

Una revisión del estado del arte sobre reconocimiento de Gestos de la Mano

Yessica Chuctaya Zamata
Ciencia de la computación
Universidad Católica San Pablo
Email:yessica.chuctaya@ucsp.edu.pe

Abstract—Desde tiempo muy antiguos el hombre se comunica de forma verbal y no verbal para esto uso partes del cuerpo como el rostro y mano. El estudio de reconocimiento de gestos parte de estos hechos y trasladar esta interacción hacia un computador. Los gestos tienen una amplias aplicaciones como en el áreas de Interacción Humano Computador (IHC) , robótica, computación gráfica y el de lenguaje de señas , la realidad aumentada basada en visión , etc.

En este documento revisaremos el reconocimiento de gestos de las manos y los sensores que se usan para capturar la información(sensores Inerciales , sensores de profundidad) posteriormente esta información sera reconocida como un gesto , en los documentos posteriores se proponen diferentes técnicas de clasificación de gestos (gestos estáticos ,gestos dinámicos).

I. INTRODUCCIÓN

Los gestos son una forma de comunicación no verbal desde hace mucho tiempo donde se utilizan varias partes del cuerpo , principalmente las manos y la cara , esta forma de comunicación muchas veces se convierte en el lenguaje principal de comunicación como es el lenguaje de sordo-mudos . Los gestos se utilizan ampliamente para diferentes aplicaciones como la interacción entre humanos y robots , el reconocimiento del lenguaje de señas ,la realidad aumentada basada en la visión , etc.Los gestos son complejos y adecuados para ambientes reales donde se utilizan mas partes del cuerpo como las articulaciones del cuerpo.

El reconocimiento de gestos de la mano ha sido utilizado con gran éxito en el uso de teclado y *mouse* pero se encontraban limitados al espacio 2D y el ambiente real nos lleva al espacio 3D . Hasta hace poco contar con herramientas de ayuda para este tipo de trabajos era complicado pero desde que se construyeron sensores capaces de capturar posiciones y rotaciones en ambientes reales ya no es tan complicado . Los sensores tienen dos enfoques : sensores para el reconocimiento de gestos basados en la visión y sensores basados en la inercia.

Los sensores basados en la inercia (intrusivos) son sensores incorporados al cuerpo capaces de capturar señales del cuerpo de forma precisa pero restringiendo la interacción , además del alto costo de equipamiento [7], dentro de estos sensores de encuentras los guantes .

Los sensores basados en la visión (no intrusivos) como las cámaras ópticas y sensores infrarrojos que proporcionan un

conjunto de datos basados en imágenes o profundidad de estos , sin embargo en este tipo de sensores son vulnerables cuando se captura la información en un ambiente que contiene ruido de fondo o la iluminación que obstaculiza la captura [10].

El problema de identificación de gestos con las manos en 3D se puede resolver en el dominio 2D , donde la extracción y comparación de características en 2D se utilizan para determinar el tipo de gestos[11].

Debido a que el reconocimiento de gestos de mano ha atraído cada vez más investigadores en los temas de visión por computador , reconocimiento de patrones y la interacción humano-computadora[2] se proponen técnicas para la clasificación de gestos donde se considera el proceso de reconocimiento de gestos.

El documento esta estructurado de la siguiente forma en la sección II se explica de forma general el proceso de reconocimiento de gestos y los sensores con los que se trabaja para capturar la información en espacio 3d , en la sección III los trabajos relacionados a la clasificación de gestos con los que se realizara la clasificación.

A. Motivación de la Investigación

La aplicación de reconocimiento de gestos de la mano es muy variado y parte de las áreas de investigación es interacción Humano computador y el de visión por computador donde se llevan acabo aplicaciones para terapias clínicas , para la interpretación de lenguaje de señas o videojuegos en un mundo virtual donde se debe considerar construir un ambiente muy similar al nuestro en 3 dimensiones y la captación de información debería ser de forma intuitiva, con el menor error posible y la captura de estos gestos se vera afectado por el dispositivo con el que se trabaje ya se con sensores pequeños acoplados a ropas(dispositivos),o sensores que no tienen ningún contacto con el usuario.

B. Contribuciones

En este estudio , nos enfocamos en realizar una revisión de investigaciones basados en el reconocimiento de gestos de la mano en 3D y los sensores ya trabajados en los diferentes documentos ya que conocerlos permitara trabajos futuras.

A continuación detallamos la contribución de esta investigación.

- Proponer criterios de clasificación , basados en los sensores que se usan para un trabajo a futuro y las técnicas para la clasificación de gestos.
- Poder entender algunas técnicas de clasificación para poder implementar en un futuro.
- Poder reconocer algunas limitaciones en el reconocimiento de gestos de la mano.

II. GESTOS DE LA MANO

El gesto de la mano se define como la combinación de todo tipo de posturas y movimientos que se producen con la mano, el gesto de la mano en general esta formado por gestos estáticos y dinámicos [2] , los gestos estáticos de la mano se refieren a la forma de la mano y los gestos dinámicos de la mano se componen de una serie de movimientos de la mano.

Además los gestos de la mano pueden estar modelado en dos modelos, temporales(dinámico) y espaciales, este ultimo se puede dividir en dos categorías , el modelo 2D y 3D.

A. Sensores

En cuanto a sensores se tiene dos enfoques :reconocimiento de gestos basados en visión (se realiza a través de sensores de profundidad usando técnicas de representación de características basadas en imágenes de profundidad 3D o esqueletos) y reconocimiento de gestos basados en inercia estos se colocan directa o indirectamente en el cuerpo humano y han sido incorporados en ropa, zapatos, relojes de pulsera, dispositivos móviles, etc. Estos sensores generan señales de aceleración y rotación correspondientes a una acción.[1]

- **Kinect:** es esencialmente un sensor de profundidad de luz estructurado, este sensor puede capturar imágenes de profundidad de 16 bits con una resolución de 320 240 píxeles, en imágenes de color a 640 480 píxeles con una resolución de 8 bits por canal. En una imagen de profundidad, el valor de cada píxel indica la distancia entre un punto de escena 3D y el sensor y devuelve un mapa de profundidad completo [1].
- **Leap Motion:** es un sensor de profundidad que devuelve un conjunto de puntos relevantes de la mano (dedos, centro de la mano) y algunas características de postura de mano es especialmente adecuado para el reconocimiento de gestos con las manos en 3D [9].

B. Arquitectura de seguimiento

Compuesto por la **adquisición de datos** se realiza a traves de los sensores mencionados anteriormente ejemplo de esto Kinect, Leap motion , cámaras de profundidad [5] .

El siguiente paso después de la adquisición de datos es el **modelamiento de gestos** depende principalmente de esta etapa el éxito del reconocimiento. Los diferentes

datos recibidos a través de los dispositivos ya están listos para ser modelados adecuadamente,dentro de esta etapa tenemos 3 pasos diferentes:Segmentación manual (se refiere a localizar las secuencias de manos en las imágenes) , eliminación de ruidos , detección de bordes o contornos [5].

Extracción de características En esta etapa se extraen características como la forma, la orientación, las texturas, el contorno, el movimiento, la distancia, el centro de gravedad, etc. para el reconocimiento de gestos de la mano se considera principalmente la siguientes características, geometría de la mano, como el contorno de mano, yemas de los dedos, detecciones de los dedos, no geométricas como el color , la silueta , la textura , en la extracción de características tenemos dos enfoques : el primer enfoque esta basado en el modelo cinemático ; en este ultimo se considera la postura de la palma y los ángulos de la articulación es adecuado en caso de interacción en tiempo real en un entorno virtuales, la dificultad se encuentra en el proceso de extracción de características es decir, bordes, porque las manos humanas tienen menos textura y no proporcionan bordes confiables internamente y el enfoque basado en la vista, en este enfoque la mano es modelado mediante una colección de imágenes de intensidad 2D.

Reconocimiento del gesto de la mano en esta etapa se ha selecciona un conjunto de datos adecuado, y los gestos se reconocen utilizando varios métodos como la coincidencia de plantillas, búsqueda de diccionarios, coincidencia estadística, coincidencia lingüística, red neuronal, etc ,

III. TRABAJOS RELACIONADOS

Gracias a los avances tecnológicos en los sensores comerciales se tienen varios enfoques diferentes de investigación como la que veremos en los siguientes artículos.

En el artículo [3] se propone trabajar con gestos dinámicos utilizando imágenes de profundidad , datos esqueléticos de la mano, primero realiza el reconocimiento de gestos dinámico basado en esqueleto mencionado en la propuesta de Quentin De Smedt, Hazem Wannous y Jean-Philippe Vandeborre, de la IMT Lille Douai / Universidad de Lille, Francia, donde un gesto dinámico se puede ver como una serie de tiempo del esqueleto de la mano que escribe el movimiento y las formas de las manos a lo largo del gesto . Finalmente clasifica la secuencia de fotogramas con una red neuronal convolucional.

En el artículo [4] se trabaja sobre datos esqueléticos de la mano proporcionados por el Kinect y clasificándola a través de la una nueva red neuronal llamada MC-DCNN que trabaja con las posiciones de las articulaciones de forma paralela .

En el artículo [8] se trabaja con el sensor Leap motion para captar los movimientos, y toma en consideración la muñeca como el centro y las posiciones de la punta de los dedos para detectar flexión de las manos , posteriormente en

la extracción de características, se incluyen características con distancia euclidiana entre los yemas de los dedos y el ángulo de inclinación para después clasificar utilizando Support Vector Machine (SVM) multiclase Y k-NN y el enfoque de vecinos mas cercanos.

En el artículo [9] se trabaja con dos sensores Leap motion y kinect para cada uno se propone algoritmos de extracción de características relevantes para el reconocimiento de gestos diferentes luego estos son depositados en vectores de características para aprovechar ambas tecnologías se concatena ambos vectores para luego ser clasificado, utilizando un clasificador SVM multiclase.

En el artículo [6] se trabaja con el sensor kinect para detectar imágenes de profundidad para luego segmentar y encontrar las distancias comunes entre la mano y el cuerpo, esto lo realiza a través del modelo hidden markov model (HMM) y the fuzzy neuronal network(FNN) para reconocer gestos dinámicos.

IV. TAXONOMÍA

La clasificación de gestos de la mano en 3D que se muestra a continuación esta dado bajo 3 criterios, Adquisición de datos, Reconocimiento de gestos de la mano.

Adquisición de datos	Reconocimiento
Cámara de profundidad	Gestos dinámicos, CNN [3]
Kinect	MC-DCNN - CNNm[4]
Leap motion	SVM and K-NN [8]
Leap motion y kinect	SVM [9]
Datos sintéticos	CNN [11]
Kinect	HMM y FNN [6]

Tabla 1: Clasificación de documentos con técnicas de clasificación de gestos y sensores.

A continuación una pequeña descripción de las técnicas para clasificación de gestos:

- Support Vector Machine (SVM) : es un métodos matemáticos clásicos y ampliamente utilizado para analizar datos y reconocer patrones, que generalmente se usan para la clasificación y el análisis de regresión.

V. CONCLUSIONES

El reconocimiento de gestos de la mano en 3D es un área de investigación vigente con la intención de reconocer un gesto como adecuado y poder realizar aplicaciones como terapias y video juegos en entornos 3D en ambientes virtual. La combinación de sensores podría ser aprovechada de una mejor forma para reconocer un gestos con mayor precisión y aceptada.

REFERENCES

[1] Chen Chen, Roozbeh Jafari, and Nasser Kehtarnavaz. A survey of depth and inertial sensor fusion for human action recognition. *Multimedia Tools and Applications*, 76(3):4405–4425, 2017.

[2] Hong Cheng, Lu Yang, and Zicheng Liu. Survey on 3d hand gesture recognition. *IEEE Trans. Circuits Syst. Video Techn.*, 26(9):1659–1673, 2016.

[3] Quentin De Smedt, Hazem Wannous, Jean-Philippe Vandeborre, Joris Guerri, Bertrand Le Saux, and David Filliat. Shrec'17 track: 3d hand gesture recognition using a depth and skeletal dataset. In *10th Eurographics Workshop on 3D Object Retrieval*, 2017.

[4] Guillaume Devineau, Fabien Moutarde, Wang Xi, and Jie Yang. Deep learning for hand gesture recognition on skeletal data. In *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on*, pages 106–113. IEEE, 2018.

[5] Edwin Escobedo and Guillermo Camara. A new approach for dynamic gesture recognition using skeleton trajectory representation and histograms of cumulative magnitudes. In *Graphics, Patterns and Images (SIBGRAPI), 2016 29th SIBGRAPI Conference on*, pages 209–216. IEEE, 2016.

[6] Xiao-Li Guo and Ting-Ting Yang. Gesture recognition based on hmm-fnn model using a kinect. *Journal on Multimodal User Interfaces*, 11(1):1–7, 2017.

[7] Pradeep Kumar, Rajkumar Saini, Santosh Kumar Behera, Debi Prosad Dogra, and Partha Pratim Roy. Real-time recognition of sign language gestures and air-writing using leap motion. In *Machine Vision Applications (MVA), 2017 Fifteenth IAPR International Conference on*, pages 157–160. IEEE, 2017.

[8] Wen-Jeng Li, Chia-Yeh Hsieh, Li-Fong Lin, and Woei-Chyn Chu. Hand gesture recognition for post-stroke rehabilitation using leap motion. In *Applied System Innovation (ICASI), 2017 International Conference on*, pages 386–388. IEEE, 2017.

[9] Giulio Marin, Fabio Dominio, and Pietro Zanuttigh. Hand gesture recognition with leap motion and kinect devices. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 1565–1569. IEEE, 2014.

[10] Zhou Ren, Junsong Yuan, Jingjing Meng, and Zhengyou Zhang. Robust part-based hand gesture recognition using kinect sensor. *IEEE transactions on multimedia*, 15(5):1110–1120, 2013.

[11] Chun-Jen Tsai, Yun-Wei Tsai, Song-Ling Hsu, and Ya-Chiu Wu. Synthetic training of deep cnn for 3d hand gesture identification. In *Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO), 2017 International Conference on*, pages 165–170. IEEE, 2017.