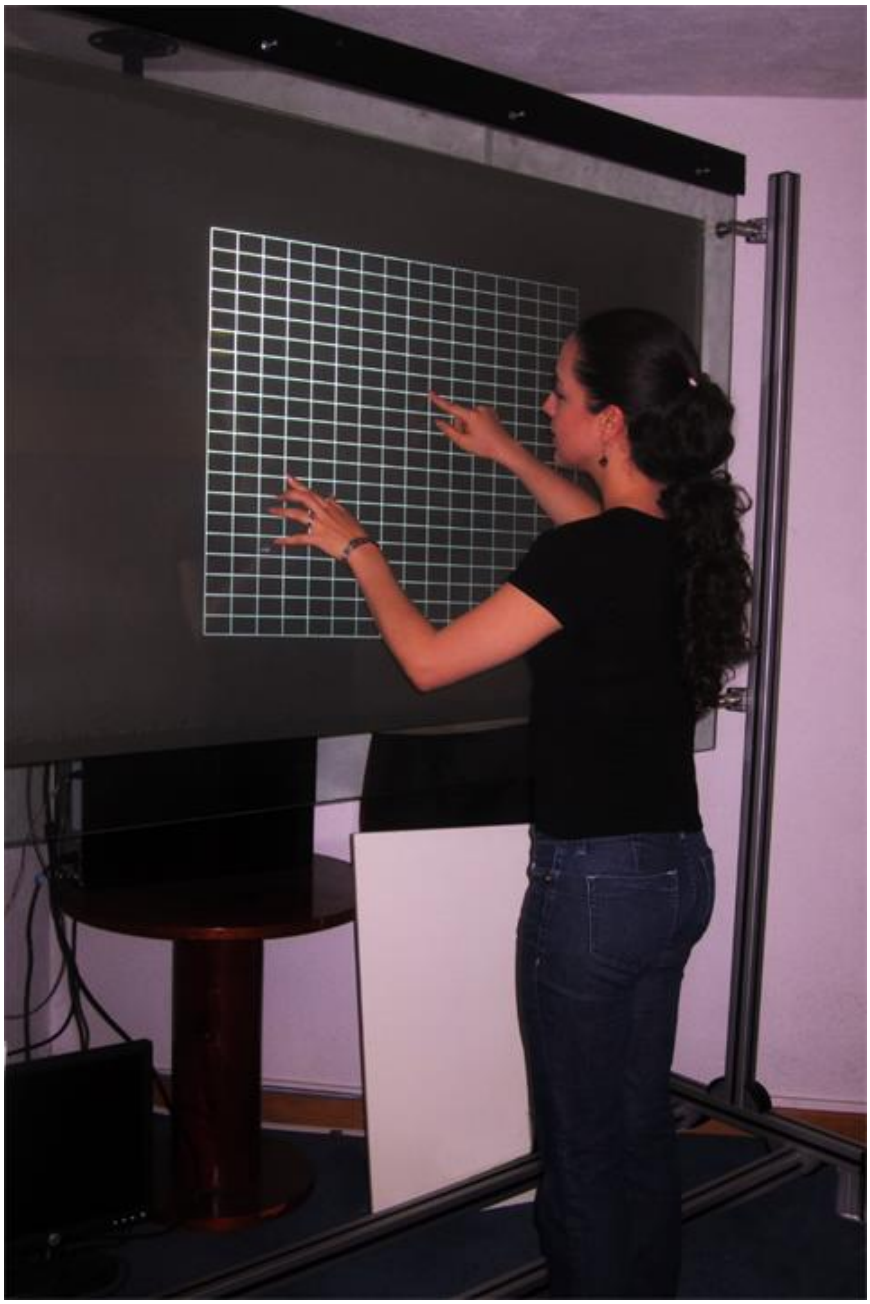

FIGURA 2 SISTEMA TÁCTIL

Al tener el registro del movimiento realizado sobre la superficie o sensor táctil, se analizan las secuencias registradas para verificar si concuerdan con alguna de las clasificaciones preestablecidas. Si el sistema concuerda con alguno de los gestos con el que el sistema responde,

se considera una acción por parte del usuario y el sistema dispara un evento en respuesta. Este tipo de retroalimentación no existe en sistemas no táctiles.

Desafortunadamente este tipo de interfaces, con esta capacidad de adaptación del sistema e interacción, tienen un alto costo. Además si tiene un uso considerado rudo, podría necesitar si no un reemplazo, un mantenimiento frecuente que sería caro. Lo cual lo hace poco accesible para considerarlo para uso de multitudes, sin mencionar lo poco higiénico y por lo mismo, inseguro. Utilizando gestos no táctiles, se permite la omisión de los periféricos de entrada de las GUI haciendo de la interacción un proceso ergonómico, higiénico y no invasivo (ver figura 3).

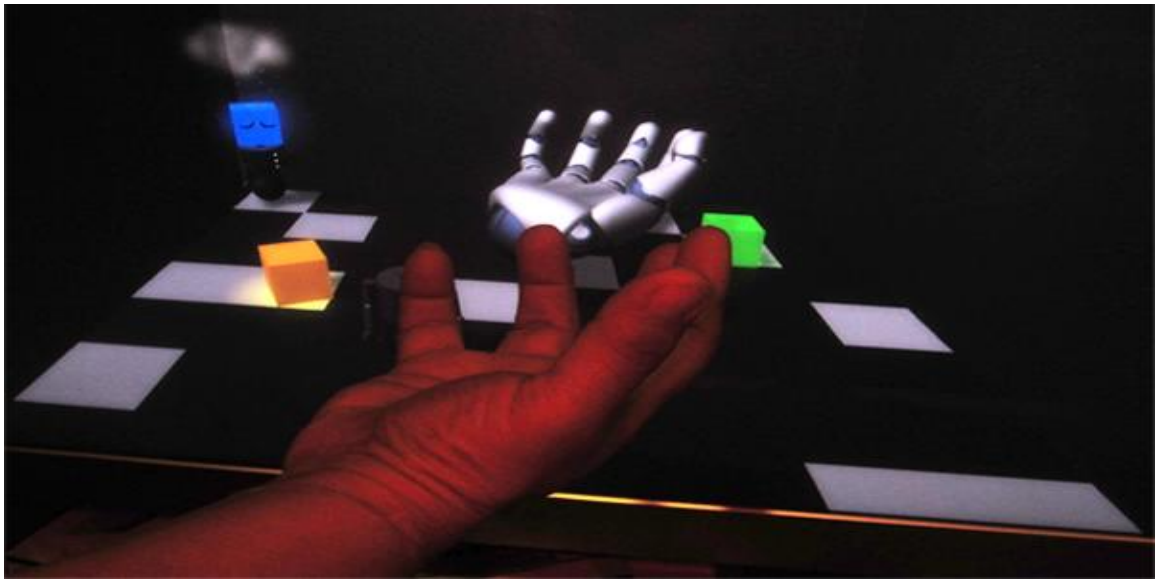

FIGURA 3 SISTEMA NO TÁCTIL

También existen métodos de interacción basados en sensores que se deben poner como accesorios o ropa (Ver figura 4).

FIGURA 4 CASCO PARA MANIPULACIÓN DE SISTEMA (INVASIVO)

Este tipo de interacción abre las puertas para controlar sistemas sin la necesidad de manipular dispositivos físicos y el usuario podría navegar y controlar un sistema de manera natural.

Muchos grupos de investigación se encuentran en la "carrera" por desarrollar el standard para reconocimiento de gestos. Existen alternativas de reconocimiento desde accesorios complejos, como trajes completos, a no invasivos como en el caso de cámaras de profundidad infrarrojas y el Kinect (Mitra & Acharya, 2007).

También existen métodos complejos como el uso de las señales de la red inalámbrica para detectar los movimientos corporales y reconocer gestos humanos (Pu et al., 2013). Mediante el análisis del movimiento por medio de una cámara web y una interfaz diseñada para la manipulación de un sistema de tareas simples de cómputo, se pondrían las nuevas tecnologías al alcance de todo el mundo. Utilizando visión por computadora sería una manera práctica de resolver el reconocimiento gestual (Chaudhary, Raheja, & Raheja, 2012).

1.1.2 Motivación

Cada vez son más frecuentes los dispositivos táctiles y los diseños orientados a gestos. Sin embargo la mayoría de estos sistemas son aún de un alto costo económico. Incrementar el alcance económico de estos sistemas podría agilizar y mejorar el tiempo de respuesta en tareas fáciles. Esto podría realizarse mediante la visión artificial debido a que carga una gran información sin ser intrusivo y de bajo costo (Murthy & Jadon, 2009).

La visión por computadora se refiere a la transformación de las imágenes digitales para tomar una decisión o una nueva representación (Bradski & Kaehler, 2011). Estas transformaciones son hechas con una meta. Los datos de entrada se encuentran en un contexto determinado, como reconocimiento del rostro de una persona o la distancia a la que se encuentre de otra cámara para determinar distancias. Las transformaciones pueden ser filtros que se encarguen de encontrar estadísticas en la imagen como la cantidad de determinado número de pixeles en la imagen completa, reducir la cantidad de colores, o determinar los histogramas de esos mismos pixeles.

En el presente trabajo, la meta y el contexto serán la interpretación de los gestos de una persona, capturando desde una cámara estática. Se traducirá e interpretará el movimiento obtenido al transformar las imágenes que se reciban en el dispositivo con los cambios que tenga con respecto al tiempo. Tratando de determinar si el movimiento registrado en las imágenes corresponde con un gesto determinado por un usuario y clasificando en tiempo real la captura.

El problema de la visión artificial es principalmente la reinterpretación de la señal capturada en video. Mientras que nosotros vemos una imagen con objetos que reconocemos fácilmente, la cámara y el sistema sólo están interpretando números que representan pixeles en un arreglo bidimensional (ver figura 5).

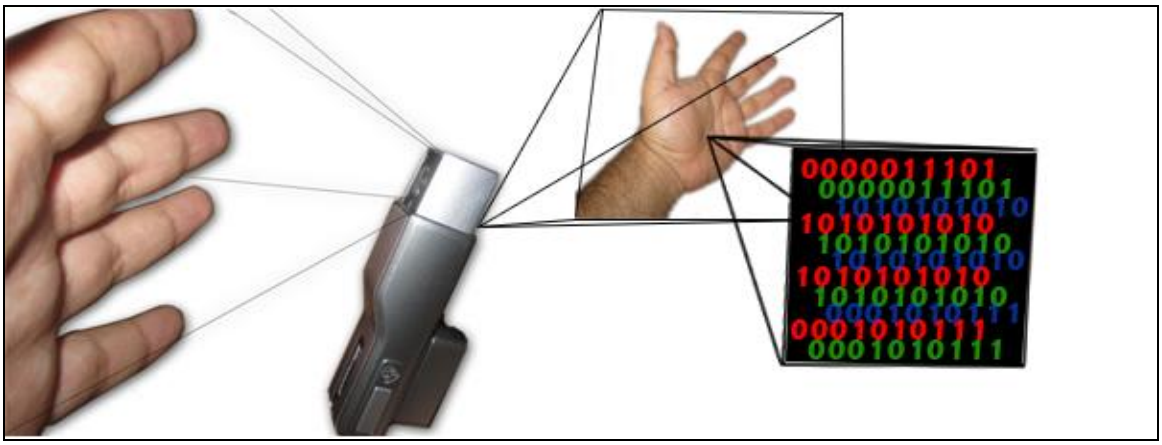

FIGURA 5 CAPTURA DE IMAGEN

Los gestos han sido considerados desde mucho tiempo una técnica de interacción que potencialmente puede resultar en una manera más natural, creativa e intuitiva como método de comunicación con las computadoras. El uso de los gestos de la mano como interfaz natural sirve como fuerza motivadora para la investigación en taxonomías de gestos. Los actos gestuales actúan como medio de comunicación no vocal en conjunto con o sin comunicación verbal. También pueden realizarse con cualquier parte del cuerpo combinado con varios de ellos. (Athavale & Deshmukh, 2015)

Por ejemplo, donde un alto número de personas necesitan consultas simples, podrían aprovechar un sistema rápido de navegación o de compra, que al ser de manejo gestual no táctil podría resultar de bajo costo de mantenimiento y por lo tanto durar más. Un ejemplo más concreto sería automatizar el sistema de compra de boleto de transporte masivo. Si el usuario ya conoce el tipo de boleto que necesita con las características que necesita, un sistema ágil de pocas opciones podría ser productivo si se lograra economizar el sistema para dispensar los boletos en lugar de tener que hacer filas.

Actualmente los gestos se reconocen mediante sistemas computacionales complejos o sensores especiales o invasivos poco prácticos, los cuales encarecen un sistema de navegación que a pesar de ser muy útil, el costo limita su replicación (Mitra & Acharya, 2007).

Generalmente los sistemas actuales pueden implicar varias de las siguientes características;

- Sensores especiales.
- Alto poder de rendimiento por parte del hardware.
- Alto costo de mantenimiento.
- Métodos invasivos (ver figura 6).
- Procesado de imagen basados en pixel (alto poder de hardware y algoritmos complejos).
- Clasificación estadística o métodos de reconocimiento de patrones.
- Entrenamiento y una base de datos con los gestos deseados.

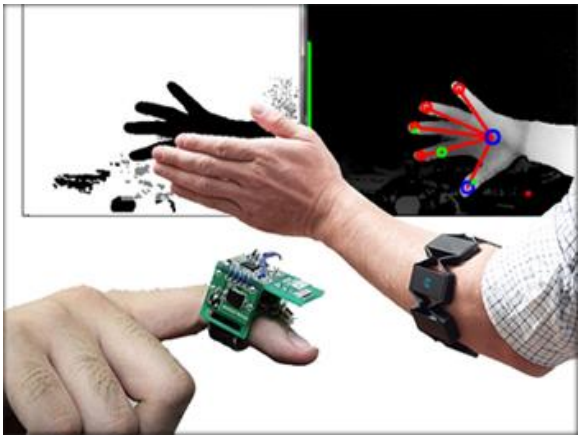

FIGURA 6 DIFERENTES SISTEMAS DE GESTOS

1.1.3 Reconocimiento de gestos

De acuerdo con Mitra (Mitra & Acharya, 2007) el reconocimiento de un gesto, es el proceso en donde un usuario hace un gesto y el receptor lo reconoce. Con esta técnica, podemos interactuar con máquinas y mandar mensajes particulares de acuerdo con el ambiente y sintaxis de la aplicación. Para utilizar visión artificial en el reconocimiento de gestos, es necesario el uso de procesado de imagen y extracción de características. La mayoría de los sistemas basados en visión son compuestos de tres etapas: detección, seguimiento y reconocimiento (Zabulis, Baltzakis, & Argyros, 2009).

Según Singh, (SINGH, 2012) existen dos tipos de interacción basada en gestos no táctiles, Directa e indirecta. Donde la manipulación de un objeto se refiere a que el objeto debe ser seleccionado antes de ser manipulado.

Singh dice que la manipulación directa se refiere a lo que en ingeniería se conoce como *control*, donde una variable manipulada influencia a una variable controlada, la cual es retroalimentada para proveer información que puede usar para reajustar la variable manipulada. Como podría ser un puntero dibujado que nos indica su posición y movimiento que al momento de mover la mano va respondiendo a la vez con nuestro movimiento y el usuario puede ir dirigiendo con precisión mediante la retroalimentación de la animación del icono.

La manipulación indirecta se refiere a cualquier tipo de interacción que se interpreta como lo sería el habla; una declaración, pregunta o comandos en general. Estos son ejemplos de actos de unidades mínimas de comunicación. De acuerdo con el paradigma de la manipulación, el usuario puede utilizar gestos con las manos para manipulación de sistemas. Los cuales permiten tener las siguientes características;

- Interacción natural, los cuales son de fácil uso.
- Concisos y poderosos porque pueden contener tanto la orden como los parámetros
- La mano se convierte en el mismo dispositivo eliminando transductores intermedios.

Los gestos dinámicos intencionales para comunicación son llamados gestos dinámicos conscientes. Por otro lado los gestos realizados sin intensión se llaman gestos dinámicos inconscientes. De acuerdo con Athavale y Deshmukh (Athavale & Deshmukh, 2015) La comunicación humana es 35% verbal y 65% no verbal basada en gestos. De los gestos dinámicos se pueden encontrar los manipulativos y comunicativos, de los comunicativos se encuentran los de acción y simbólicos. Los gestos de acción se pueden subdividir en *mimicos* y *deicticos*.

1.1.4 Detección del movimiento del gesto

El primer paso para el reconocimiento de gestos es la detección de las manos y la segmentación de la región de interés de la imagen. Este proceso es crucial para eliminar la información de fondo irrelevante y poder hacer un seguimiento de acuerdo a la secuencia. Varias características visuales han sido tomadas en consideración en diferentes métodos como color forma, movimiento o modelos (Zabulis et al., 2009) (ver figura 7).

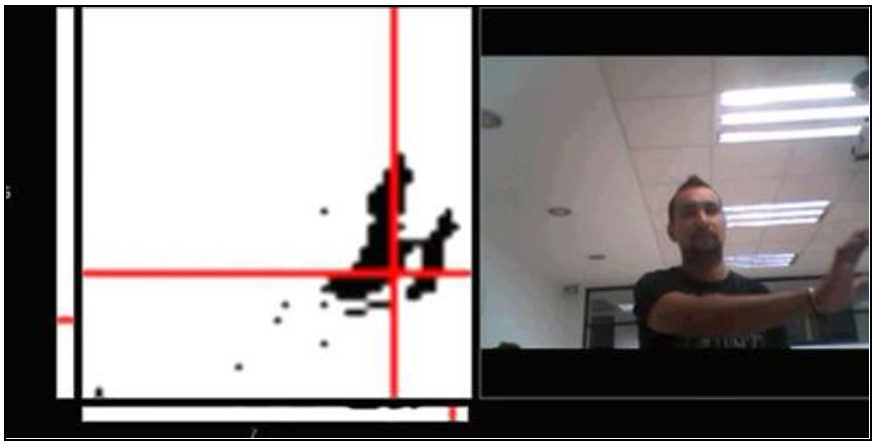

FIGURA 7 DETECCIÓN DE MOVIMIENTO

1.1.5 Rastreo del movimiento

El seguimiento es el siguiente paso en el reconocimiento de gestos. Este paso se refiere a la correspondencia de cuadro a cuadro del objeto en movimiento. Este seguimiento nos formara la trayectoria que deberá ser clasificada para determinar el gesto detectado. Puede ser utilizado de manera directa o analizado después de terminar el gesto. (Zabulis et al., 2009) (Ver figura 8).

FIGURA 8 DETECCIÓN Y RASTREO DEL MOVIMIENTO

1.1.6 Reconocimiento

La meta de un sistema de reconocimiento de gestos; es la interpretación de la semántica de la posición, postura y movimientos de la mano. Básicamente existen dos tipos de interacción de gestos con la computadora. El primero es control de aplicaciones basados en el movimiento en tiempo real. El segundo tipo es el que involucra el reconocimiento de la postura de la mano, señas o gestos y el reconocimiento de los mismos tendrán una dependencia directa con la aplicación en la que se use (Zabulis et al., 2009).

1.2 Propósito

Debido a la complejidad de los sistemas actuales para detectar y clasificar gestos, así como de la especialización de los sensores necesarios, este proyecto de tesis se propone investigar y diseñar un prototipo de sistema que permita analizar y reconocer gestos básicos utilizando visión artificial con baja complejidad de análisis y que tenga su tiempo de respuesta en tiempo real. Este reconocimiento puede ser útil en sistemas de tareas simples donde tres comandos puedan ser suficientes para cumplir con su objetivo.

1.2.1 Posibles aplicaciones

A continuación se describen algunas posibles aplicaciones de los sistemas basados en gestos.

- Algunas posibles aplicaciones serían sistemas de ventas de boletos en el metro de la ciudad de México. Lo conveniente de este tipo de aplicación podría observarse en la cantidad de gente que estaría usándolo sin tener que tocar el dispensador hasta recoger el boleto.

- Otro ejemplo de una aplicación sería en el uso de un sistema para asistir en un consultorio dental donde las manos del dentista no deben tocar nada mientras se encuentran trabajando con el paciente y poder hacer gestos a distancia para la manipulación del sistema.
- El uso general de interacción con aplicaciones en ambientes virtuales, en museos o casetas de información donde no se requiera de un uso extenso o de movimientos complejos para hacer funcionar el sistema, además al tener pocas opciones iniciales puede favorecer al entendimiento y uso del sistema.
- Juegos virtuales, teniendo reconocimiento de gestos mediante visión artificial en tiempo real permite interacción suficientemente hábil para el uso en videojuegos.
- Logrando un sistema robusto de reconocimiento de gestos, también sería posible tener en un dispositivo móvil con el sistema e interactuar o comunicarse más naturalmente con personas con problemas de habla mediante lenguaje de señas.

1.2.2 Aportaciones

Las posibles aportaciones del presente trabajo pretenden desarrollar en el campo del análisis del movimiento, posiblemente regresando a proyectos y soluciones ya encontradas y mejorarlas mediante la inclusión del análisis visto desde el punto de vista de distancia, posición y velocidad en visión artificial en las soluciones donde fue excluido, particularmente los que se alejaron conscientemente del uso de la diferencia de cuadros.

Especulando económicamente, si se usaran para instalarse diez sensores de reconocimiento de gestos en 195 estaciones de metro el costo utilizando un dispositivo Kinect sería de USD292,480.5 en cambio utilizando una cámara web la misma cantidad en sensores sería de

USD23,400. De ser así y manejar estos costos, el tiempo de espera de cada persona que conoce su destino o la cantidad de boletos que requiere al momento de acercarse a conseguir sus boletos podría reducirse considerablemente. Tampoco sería necesario una caseta de venta de boletos con una persona encargada de la transacción e incluso la conveniencia de la higiene en un uso tan recurrente podría ser también de alta relevancia.

La misma idea de utilizar una cámara web podría facilitar al momento de hacer la instalación o reparación de un lector, una tarea mucho más simple de sustituir o calibrar.

1.3 Objetivos

Se diseñara un método para reconocer patrones en tiempo real pertenecientes a gestos analizando el movimiento generado por el usuario utilizando análisis de imagen y clasificación de patrones. El método se implementará en un prototipo para analizar el rendimiento del método. El usuario moverá la mano dentro del rango de visión de la cámara y el sistema se encargará de la detección y clasificación del movimiento.

1.3.1 Objetivos Generales

Diseñar un método de reconocimiento de gestos basado en reconocimiento de patrones sin la necesidad de entrenamiento para detectar los gestos, de baja complejidad de cómputo y de respuesta en tiempo real.

1.3.2 Objetivos Específicos

- Estudiar y analizar las tecnologías existentes (lenguajes, sensores, métodos, etc.) para el desarrollo de aplicaciones basadas en gestos.
- Analizar los requerimientos generales (disponibilidad, desempeño, etc.) de métodos para el análisis de los gestos.
- Estudiar y analizar las técnicas de reconocimiento de patrones utilizados en clasificación de gestos.
- Seleccionar prototipos y algoritmos para proponer un método de tiempo real.
- Propuesta de arquitectura y solución.
- Diseñar un sistema prototipo que utilice el método propuesto para evaluar el desempeño.
- Interpretar el movimiento gestual de una mano mediante un lenguaje de programación.
- Evaluar el desempeño del sistema mediante pruebas con usuarios de diferentes perfiles.
- Documentación de las pruebas y correcciones de los métodos probados.
- Definición de trabajo a futuro

1.4 Alcances y Limitaciones

A continuación se describen los alcances y limitaciones del proyecto y de los objetivos planteados en el proyecto.

1.4.1 Alcances

La aportación de este proyecto serán los resultados obtenidos de las observaciones del comportamiento de las diferentes técnicas analizadas para diseñar el método de reconocimiento. También, el diseño e implementación del sistema prototipo para el análisis del método y propuestas de análisis para mejoras del mismo.

- La activación de los gestos será de manera automática.
- Los gestos serán reconocidos utilizando visión artificial.
- El análisis del movimiento será mediante una cámara web.
- No se utilizará hardware especializado para la detección del movimiento o la profundidad.
- Se implementarán tres gestos; derecha, izquierda y arriba (Enter).
- El evento se generará automáticamente al reconocer el gesto.
- La cámara será de baja resolución (640x480) o menos.
- El prototipo se implementara en ambiente Windows.
- La parte que reconozca el gesto funcionará reconociendo el movimiento.
- El método buscará la menor cantidad de pasos posibles para hacer el reconocimiento.

1.4.2 Limitaciones

Cabe mencionar que el análisis, metodología propuesta y prototipo, no pretenden en ningún momento llegar a nivel de producto de venta o competencia completa con productos existentes en el mercado.

- El reconocimiento en el prototipo será para una sola mano.
- No se buscará eliminar el ruido de la señal de video (movimiento de fondo)
- Se necesitará iluminación frontal.
- El reconocimiento será para un alcance de un metro a la cámara.
- El prototipo reconocerá una persona por vez.
- Las pruebas serán en ambientes controlados.

1.4.3 Implicaciones generales

La segmentación correcta de la mano para reconocer el gesto y seguir el movimiento sigue siendo un reto, debido a las variaciones espacio-temporal. Errores en el proceso del seguimiento causan fallas de la estimación del movimiento desviándola de la trayectoria real (Bhuyan, Ajay Kumar, MacDorman, & Iwahori, 2014).

En la figura 9 podemos ver la implicación básica de analizar video en tiempo real:

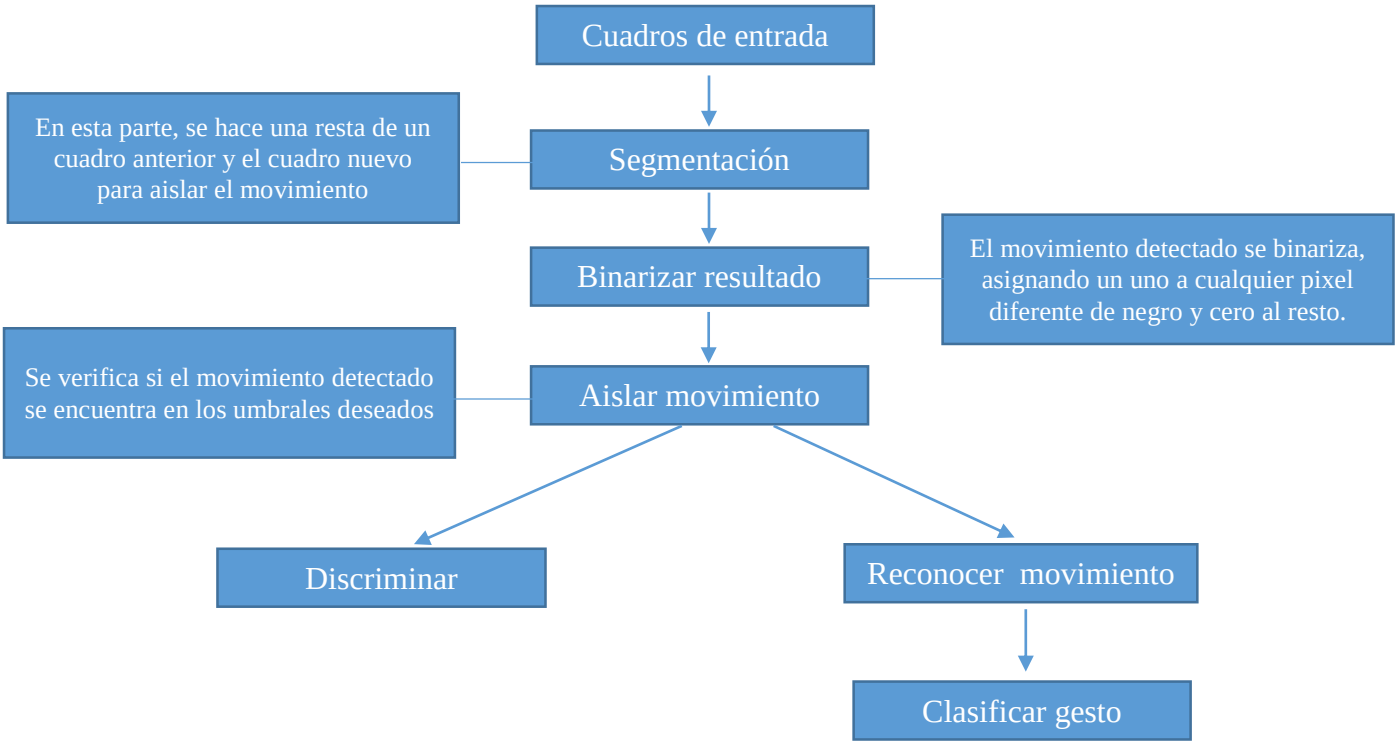

FIGURA 9 IMPLICACIONES DE ANÁLISIS DE VIDEO

Todos los módulos se implementarán en el lenguaje de C# debido al balance que posee entre velocidad de procesado y versatilidad de desarrollo, así como su robustez en librerías para manejo de procesado de imagen. La recomendación final sería portar la solución a un lenguaje de bajo nivel como C++ que nos permite sobre todo utilizar desarrollo sobre GPU, lo cual agiliza las operaciones sobre matrices, además que al probar que el método tenga funcionalidad sobre un lenguaje sobre C# casi nos asegura que su total estabilidad y velocidad de desempeño al portarlo a un lenguaje así.

1.5 Software y Hardware a utilizar

A continuación se describen las propiedades tanto de software como hardware para montar y desarrollar pruebas y prototipos.

1.5.1 Software

Para el desarrollo del prototipo propuesto como método, se utilizó el siguiente software. Se utilizó este software por ser las características básicas disponibles en el momento de realizar este trabajo para probar el comportamiento final del método implementado en esta plataforma.

- Visual Studio 2010
- Windows 7 Pro 32 bits
- Microsoft Visual Studio 2010

Además se utilizará una librería de uso libre porque está desarrollada completamente en .net. Con esta librería no es necesario instalar OpenCV en las máquinas para estar desarrollando y probando, además de no tener que hacer configuraciones para tener compatibilidad entre sistemas de 32 y 64 bits. Aun así la librería es compatible con los xml de características ya entrenadas disponibles en OpenCV.

1.5.2 Hardware

- Para el desarrollo de este proyecto se utilizó una computadora y una cámara web con las siguientes características:

- Computadora: Procesador Intel Core vPro i5 a 2.67 GHz, 4 GB de RAM y Disco duro de 150 GB
- Webcam Microsoft NX-6000 con video de alta definición (2.0 mega píxeles)

Se debe mencionar que se buscó una cámara común disponible con la menor capacidad de video compatible al momento de hacer el proyecto para tomarlo como base mínima de requerimientos para probar el método propuesto y sistema con el hardware también mínimo disponible para hacer tanto el diseño como las pruebas del proyecto completo.

En general, podemos ver la tendencia actual a sistemas basados en gestos o el uso del reconocimiento de gestos como complemento de los actuales métodos de interacción. Describimos las partes básicas de un sistema visual para reconocimiento de gestos y sus implicaciones. Si se logra desarrollar un método eficaz que permita el uso de gestos sin requerimientos especializados de hardware como sensores especiales o equipos de cómputo con capacidades por encima de la media. Podremos llegar a satisfacer los propósitos del presente trabajo y plantear a futuro el rebasar las condiciones que nos limitan para poder competir con las actuales soluciones existentes tanto comerciales como en investigación.