

Reconocimiento de Gestos para la Enseñanza de los Números a Párvulos Usando Sensor Kinect.

Sindy T. Fuentes Gutiérrez, Jhonatan E. Ríos Meza, Carlos A. Madrigal

Instituto Tecnológico Metropolitano
Medellín, Colombia

Keywords: Mapa de profundidad, Sensor Kinect, reconocimiento, enseñanza.

Resumen

Este artículo presenta una aplicación para la enseñanza de números a párvulos usando imágenes de profundidad obtenidas con el Sensor Kinect. La metodología propuesta consiste en localizar la mano y utilizar algoritmos de segmentación que permitan aislar la mano del niño del resto del cuerpo y de los objetos presentes en la escena, además son extraídas 14 características que fueron seleccionadas con la ayuda del toolbox Balu. Finalmente, mediante el clasificador LDA se reconocen de manera precisa y eficiente los ademanes realizados por los niños. En las pruebas realizadas se pudo evidenciar que el algoritmo de clasificación permite reconocer los gestos en un 95% de los casos. Como interface con el Kinect se utilizó la librería OpenNI y para el procesamiento de las imágenes se utiliza la librería OpenCV.

1 Introducción

La enseñanza a párvulos cada día se convierte en un reto más grande para educadores y padres, ya que los niños buscan activamente el conocimiento a través de sus interacciones con el entorno y se comportan como “pequeños científicos” que tratan de interpretar el mundo [1], a esto se le suma que recientemente no se han hecho avances significativos en cuanto a los mecanismos de aprendizaje. En particular, la enseñanza de los números se realiza comúnmente de forma mecánica y repetitiva, lo que puede provocar que el niño sepa los números de memoria durante un tiempo pero que difícilmente los recuerde y comprenda su significado. Esto demuestra la importancia de desarrollar estrategias de aprendizaje de los números que se ajusten a la forma en que los niños asimilan la información.

Hoy en día existen diferentes estrategias para la enseñanza a párvulos tales como libros, videos, canciones, material audiovisual y juegos didácticos. En internet se encuentra gran variedad de recomendaciones que buscan hacer del aprendizaje de los números una tarea fácil y divertida. Por ejemplo el portal Web guiainfantil.com propone una serie de juegos, cuentos y adivinanzas para la enseñanza de los números, uno de ellos consiste en dibujar los números en cartulinas y jugar con los niños hasta que aprendan reconocer cual es cada número, este juego tiene como ventaja su bajo costo [2]. El sitio Web crecerfeliz.es sugiere utilizar la vida

cotidiana para enseñar y poner en práctica la habilidad de contar, poner la mesa y pedirles ayuda en las compras son algunas de las propuestas que se encuentran en este sitio Web [3]. Así mismo, diferentes autores han propuesto gran variedad de metodologías para la enseñanza de competencia numéricas en los niños. En [4] examinan si un juego de mesa resulta eficaz para mejorar el conocimiento de las magnitudes numéricas de los prescolares en dos circunstancias diferentes. El primer escenario consiste en que los niños jueguen en pequeños grupos, el segundo se trata de que el adulto en el juego sea un profesional que recibió un breve entrenamiento acerca del juego. El estudio dio como resultado que un juego de mesa cuando se juega en pequeños grupos y acompañados de docentes puede promover el conocimiento de los niños. En [5] proponen el uso de una herramienta didáctica llamada Numicon para la enseñanza de habilidades y conceptos numéricos a niños con síndrome de Down, el enfoque de Numicon es que los niños realmente sientan que están aprendiendo mediante la manipulación física y visualizando cálculos.

Por otro lado, recientemente se han desarrollado diferentes trabajos basados en el reconocimiento de gestos con las manos. En [6] proponen un esquema de reconocimiento de gestos con las manos basados en datos de profundidad adquiridos con el sensor Kinect, este trabajo se centra en dos conjuntos de características, uno basado en la distancia de la yema de los dedos al centro de la mano y otro en la curvatura del contorno de la mano, finalmente utilizan una máquina de soporte vectorial SVM para realizar el proceso de clasificación. Sin embargo, en este trabajo se localiza la mano a partir del punto más cercano a la cámara, lo que puede ser una limitación ya que se debe utilizar la aplicación en un espacio totalmente despejado. En [7] al igual que en [6] se utilizan mapas de profundidad obtenidos con el Kinect, los autores de este artículo exponen dos métodos para realizar el reconocimiento, el primero consiste en dibujar un círculo en el centro de la mano que corta los dedos de la mano, la intersección del círculo con la silueta representa la detección de los dedos y el segundo método utiliza descriptores de Fourier, los resultados de este trabajo son bastante buenos pero solo permiten el reconocimiento de los números del 1 al 5 de la forma en que se realizan habitualmente con las manos.

Aunque se han propuesto diferentes mecanismos para la enseñanza de números a párvulos aún no hay uno realmente eficiente que sea capaz de lograr que los niños aprendan los números y los entiendan. Adicionalmente, en cuanto al reconocimiento de gestos con las manos, todavía no se ha

logrado obtener un método que dé como resultado un reconocimiento de gestos eficiente y preciso.

Este trabajo propone una metodología para el reconocimiento de gestos con las manos, mediante mapas de profundidad obtenidos con el sensor Kinect y la extracción de una serie de características geométricas que permiten identificar los números del 1 al 5. Este reconocimiento de gestos será utilizado como herramienta para enseñar los números a párvulos. La aplicación contempla el uso de un agente virtual (animaciones virtuales que inicialmente le explique al usuario mediante imágenes y sonidos los números del 1 al 5).

El artículo se divide en 4 secciones. La sección 1 contiene los trabajos relacionados y la presente introducción, en la sección 2 se expone la metodología utilizada para el desarrollo del proyecto, en la sección 3 se muestran los resultados obtenidos y finalmente en la sección 4 se encuentran las conclusiones.

2 Metodología

A continuación se describe la metodología propuesta para la clasificación de gestos con las manos, nuestro sistema de reconocimiento se compone de las siguientes etapas:

2.1 Adquisición de datos

El sensor Kinect es un dispositivo que incorpora diferentes hardware para la detección avanzada, en particular contiene un sensor de profundidad, una cámara de color y una serie de cuatro micrófonos. El sensor de profundidad consiste en un proyector IR combinado con una cámara IR [8].

Usando OpenNI como interface con el Kinect se obtiene el mapa de profundidad que permitirá seguir el movimiento de la mano de una forma mucho más sencilla y eficiente. En la Fig. 1 se puede observar el mapa de profundidad que proporciona el sensor Kinect, el valor de profundidad se codifica en escala de grises es decir, cuanto más cercanos estén los elementos presentes en la escena al sensor, más claros serán los píxeles, sin embargo los píxeles con nivel de intensidad 0 indican que no hay valores de profundidad.



Figura. 1. Mapa de Profundidad obtenido con el Sensor Kinect.

Con el fin de facilitar el procesamiento de las imágenes, el mapa de profundidad obtenido con OpenNI fue invertido, es decir, los píxeles más cercanos al sensor serán los que tenga un nivel de intensidad cercano a 0.

2.2 Segmentación

El segundo paso para el reconocimiento de gestos es la segmentación, el procesamiento de los datos obtenidos con el sensor Kinect es realizado con la librería OpenCV. El método propuesto se plantea mediante la Ecuación (1), consiste en localizar la mano en una región R_k trazada alrededor del punto central P_i de la mano, este punto es obtenido con la ayuda de OpenNI. Son considerados como parte de la mano M , aquellos puntos cuya distancia entre P_i se encuentre dentro de un rango U establecido empíricamente.

$$P_i \wedge P_j \in R_k \because P_j \in M \forall U_- \leq dist(P_i, P_j) \leq U_+ \quad (1)$$

En la siguiente figura se puede observar el resultado del método de segmentación propuesto.



Figura. 2. Localización y segmentación de la mano

2.3 Extracción de características

Utilizando el toolbox Balu [9], se extrajeron 44 características, geométricas 2D, geométricas 3D, y topológicas, que luego pasaron por un proceso de selección basado en el criterio de Fisher, 14 características fueron elegidas como las más relevantes para la clasificación. Luego de conocer las características adecuadas usando Balu y Matlab, estas fueron implementadas en C++ utilizando la librería cvblobslib.

Para la extracción de las 5 características que se explicaran a continuación se realizó un procesamiento de blobs. Un blob (Binary Large Object) es un conjunto conectado y amorfo de píxeles de una imagen cuya intensidad es superior a un umbral determinado y que están unidos entre sí [10]. En la Fig 3, la región azul representa el blob de mayor área obtenido en la imagen.

El procesamiento de blobs fue realizado con la librería cvblobslib, la cual permite extraer distintas características geométricas.

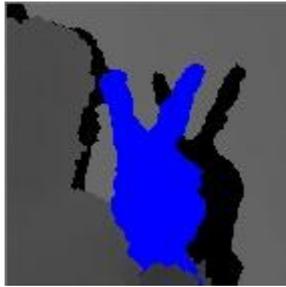


Figura. 3. Extracción de blobs

2.3.1 Área

El área representa el número de píxeles que hacen parte del blob. Para números como el 1 la mano se encuentra cerrada, por lo que el área no será tan grande como para el 5 en el que la mano se encuentra completamente abierta y se muestran todos los dedos. La Ecuación (2) especifica el procedimiento utilizado para hallar el área.

$$A = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} M(i, j) \quad (2)$$

2.3.2. Perímetro

Esta característica está representada mediante la Ecuación (3) y nos entrega la longitud alrededor del contorno del blob. La forma de hallar el perímetro consiste en calcular la distancia entre cada par de píxeles adyacentes alrededor del borde de la región.

$$P = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} dist(M(i, j), M(i + 1, j + 1)) \quad (3)$$

2.3.3 Alto y ancho del blob

El ancho se define como la distancia entre el borde interior izquierdo y derecho del blob, por otro lado, el alto es la distancia entre el borde interior superior e inferior del blob. Las Ecuaciones (4) y (5) muestran la forma en que es hallado el alto y ancho del blob respectivamente.

$$W = X_{max} - X_{min} \quad (4)$$

$$H = Y_{max} - Y_{min} \quad (5)$$

2.3.4 Redondez

La redondez nos da una medida de que tan cerca es la forma de un objeto a un círculo. Un objeto redondo perfecto, tendrá

una redondez de 1.0, mientras que los objetos rectos y alargados se acercaran a cero [11], en nuestro caso para números como el 1 y 2 la redondez tendrá valores mucho más cercanos a cero y para números como el 5 esta característica nos dará valores un poco más altos.

La redondez fue hallada mediante la siguiente fórmula:

$$r = \frac{4 \cdot \pi \cdot A}{P^2} \quad (6)$$

2.3.5. Eje mayor del blob

Finalmente, con la ayuda de la librería cvblobslib se extrae esta característica, la cual corresponde a la longitud del eje mayor del blob.

Las características que se muestran a continuación fueron extraídas a partir del contorno de la mano. Para obtener el contorno, se aplicó un filtro con el fin de reducir el ruido en la imagen y luego se utilizó OpenCV para extraer el contorno. En la Fig 4, se muestra el resultado de la extracción de contornos con esta librería.



Figura. 4. Contorno de la mano obtenido con OpenCV

2.3.6 Diámetro Equivalente

El diámetro equivalente hace referencia al diámetro del círculo el cual tiene un área igual a la zona del contorno de la mano. En la Ecuación (7) se muestra la forma que se halló esta característica.

$$D_e = \sqrt{\frac{4 \cdot A_c}{\pi}} \quad (7)$$

2.3.7 Eje mayor de la elipse

Esta característica hace referencia a la longitud en píxeles del eje mayor de la elipse que se ajusta al contorno de la mano, El eje mayor de la elipse es hallado sumando las distancias desde cada foco hasta cualquier punto de la elipse. En la Fig 5 se puede observar la elipse obtenida alrededor del contorno de la mano y la representación gráfica de la Ecuación (8).

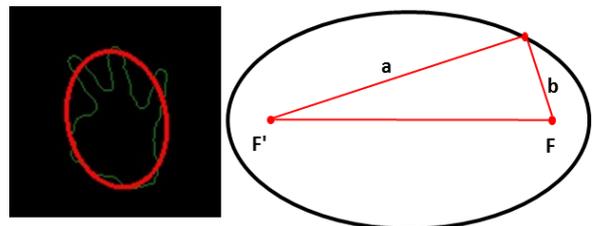


Figura. 5. Extracción del eje mayor de la Elipse

$$Major_{axis} = a + b$$

2.3.8 Descriptores de Fourier

La transformada de Fourier puede ser utilizada para la descripción de formas, debido a que los primeros valores de dicha transformada, los cuales corresponde a las frecuencias bajas, permiten diferenciar objetos de forma muy precisa. El proceso de extracción de esta característica consiste en introducir como valor complejo cada punto del contorno (ver Fig 6).

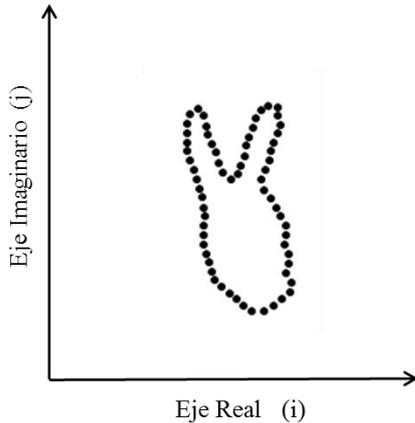


Figura. 6. Contorno de una imagen en el espacio de los números complejos

Matemáticamente para imágenes bidimensionales la transformada de Fourier es:

$$F(k, l) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i, j) e^{-i2\pi\left(\frac{ki}{N} + \frac{lj}{N}\right)} \quad (9)$$

En la Ecuación (9), f es el nivel de intensidad de la imagen en su dominio de espacial y F en su dominio de la frecuencia. El resultado de la transformación son números complejos. Esto se puede interpretar como una imagen real y una imagen compleja o como una imagen magnitud y una imagen fase. Sin embargo, a través de los algoritmos de procesamiento de imágenes, solo la imagen magnitud es interesante ya que contiene toda la información que se necesita sobre la estructura geométrica de las imágenes [12].

Los descriptores de Fourier utilizados en este proyecto son: descriptor de Fourier 1, 2, 3, 5, 6, 8 [13].

2.4 Clasificación

Para el desarrollo de este proyecto se probaron diferentes clasificadores (SVM, LDA, KNN y Redes Neuronales) y con el que mejor resultados se obtuvieron fue con el clasificador LDA. LDA o Linear Discriminant Analysis es una de las técnicas de clasificación de datos más utilizadas. La idea central de LDA es obtener una proyección de los datos en un espacio de menor (o incluso igual) dimensión que los datos entrantes, con el fin de que la separabilidad de las clases sea lo mayor posible. Es una técnica supervisada ya que para

- (8) poder buscar esa proyección se debe entrenar el sistema con patrones etiquetados [13].

3 Resultados

Para evaluar el rendimiento de nuestro clasificador se utilizó una base de datos de 779 imágenes de 5 gestos realizados con las manos.

Con el fin de determinar la eficiencia del clasificador se utilizó validación cruzada, la cual consiste en que para cada k muestras se utiliza $k-1$ para entrenar y el restante para pruebas, este tipo de validación tiene como ventaja que la estimación del error es mucho más estable. El tamaño de k elegido en este proyecto es 10.

Predicción \ Real	1	2	3	4	5
1	235	11	0	0	0
2	3	139	5	0	0
3	0	2	153	3	0
4	0	0	2	105	3
5	0	0	0	6	114

Tabla 1. Matriz de confusión para el reconocimiento de gestos

En la Tabla 1 se puede observar la matriz de confusión, en la cual se muestra la correspondencia entre la clase asignada a una muestra y la verdadera clase a la que pertenece. El porcentaje de acierto del clasificador corresponde a un 95%, lo que indica que la metodología propuesta en este artículo resulta efectiva para el reconocimiento de gestos con las manos.

4 Conclusiones

En este trabajo se presentó una metodología que permite identificar eficazmente los números del 1 al 5 realizados con las manos, es importante resaltar que el uso del Sensor Kinect garantizó en este proyecto la localización de la mano de una forma mucho más fácil, rápida y precisa gracias al mapa de profundidad que proporciona dicho dispositivo y a la librería OpenNI. El proceso de segmentación resultó ser bastante efectivo cuando la mano del niño se encuentra separada del resto del cuerpo, es decir, cuando en el mapa de profundidad se puede observar un contraste de color entre la mano y el cuerpo. Adicionalmente, las características seleccionadas con la ayuda del Toolbox Balu resultaron ser bastante efectivas para el reconocimiento de gestos. Finalmente, el clasificador y el método de validación utilizado arrojaron resultados bastante favorables.

Teniendo en cuenta que uno de los objetivos principales de este trabajo es la enseñanza de los números a párvulos, actualmente se está desarrollando una aplicación de realidad aumentada utilizando el método de reconocimiento de gestos propuesto en este artículo, esta aplicación permitirá a los niños interactuar con su propio aprendizaje, a través de proyecciones en tiempo real. Así mismo, este proyecto está enfocado en realizar una medición del impacto de esta herramienta en la enseñanza de los números a párvulos, lo

que permitirá obtener nuevos mecanismos de enseñanza no solo de los números sino de diferentes áreas del aprendizaje.

En futuros trabajos se pretende trabajar en un mecanismo de segmentación que permita extraer la mano cuando se encuentra muy cerca al cuerpo del usuario, además de incorporar más características que permitan la clasificación de los gestos con un porcentaje de reconocimiento más alto y que permita reconocer los dedos aun cuando estos se encuentran completamente juntos.

Referencias

- [1] A. Rafael Linares, «Desarrollo Cognitivo: Las Teorías de Piaget y de Vygotsky,» [En línea]. Available: http://www.paidopsiquiatria.cat/files/teorias_desarrollo_cognitivo_0.pdf. [Último acceso: 15 Junio 2015].
- [2] L. Vélez, «Guiainfantil.com,» [En línea]. Available: <http://www.guiainfantil.com/articulos/educacion/aprendizaje/juegos-y-tecnicas-para-ensenar-colores-y-numeros-a-ninos-de-3-anos/>. [Último acceso: 15 Junio 2015].
- [3] E. García, «Crecefeliz.es,» [En línea]. Available: <http://www.crecefeliz.es/Ninos/Desarrollo-y-aprendizaje/Aprender-a-contar-jugando>. [Último acceso: 15 Junio 2015].
- [4] G. Ramani, R. Siegler y A. Hitti, «Taking It to the Classroom: Number Board Games as a Small Group,» *Journal of Educational Psychology*, vol. 104, n° 3, pp. 661-672, 2012.
- [5] T. Wing y R. Tacon, «Teaching number skills and concepts with Numicon materials,» *Down Syndrome Research and Practice*, vol. 12, n° 1, pp. 22-26, 2007.
- [6] Dominio, Fabio; Donadeo; Marin, Giulio; Zanuttigh, Pietro; Cortelazzo, Guido Maria, "Hand Gesture Recognition with Depth Data," ARTEMIS 2013 - Proceedings of the 4th ACM/IEEE International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Stream, pp. 9- 16 October 21, 2013.
- [7] C. Chan y S. S. Mirfakhraei, «Hand Gesture Recognition using Kinect,» Boston , 2013.
- [8] Zhang, Z., (2010). Microsoft Kinect Sensor and Its Effect, *Multimedia at Work*, vol.19, 4-10.
- [9] Mery, D. (2011): BALU: A Matlab toolbox for computer vision, pattern recognition and image processing. <http://dmery.ing.puc.cl/index.php/balu>
- [10] A. Jiménez González, «Percepción Activa para Seguimiento de Objetos en Entornos Urbanos; Capitulo 3: Detección y seguimiento de,» [En línea]. Available: <http://bibing.us.es/proyectos/abreproy/70077/fichero/capitulo3.pdf>. [Último acceso: 20 Junio 2015].
- [11] A. González Marcos, F. Martínez, A. Pernía Espinoza, F. Alba Elías, M. Castejon Limas, J. Ordieres Meré y E. Vergara González, «Técnicas y Algoritmos Básicos de Visión Artificial,» 2006. [En línea]. Available: <https://publicaciones.unirioja.es/catalogo/online/VisionArtificial.pdf>. [Último acceso: 10 Junio 2015].
- [12] OpenCV, «OpenCV.org,» 25 Febrero 2015. [En línea]. Available:http://docs.opencv.org/doc/tutorials/core/dscrite_fourier_transform/discrete_fourier_transform.html. [Último acceso: 10 Junio 2015].
- [13] Zahn, C; Roskies, R.: Fourier Descriptors for Plane Closed Curves, *IEEE Trans on Computers*, C21(3):269-281, 1972.