

Capítulo 2 Estado del arte

Existen métodos de reconocimiento de gestos tanto comerciales con sensores especializados como métodos aún en desarrollo e investigación académica. En este análisis me enfocaré principalmente en los métodos que son no invasivos, es decir, que el usuario no requiere equiparse con artículos especiales para el seguimiento o reconocimiento y segmentación.

2.1 Soluciones Comerciales Existentes

A continuación se presenta en la Tabla 1 una comparación de costos de los dispositivos no invasivos más comunes utilizados para detección de gestos. Con el fin de demostrar la razón por la cual estaremos enfocados en visión artificial.

Además, haciendo una comparación con los requerimientos del sistema de cada uno de los dispositivos, se escogió una cámara que se encontrara disponible y compatible con los sistemas actuales pero lo suficientemente básica para demostrar que el sistema podría funcionar de manera deseada con muy pocos requerimientos de procesado y de hardware en general.

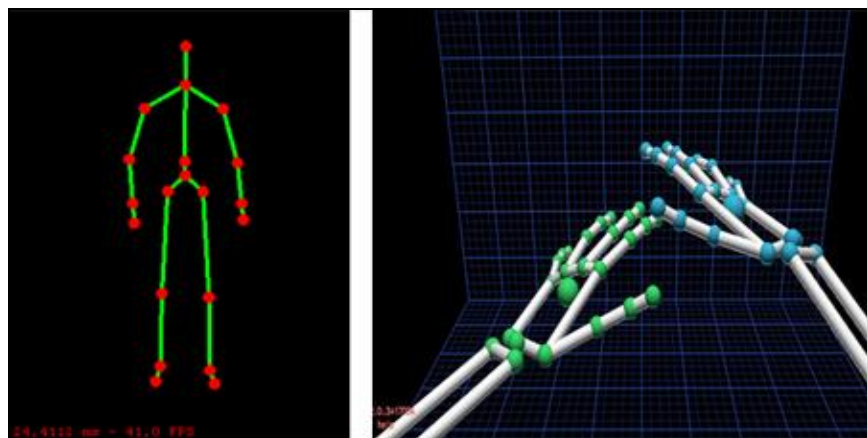


FIGURA 10 ESQUELETO KINECT (IZQUIERDA) ESQUELETO LEAP MOTION (DERECHA)

TABLA 1 COMPARACIÓN DE COSTO – VIDA DE DISPOSITIVOS Y REQUERIMIENTOS.

DISPOSITIVO	COSTO	GARANTÍA	Referencia	Hardware
Kinect	USD149.99	1 AÑO	(Microsoft, 2015)	<ul style="list-style-type: none"> •64-bit (x64) processor •Dual-core 3.2 GHz or faster processor •Dedicated USB 3.0 bus •2 GB RAM •A Kinect for Windows v2 sensor
Leap motion	USD79.99	1 AÑO	(Leap Motion, 2015)	<ul style="list-style-type: none"> • AMD Phenom™ II or Intel® Core™ i3 / i5 / i7 Processor • 2 GB RAM • USB 2.0 port
Miscrosoft LifeCam NX-6000	USD12.0	1 AÑO	(Newegg, 2015; Amazon, 2015; (Microsoft Inc., 2010))	<ul style="list-style-type: none"> •Intel Pentium® III 700 MHz (Intel Pentium 4 2.4 GHz recommended) •256 MB RAM (256 MB RAM recommended) •2 MB or higher video memory •USB 1.1 (USB 2.0 recommended).

Si tuviéramos que dar mantenimiento cada que vence la garantía, suponiendo que cada dispositivo caduca en ese tiempo. La diferencia en costos se vuelve notoria sobre todo si los sistemas son replicados para uso masivo. Suponiendo que necesitaríamos 20 estaciones de reconocimiento en una instalación (como un museo) el costo sólo de los sensores sería el siguiente.

TABLA 2 EJEMPLO DE PROYECCIÓN DE COSTOS

DISPOSITIVO	PRECIO	CANTIDAD	TOTAL
KINECT	USD 149.99	20	USD 2999.8
LEAP MOTION	USD 79.99	20	USD 1599.8
WEBCAM	USD 12.0	20	USD 240

Se ha encontrado mucha información sobre procesos para detección de gestos con múltiples metodologías, reconocimiento mediante análisis y reconocimiento de la piel, clasificadores estadísticos, y filtros de movimientos. El reconocimiento se puede dividir en diferentes etapas y diferentes tipos de clasificaciones. De las diferentes clasificaciones de gestos que existen estas pueden ser realizados con una o dos manos (Athavale & Deshmukh, 2015).

En este proyecto el foco será en los trabajos relacionados a visión artificial con reconocimiento de una sola mano sin la necesidad de equiparse con artículos extras como ropa, pulseras o guantes. Por esta razón la comparación de la cámara web básica con los dispositivos más comunes de reconocimiento de gestos disponibles en el mercado y su costo de funcionamiento así como su posible rendimiento o tiempo de vida.

2.2 Trabajo relacionado general

Existen diferentes métodos en estudio para la búsqueda del reconocimiento de gestos. La mayoría se dividen en dos tipos de gestos; estáticos y dinámicos (Athavale & Deshmukh, 2015). A pesar de que nos enfocamos en los gestos dinámicos por parte del usuario, mencionaremos algunos estudios basados en los gestos estáticos para tener una visión general de los diferentes métodos utilizados hasta el momento.

A continuación se resume el análisis de algunos de los métodos más usados en los últimos cinco años para reconocimiento de gestos de acuerdo con Athavale (Athavale & Deshmukh, 2015). Todos los métodos funcionan en tiempo real, buscando y detectando el color de la piel, lo cual los hace especialmente sensibles a la imagen dependiendo de los colores y la iluminación junto con la condición de que el sujeto tenga la mano descubierta para su análisis a nivel pixel (ver Tabla 3).

TABLA 3 MÉTODOS DE RECONOCIMIENTO DE GESTOS

Referencia Características	(Bhuyan et al., 2014)	(Kshirsagar, Sahu, Bankar, Moje, & Doye, 2015)	(Vitaly, Alexander, & Andrey, 2014)	(Rios-Soria, Schaeffer, & Garza-Villarreal, 2013)	(Lee, Sohn, Kim, Kim, & Kim, 2013)
Ventajas	Soporta gestos complejos	Fácil de implementar	Fácil de implementar	Fácil de implementar	Veloz y preciso para aplicaciones
Desventajas	No soporta gestos largos como el número 1	No es robusto	Sólo detecta gestos estáticos	Sólo detecta gestos estáticos	Costoso
Parámetros de clasificación del gesto	Vector de trayectoria	Vector de trayectoria	Huecos encontrados con Convex-hull	Huecos encontrados con Convex-hull	Vector de trayectoria
Técnica de extracción de características para análisis de movimiento	CRF ¹	FSM ²	Viola-Jones	Heurístico	SVM ³
Detecta rostro	Si, para eliminarlo y después buscar la mano	NO	NO	NO	SI
Principio de rastreo del movimiento	Movimiento y características locales	Estados	Detección de la mano, la piel y Lucas-Kanade	Detección de la mano y la piel	Kanade-Lucas-Tomasi
Tipo de reconocimiento del gesto	Clasificador	Comparación	Análisis de parámetros	Análisis de parámetros	Comparación
Requiere entrenamiento	SI	NO	SI	NO	SI
Analiza movimiento para gesto dinámico	SI	SI	NO	NO	SI
Tipo de fondo soportado	Simple	Complejo	Simple	Simple	Complejo
Porcentaje de éxito Reportado	96 %	65.702%	80%	93.17%	80%

¹ Conditional Random Fields² Finite State Machine³ Support Vector Machine

Los trabajos que se relacionan con el reconocimiento de gestos de la mano basados en visión artificial son variados y muchos grupos de investigación se encuentran actualmente mejorando respuestas de sus métodos. Como lo mencione anteriormente, en este caso, todos dependen directamente de la detección del color de la piel para poder hacer la segmentación o el rastreo o ambos procesos. Lo cual lo hace dependiente del color de la imagen en lugar de la forma si es que detectaran o rastrearán de este modo. Una posible desventaja es que para el reconocimiento, el usuario debe tener al menos las manos descubiertas para el análisis del color.

En la parte de la clasificación se encuentran más variantes como árboles de decisión, comparación de plantillas, estadística o clasificadores no lineales. El problema principal de este último es el entrenamiento que si bien no sólo es tardado también requiere de un repositorio grande de ejemplos para poder funcionar correctamente.

En el método propuesto, tratamos de hacer la segmentación y el rastreo del movimiento con el mismo método en un solo paso. Para la clasificación del gesto utilizamos heurísticas y la velocidad de respuesta es también en tiempo real. Además, con este método es posible hacer reconocimiento de gestos continuos sin utilizar un clasificador estadístico como los demás métodos.

2.3 Propuesta de Solución

Pocos enfoques toman en consideración el uso de la diferencia de cuadros (uso de movimiento) para determinar la detección de la mano y su segmentación. Esto es debido a que para funcionar se considera que el movimiento de la mano sería el único frente a la cámara para

poder ser detectado (Zabulis et al., 2009). A continuación se presentaran dos propuestas basadas en este mismo principio.

La primera propuesta buscará aislar el movimiento y utilizarlo también mediante diferencia de cuadros para rastrear además de segmentar. Aquí la diferencia estará en tratar de clasificar el movimiento de manera directa sin utilizar ningún otro tipo de referencia. Se intentará buscar patrones posibles en los gestos o diseñar la manera de reconocerlos a partir del movimiento que generan (Choi & Choi, 2013).

En la segunda propuesta también se utilizará la diferencia de cuadros para detectar la mano, segmentarla y hacer el rastreo. Los primeros dos pasos de los tres generales que son para la detección de gesto en un único proceso, la implementación se describirá más adelante en el capítulo 4.2. Para determinar la clasificación usaremos un método heurístico y aprovecharemos el rendimiento que ya existe para la detección rápida de rostros de Viola-Jones (Viola & Jones, 2004) para tener un punto de comparación a partir del cual se clasificará el movimiento registrado.

Los métodos existentes aún se encuentran en la parte del reconocimiento preestablecido de gestos, es necesario agilizar las tres partes principales para poder expandir el reconocimiento de los movimientos a los diferentes tipos de gestos tanto estáticos como dinámicos y que sean de fácil entrenamiento o aprendizaje para el sistema y para el manejo del usuario.