

ANTEPROYECTO DE TESIS ELABORADO POR

América Ivone Mendoza Morales

Y APROBADO POR EL SIGUIENTE COMITÉ

Dr. Vitaly Kober

Director del Comité

Dr.

Miembro del Comité

Dr.

Miembro del Comité

Dr.

Miembro del Comité

28 de agosto de 2014

**CENTRO DE INVESTIGACIÓN CIENTÍFICA Y DE
EDUCACIÓN SUPERIOR DE ENSENADA**



**PROGRAMA DE POSGRADO EN CIENCIAS
EN CIENCIAS DE LA COMPUTACIÓN**

**CONTROL DE COMPUTADORA BASADO EN GESTOS EN
CIRCUNSTANCIAS DE BAJA ILUMINACIÓN**

ANTEPROYECTO DE TESIS

que para cubrir parcialmente los requisitos necesarios para obtener el grado de

MAESTRO EN CIENCIAS

Presenta:

AMÉRICA IVONE MENDOZA MORALES

Ensenada, Baja California, México, agosto de 2014

RESUMEN del anteproyecto de tesis de **AMÉRICA IVONE MENDOZA MORALES**, presentado como requisito parcial para la obtención del grado de MAESTRO EN CIENCIAS en CIENCIAS DE LA COMPUTACIÓN . Ensenada, Baja California, agosto de 2014.

CONTROL DE COMPUTADORA BASADO EN GESTOS EN CIRCUNSTANCIAS DE BAJA ILUMINACIÓN

Resumen aprobado por:

Dr. Vitaly Kober

Director de Tesis

El reconocimiento de gestos con las manos ha sido un tema relevante en distintas áreas de las ciencias de la computación, por ejemplo en HCI es importante pues ayuda a crear una interacción natural entre la computadora y el usuario, por lo que se han desarrollado diversos métodos para encontrar el modelo que funcione en tiempo real y en diversas circunstancias. De manera que se pretende crear un modelo que fusione la información proporcionada por el dispositivo Kinect y haga el reconocimiento de gestos estáticos y dinámicos en tiempo real en circunstancias de baja iluminación y cuando existe oclusión. Dicho modelo será aplicado para crear un sistema que sirva como control de una computadora, osea que los gestos puedan ser utilizados como el cursor de esta.

Palabras Clave: Reconocimiento de gestos con las manos, Baja iluminación, Oclusion, Kinect

Contenido

	Página
Resumen en español	i
I. Introducción	1
II. Marco teórico	3
II.1 Reconocimiento de gestos con las manos	3
II.2 Métricas de distancia	8
II.3 Fusión de datos	8
III. Investigación previa relevante	10
III.1 Modelos de contacto	10
III.2 Modelos basados en la visión	10
III.3 Sistemas comerciales	11
IV. Importancia de la investigación	13
V. Limitaciones y suposiciones	14
VI. Contribución al conocimiento	15
VII. Objetivos	16
VII.1 Objetivo general	16
VII.2 Objetivos específicos	16
VII.3 Preguntas de investigación	17
VIII Metodología de la investigación	18
IX. Calendario de actividades	19

I. Introducción

La interacción del humano y la computadora (HCI, por sus siglas en inglés) es una rama de las ciencias computacionales dedicada al estudio, diseño e interacción de un humano con la computadora. El objetivo principal es que, para el humano la interacción entre ellos sea natural.

Los humanos para expresarnos utilizamos la escritura y el habla pero siempre se utiliza como complemento de la expresión los gestos corporales de cara, manos, etc. De manera que no es extraño que los investigadores de HCI se hayan interesado en los gestos corporales, en especial los gestos realizados con las manos, para crear un ambiente natural entre el usuario y la computadora. Pero para realizar esta interacción natural se necesita hacer el reconocimiento de los gestos, esto ha sido cada vez más sencillo gracias al avance de la tecnología, en especial con dispositivos de visión como distintos tipos de cámaras, y al crecimiento del procesamiento de las computadoras. Aunque existen diversos métodos para lograr el reconocimiento, no hay alguno que nos pueda dar un reconocimiento totalmente preciso en todas las situaciones que se presentan en el mundo real.

Es por eso que se propone crear un sistema que haga el reconocimiento de gestos con las manos, en situaciones donde existe baja iluminación y cuando tenemos oclusión causada por los dedos. El sistema se enfoca mayormente a solucionar los problemas de gestos con las manos que no tienen movimiento, y después se abordarán los gestos con las manos que involucran movimiento. El sistema aplicara los gestos como control de la computadora, esto con ayuda del dispositivo Kinect como herramienta para capturar la información de entrada.

La propuesta se encuentra dividida por secciones, enseguida se muestra la distribución de cada una de estas: la segunda sección se concentra en el marco teórico, la tercera muestra la

investigación previa relevante, en la cuarta la importancia de la investigación, la quinta las limitaciones y suposiciones, la sexta contiene la contribución al conocimiento, la séptima los objetivos, la octava la metodología de la investigación y por último la novena sección muestra el calendario de actividades que se lleva acabo a lo largo de la investigación.

II. Marco teórico

En esta sección se encuentran la información necesaria para entrar en contexto con el tema de reconocimiento de gestos con las manos.

II.1 Reconocimiento de gestos con las manos

La definición de gestos (?) son movimientos del cuerpo expresivos y significativos que involucran a los dedos, manos, brazos, cabeza, cara o cuerpo con la intención de transmitir información relevante o de interactuar con el ambiente. De acuerdo con la literatura (?) los gestos con las manos se clasifican en estáticos y dinámicos, los primeros están definidos como la posición y orientación de la mano en el espacio manteniendo esta pose durante cierto tiempo, por ejemplo para hacer una señal de aventón, la diferencia de los gestos dinámicos es que no se mantiene la pose si no que hay movimiento de la mano, un ejemplo cuando mueves tu mano en señal de adiós. De aquí en adelante entiéndase el término gestos con las manos, como gestos.

El reconocimiento de gestos lo podemos dividir en tres fases ?), detección o segmentación; seguimiento; dependiendo si los gestos son dinámicos, por último la etapa final el reconocimiento del gesto.

Los primeros acercamientos para llevar acabo el reconocimiento de gestos fue usando modelos de contacto ?) y ?), como su nombre lo dice son los que están en contacto físico con el usuario para capturar la posición y el movimiento de la mano, por ejemplo existen los guantes de datos, marcadores de colores, acelerómetros y pantallas multi-touch, aunque estos no son tan aceptados pues entorpecen la naturalidad entre la interacción del humano

y la computadora. Posteriormente se desarrollaron los modelos basados en la visión, el cual utiliza cámaras para extraer la información necesaria para realizar el reconocimiento, desde webcams hasta algunas más sofisticadas por ejemplo Kinect.

A continuación se describirán cada uno de estos métodos, pero solo se tomara los que están basados en la visión, pues estos son mejores para interactuar entre el usuario y la computadora, ya que crean un ambiente natural, aunque estos tienen mayor complejidad para lograr la implementación.

Los métodos basados en la visión se pueden representar por dos modelos (?), los basados en 3D, da una descripción espacial en 3D de la mano, y los basados en apariencia, como su nombre lo dice se basan en la apariencia de la mano, se puede dividir en dos categorías, los estáticos (modelo de silueta, de contorno deformables) y de movimiento (de color y movimiento).

Etapas del reconocimiento de gestos

Enseguida se describen las etapas del reconocimiento de gestos (detección, seguimiento y reconocimiento), con los métodos para llevar cada una de estas.

Detección

En esta etapa se detecta y segmenta la información relevante de la imagen (la mano), con la del fondo, existen distintos métodos para obtener dichas características como la de color de la piel, forma, movimiento, entre otras que generalmente son combinaciones de alguna de estas, para obtener un mejor resultado. Enseguida se describe brevemente cada una de estas.

- Color de la piel: Se basa principalmente en escoger un espacio del color, es una organización de colores específica; como; RGB (rojo, verde, azul), RG (rojo, green), YCrCb

(brillo, la diferencia entre el brillo y el rojo, la diferencia entre el brillo y el azul), etc. La desventaja es que si el color de la piel es similar al fondo, la segmentación no es buena, para corregir esto supone que el fondo no se mueve con respecto a la cámara.

- Forma: Extrae el contorno de las imágenes, si se realiza correctamente se obtiene el contorno de la mano. Aunque si se toman las yemas de los dedos como características, estas pueden ser ocluidas por el resto de la mano, una posible solución es usar más de una cámara.
- Valor de píxeles: Usar imágenes en tonos de gris para detectar la mano en base a la apariencia y textura, esto se logra entrenando un clasificador con un conjunto de imágenes.
- Modelo 3D: Depende de cual modelo se utilice, son las características de la mano requeridas.
- Movimiento: Generalmente esta se usa con otras formas de detección ya que para utilizarse por sí sola hay que asumir que el único objeto con movimiento es la mano.

Seguimiento

Consiste en localizar la mano en cada cuadro (imagen). Se lleva a cabo usando los métodos de detección si estos son lo suficientemente rápidos para detectar la mano cuadro por cuadro. Se explica brevemente los métodos para llevar a cabo el seguimiento.

- Basado en plantillas: Este se divide en dos categorías (Características basadas en su correlación y basadas en contorno), que son similares a los métodos de detección, aunque supone que las imágenes son adquiridas con la frecuencia suficiente para llevar a cabo el seguimiento. Características basadas en su correlación, sigue las características a

través de cada cuadro, se asume que las características aparecen en mismo vecindario. Basadas en contorno, se basa en contornos deformables, consiste en colocar el contorno cerca de la región de interés e ir deformando este hasta encontrar la mano.

- Estimación óptima: Consiste en usar filtros Kalman, un conjunto de ecuaciones matemáticas que proporciona una forma eficiente computacionalmente y recursiva de estimar el estado de un proceso, de una manera que minimiza la media de un error cuadrático, el filtro soporta estimaciones del pasado, presente y futuros estados, y puede hacerlo incluso cuando la naturaleza precisa del modelo del sistema es desconocida; para hacer la detección de características en la trayectoria.
- Filtrado de partículas: Un método de estimación del estado de un sistema que cambia a lo largo del tiempo, este se compone de un conjunto de partículas (muestras) con pesos asignados, las partículas son estados posibles del proceso. Es utilizado cuando no se distingue bien la mano en la imagen. Por medio de partículas localiza la mano la desventaja es que se requieren demasiadas partículas, y el seguimiento se vuelve imposible.
- Camshift: Busca el objetivo, en este caso la mano, encuentra el patrón de distribución mas similar en una secuencia de imágenes, la distribución puede basada en el color.

Reconocimiento

Es la clasificación del gesto, la etapa final del reconocimiento, la clasificación se puede hacer dependiendo del gesto. Para gestos estáticos basta con usar algún clasificador o empatar el gesto con una plantilla. En los dinámicos se requiere otro tipo de algoritmos de aprendizaje de máquina. A continuación se encuentran los principales métodos para llevar acabo el reconocimiento del gestos.

- K-medias: Consiste en determinar los k puntos llamados centros para minimizar el error de agrupamiento, que es la suma de las distancias de todo los puntos al centro de cada grupo. El algoritmo empieza localizando aleatoriamente k grupos en el espacio espectral. Cada píxel en la imagen de entrada es entonces asignadas al centro del grupo mas cercano
- K-vecinos cercanos (KNN): Este es un método para clasificar objetos basado en las muestras de entrenamiento en el espacio de características.
- Desplazamiento de medias: Es un método iterativo que encuentra el máximo en una función de densidad dada una muestra estadística de los datos.
- Máquinas de soporte vectorial (por sus siglas en inglés SVM). Consiste en un mapeo no lineal de los datos de entrada a un espacio de dimensión más grande, donde los datos pueden ser separados de forma lineal.
- Modelo oculto de Markov (por sus siglas en inglés HMM) es definido como un conjunto de estados donde un estado es el estado inicial, un conjunto de símbolos de salida y un conjunto de estados de transición. En el reconocimiento de gestos se puede caracterizar a los estados como un conjunto de las posiciones de la mano; las transiciones de los estados como la probabilidad de transición de cierta posición de la mano a otra; el símbolo de salida como una postura específica y la secuencia de los símbolos de salida como el gesto de la mano.
- Redes neuronales con retraso: Son una clase de redes neuronales artificiales que se enfocan en datos continuos, haciendo que el sistema sea adaptable para redes en línea y les da ventajas sobre aplicaciones en tiempo real.

II.2 Métricas de distancia

Existen diversas métricas, formas de medición, de distancia que nos sirven para comparar imágenes como: la distancia Euclidiana, Mahalanobis, Manhattan y Hausdorff. A continuación se mencionan las dos primeras ya que son utilizadas con mayor frecuencia.

Distancia Euclidiana

Sea X, Y dos imágenes, la distancia euclidiana se define como:

$$d(X, Y) = \sqrt{(X - Y)^t(X - Y)}$$

Distancia Mahalanobis

$$d(X, Y) = \sqrt{(X - Y)^t \Sigma^{-1} (X - Y)}$$

donde $\Sigma^{-1} = \sum_{k=1}^N \frac{(x_k - x_m)(y_k - y_m)}{N}$, x_m, y_m es la media.

II.3 Fusión de datos

La fusión de datos de multi-sensores se define (?) la teoría, técnicas y herramientas que son usadas para combinar datos del sensor, o datos provenientes de las lecturas del sensor, a un formato común. Se puede clasificar la fusión de multi-sensores de diferentes maneras como: aspecto del sistema se fusiona, debido a la configuración del sensor y a las características de entrada y salida. Enseguida se describe brevemente la clasificación dependiendo de que aspecto del sistema se fusiona.

Fusión a través del sensor

Se realiza con cierto número de sensores que miden la misma propiedad.

Fusión a través de atributos

Se lleva acabo con cierto número de sensores que miden diferentes cantidades asociadas a mismo experimento.

Fusión a través del dominio

Cierto número de sensores miden el mismo atributo sobre en diferentes rangos o dominios.

Fusión a través del tiempo

En este tipo de fusión las medidas recientes son fusionadas con información obtenida anteriormente.

III. Investigación previa relevante

La sección anterior explica los distintos enfoques para llevar acabo el reconocimiento de gestos, a continuación se encuentran los trabajos relevantes de cada uno de estos enfoques.

III.1 Modelos de contacto

(?) propone un modelo de mezclas adaptativo, usando un guante de datos, la principal limitante para este sistema es que solo reconoce gestos estáticos.

Aunque estos sistemas nos evitan algunos problemas que son consecuencia de los modelos basados en la visión, nos son perfectos, lo cual veremos enseguida.

Uno de los dispositivos recientes es MYO ¹, aunque de este se hablará en la ultima parte de esta sección.

Como se describió en la sección anterior en los modelos de contacto la principal limitante es el uso de dispositivos en el cuerpo para el reconocimiento de los gestos, por esta razón la mayoría de los sistemas para el reconocimiento estan enfocados en modelos basados en visión. Por lo que resulta natural que la investigación propuesta tome un enfoque basado en la visión.

III.2 Modelos basados en la visión

?) realizan un modelo de reconocimiento de gestos estático y dinámico basados en el algoritmo de Lucas-Kanade. Las principales ventajas de este método son que es invariante a rotación, escala y al fondo. Aunque el modelo es afectado por los cambios en la iluminación.

¹<https://www.thalmic.com/en/myo/>

(?), propone un método para calcular gestos estáticos y dinámicos usando los filtros Gabor y haciendo una estimación del ángulo en el que se encuentra la mano. Las principales ventajas son que el sistema funciona con cambios en la iluminación y es robusto a la rotación y escala. La desventaja es que el problema de oclusión no es tratado.

(?) hacen el seguimiento de la mano para identificar los gestos dinámicos usando los filtros adaptativos Kalman y el método Eigenhand. Con esta combinación obtienen un excelente resultado pues el sistema es robusto a la iluminación, cambio de pose, y a la oclusión causada la mano oculta por algún objeto en movimiento.

A pesar que la mayoría de los modelos vistos en la parte de arriba solucionan muchos de los problemas de los modelos basados en la visión. Ninguno de ellos puede resolver el problema de iluminación y oclusión, formada por lo dedos. Allí la importancia de la investigación propuesta, pues dará solución a estos inconvenientes al momento de reconocer los gestos.

III.3 Sistemas comerciales

Existen dispositivos como: Leap Motion ², MYO, y software, como Flutter ³, que realizan el reconocimiento de gestos, y estos los utilizan como reemplazo del ratón de la computadora.

Leap Motion es un dispositivo que detecta los movimientos de manos y dedos por medio de sensores infrarrojos. Este es robusto con el fondo, escala y rotación, pero no cuando existe oclusión pues cuando se realiza un zoom, como el que se hace en cualquier dispositivo touch, produce un error, y se presenta cuando un dedo es cubierto por otro, un problema grave es que tiene problemas de reconocimiento en circunstancias normales de luz.

MYO este dispositivo, solo se encuentra en pre-ordenamiento, detecta los impulsos eléctricos

²<https://www.leapmotion.com/>

³<https://flutterapp.com/>

de tus músculos mediante tres sensores, giroscopio, acelerómetro y magnetómetros. MYO es un brazalete que promete controlar aparte de la computadora, dispositivos tales como el celular o la tableta. La principal desventaja del sensor es que gestos involuntarios pueden producir acciones no deseadas.

Flutter es un software que reconoce cuatro gestos estáticos detectando la palma de la mano, usando la cámara web como dispositivo de entrada. Flutter permite controlar aplicaciones multimedia de la computadora. Las limitaciones del software son que solo reconoce gestos estáticos, realiza acciones no deseadas al hacer gestos involuntarios y no siempre reconoce los gestos.

Aunque estos dispositivos y software para reconocer gestos solucionan algunos problemas importantes en el área, sigue existiendo el problema de oclusión e iluminación. De allí la importancia que existan modelos que puedan resolver estos problemas se presentan frecuentemente en el reconocimiento.

IV. Importancia de la investigación

HCI, se dedica a que la interacción entre el usuario y la computadora sea natural y amigable. Y una manera de lograr u obtener un acercamiento a esta interacción es con el reconocimiento de gestos, en especial el de las manos. Aunque existen diversos métodos para realizar el reconocimiento, que funcionan en distintas circunstancias, fondo estático, fondo complejo, baja iluminación, oclusión del gesto.

Realidad aumentada, el mouse y el teclado no siempre pueden transmitir la información necesaria cuando se trata de obtener un modelo 3D en la computadora por eso es necesario encontrar una manera de transmitir esta información y que mejor que con el uso de un espacio 3D por lo que en este caso es de gran utilidad el reconocimiento de gestos con las manos.

V. Limitaciones y suposiciones

Los sistemas de reconocimiento de gestos trabajan con distintos fondos, en este caso el sistema propuesto trabajará con fondos estáticos.

Gran porcentaje de los trabajos realizados anteriormente para el reconocimiento de gestos basado en el modelo de la visión utilizan cámaras digitales o cámaras web, en la investigación se utilizará el dispositivo Kinect, para tomar la información de entrada (RGB, infrarroja y profundidad). De tal manera que las limitaciones del sistema propuesto están dadas por dicho dispositivo, tales como la distancia a la que se encontrara, la resolución de las imágenes, en especial a la baja resolución de la información infrarroja. También el sistema depende de dos Kinect, que se utilizarán en el caso que exista oclusión.

VI. Contribución al conocimiento

Se desarrollará un modelo que fusione 3 tipos de datos (RGB, infrarroja y de profundidad) de una imagen de entrada en tiempo real, con esta fusión se pretende dar solución a los problemas de baja iluminación y oclusión, que se presentan al hacer el reconocimiento de gestos.

Una aportación en las áreas de HCI, realidad aumentada y reconocimiento de patrones ya que con el modelo propuesto se podrá dar solución a algunas de las limitaciones importantes para el reconocimiento de gestos.

VII. Objetivos

VII.1 Objetivo general

Desarrollar un sistema que permita controlar la computadora haciendo uso de gestos con las manos, estáticos y dinámicos. El sistema debe ser robusto, funcionar en circunstancias de baja iluminación, cuando exista oclusión en gestos dinámicos de forma que el sistema debe reconocer el tipo de gesto que se utiliza y solucionar el problema con ayuda de dos dispositivos kinect.

VII.2 Objetivos específicos

1. Identificar los principales métodos de reconocimiento de gestos, estáticos y dinámicos cuando existe baja iluminación y en el caso de los gestos dinámicos cuando existe oclusión.
2. Proponer un algoritmo que reconozca gestos estáticos en situaciones de baja iluminación mediante la fusión de la información de entrada (RGB, IR y profundidad).
3. Proponer un modelo basado en los métodos de predicción para reconocer gestos dinámicos, este debe resolver los problemas de oclusión que causan los dedos.
4. Analizar el sistema diseñado, en cuanto a su eficiencia presentada en cuanto al reconocimiento de los gestos, tiempo de respuesta, en circunstancias de baja iluminación y oclusión.
5. Comparar los modelos propuestos con los existentes, en base al tiempo de respuesta y

la eficiencia en cuanto al reconocimiento del gesto.

VII.3 Preguntas de investigación

1. ¿Cuáles son los mejores modelos de reconocimiento de gestos estáticos y dinámicos?
2. ¿Cuáles son sus limitantes?
3. ¿Cuáles son los modelos de reconocimiento de gestos que obtienen mejores resultados en circunstancias de baja iluminación?
4. ¿Cuáles son los modelos de reconocimiento de gestos que obtienen mejores resultados cuando existe oclusión?

VIII. Metodología de la investigación

1. Realizar una revisión bibliográfica de los métodos existentes de reconocimiento de gestos e identificar principales métodos en base a los problemas que atacan.
2. Estudiar y comprender los modelos basados en visión usados para el reconocimiento de gestos, los cuales fusionan diferentes tipos de información de entrada.
3. Crear una base de datos con los gestos necesarios para el sistema.
4. Implementar los modelos relevantes de fusión, estudiados previamente.
5. Desarrollar un algoritmo de reconocimiento de gestos estáticos basado en los modelos relevantes que solucione las desventajas de estos.
6. Estudiar y comprender los modelos estadísticos de predicción de gestos dinámicos.
7. Implementar los modelos estadísticos relevantes.
8. Desarrollar un modelo que solucione el problema de oclusión, basado en los existentes.
9. Probar y evaluar los modelos desarrollados en circunstancias distintas de iluminación en base al tiempo de respuesta, a la precisión del reconocimiento del gesto.
10. Analizar y evaluar los modelos existentes conforme al tiempo de respuesta y la precisión del reconocimiento del gesto.

IX. Calendario de actividades

Actividades	Septiembre	Octubre	Noviembre	Diciembre	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto
Revisión bibliográfica.												
Estudiar y comprender los modelos de reconocimiento de gestos que se basan en fusión de datos y en modelos estadísticos.												
Construir una base de datos con los gestos a utilizar en el sistema.												
Implementar los modelos relevantes estudiados anteriormente.												
Primer avance de tesis.												
Diseñar e implementar los modelos propuestos.												
Entrega del resumen del artículo al congreso de SPIE.												
Segundo avance de tesis.												
Evaluar los modelos propuestos.												
Comparar el sistema con los ya existentes.												
Tercer avance de tesis.												
Escritura preliminar de tesis.												
Corrección de tesis.												
Defensa de tesis.												
Presentación de artículo en congreso SPIE.												

. Calendario de actividades del proyecto de investigación