

CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO
NACIONAL
ROBÓTICA Y MANUFACTURA AVANZADA



Visión por Computadora

PROYECTO

Hand Gesture Detection and Recognition

Integrantes:
Hilario Acuapan Gabriela
Pineda Gómez Luis Alberto

RAMOS ARIZPE, COAHUILA, APRIL 2023



Contents

| | | |
|----------|---|-----------|
| 1 | Introducción | 3 |
| 2 | Fundamentos teóricos | 3 |
| 2.1 | Análisis de componentes principales | 3 |
| 2.2 | Reconocimiento de gestos de mano | 6 |
| 3 | Desarrollo de la metodología | 6 |
| 3.1 | Adquisición de datos | 7 |
| 3.2 | Training set | 7 |
| 3.3 | Eigenhands con PCA | 9 |
| 3.4 | Clasificación | 11 |
| 4 | Resultados | 12 |
| 5 | Conclusiones | 12 |

1 Introducción

El objetivo principal de este trabajo es detectar y reconocer los gestos de mano en 2D (en imágenes) a partir de su forma, comparándolos con gestos existentes en una base de datos. Para cumplir este objetivo se utiliza un algoritmo para la detección y reconocimiento de gestos basado en los componentes principales (PCA). El algoritmo a implementar se describe en el artículo de Turk y Pentland [3] (*Face Recognition Using Eigenfaces*), pero en lugar de aplicarlo a la detección de rostros humanos, se utilizará para la detección de gestos de mano.

El algoritmo de *Eigenfaces*, fue presentado por primera vez por Sirovich y Kirby en su artículo de 1987 [2], y posteriormente formalizado por Turk y Pentland en su artículo de 1991 [3]. El procedimiento general indica que cada cara se almacena en un vector de dimensión N^2 . Y el Análisis de los Componentes Principales (PCA) se utiliza para encontrar un subespacio de dimensión M cuyos vectores de la base corresponden a las direcciones de máxima varianza en el espacio original de la imagen (N^2). Este nuevo subespacio es normalmente de dimensión más baja ($M \ll N^2$).

El enfoque del algoritmo es descomponer las imágenes de los rostros en un conjunto más pequeño de rasgos característicos llamados "eigenfaces", que pueden considerarse como los componentes principales de la base de datos original. Para la parte del reconocimiento se proyecta una nueva imagen en el subespacio generado por las eigenfaces (llamado "face space"), para luego clasificar el rostro comparando su posición en el *face space* de las personas conocidas (o clases conocidas).

Este trabajo aquí presentado trata de replicar este procedimiento pero en lugar de reconocer rostros, se utilizará para reconocer y clasificar gestos de la mano. . . .

2 Fundamentos teóricos

2.1 Análisis de componentes principales

El objetivo de implementar un análisis por componentes principales es realizar una **reducción de dimensionalidad** ó del número de variables en una base de datos. Sin embargo, el realizar este análisis viene acompañado de un precio a pagar y esto es la relación **precisión - simplicidad**. Por un lado, al reducir la dimensión de nuestra base de datos, el realizar cálculos con ella será mucho más rápido y eficiente, además de que la interpretación de los datos es más sencilla a expensas de perder precisión y exactitud de la base original.

El análisis por componentes principales puede ser implementado en 5 pasos:

Paso 1.- Estandarización

En este primer paso realizamos una estandarización de los datos, esto con el objetivo de evitar que valores grandes, por ejemplo, que se encuentren en un rango de (100 a 1000) no dominen entre valores menores, por ejemplo (0 a 10). Esto es para evitar que nuestra base de datos presente lo que conocemos como **sesgo o bias**.

La estandarización de los datos se logra aplicando la siguiente expresión 1:

$$Z = \frac{\text{valor} - \text{media}}{\text{desviación estándar}} \quad (1)$$

Paso 2.- Obtención de la matriz de covarianza

El propósito por el cual se calcula la matriz de covarianza es para obtener la relación existente entre las vari-

ables de la base de datos. Se busca encontrar cómo es que varían entre ellas. En algunas ocasiones, las variables se encontrarán relacionadas de tal forma, que tendrán información **redundante**, esto es, que varias variables contengan información similar. Este tipo de relaciones son las que se encuentran en la matriz de covarianza.

La matriz de covarianza es una matriz **simétrica** de dimensiones $p \times p$. Por ejemplo, supongamos que tenemos una matriz de covarianza de 3×3 :

$$\begin{bmatrix} Cov(x, x) & Cov(x, y) & Cov(x, z) \\ Cov(y, x) & Cov(y, y) & Cov(y, z) \\ Cov(z, x) & Cov(z, y) & Cov(z, z) \end{bmatrix}$$

Donde los elementos de la diagonal de la matriz $Cov(x, x) = Var(x)$ y también es importante mencionar que $Cov(x, y) = Cov(y, x)$.

La información obtenida por la matriz de covarianza viene dada principalmente por el **signo**. Esto es:

- Si es **Positivo**: Las dos variables incrementan o decrecen de la misma forma (Correlacionadas).
- Si es **Negativo**: Una variable incrementa cuando la otra decrece (Inversamente correlacionadas).

Paso 3.- Obtención de los eigenvalores y eigenvectores de la matriz de covarianza

Esto se realiza con el objetivo de obtener los **componentes principales** de los datos. En los componentes principales, se encuentra contenida la información principal o más significativa.

Por ejemplo, en la figura 1, se muestra una gráfica de la varianza dado el número de componentes principales. En este caso, en el componente principal de mayor importancia se encuentra codificada el 40% de la covarianza. Esto es que dicho componente contiene información *única*.

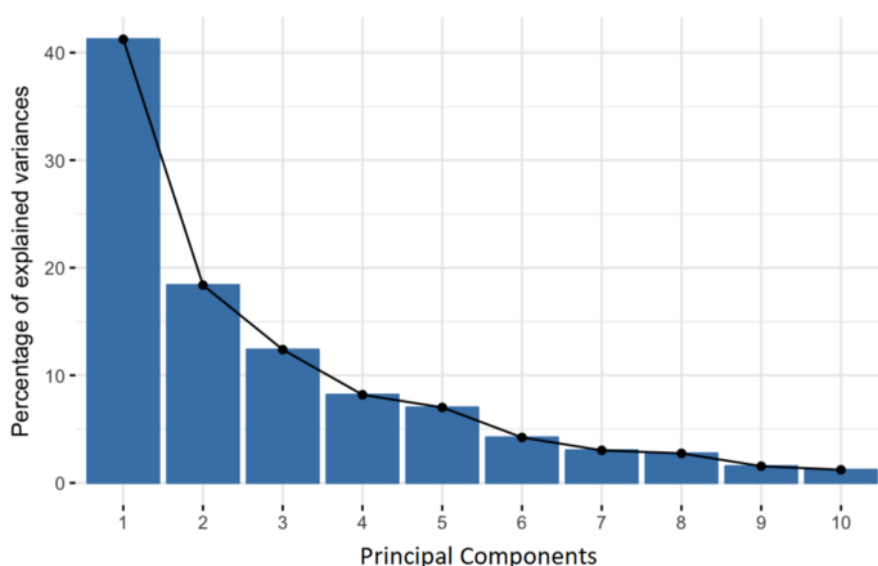


Figure 1: Comparación del Porcentaje de la varianza vs el número de componentes principales

Con este procedimiento, reducimos la dimensionalidad de la base de datos original sin tener un impacto sustancial en la pérdida de información. La parte sustancial del PCA, es que los componentes principales **carecen** de interpretación. Esto quiere decir que no se le puede atribuir cierto atributo o explicación a un componente principal.

Debido a que existe tantos componentes principales como número de variables en la base de datos, el PCA se construye de tal forma que el componente primario contenga la mayor cantidad de varianza posible de la base de datos. El componente secundario contendrá la mayor cantidad de varianza posible, pero que no se encuentra en la primer componente y así sucesivamente con el resto de componentes.

Los **eigenvectores** juegan el papel de indicar los ejes en **donde** se encuentra la mayor cantidad de varianza, esto es, la mayor cantidad de información de la base de datos. Mientras que los **eigenvalores**, dan la **cantidad** de varianza que contiene cada componente principal.

Paso 4.- Construcción del Vector característico

En este paso, es donde se toma la decisión de conservar o descartar aquellos **eigenvectores** que no aporten información relevante de la base de datos. Con los vectores que deseemos conservar, se construirá una nueva matriz, conformada por los vectores columna de los eigenvectores más representativos.

Paso 5.- Reconformar la base de datos a lo largo de los ejes de los componentes principales

Para este último paso, tenemos que **reorientar** los datos a lo largo de los nuevos eigenvectores en donde se encuentra concentrada la mayor cantidad de información. Esto se obtiene de la siguiente forma:

$$Final\ data\ set = Matriz\ caracteristica^T * Data\ set\ estandarizado$$

2.2 Reconocimiento de gestos de mano

3 Desarrollo de la metodología

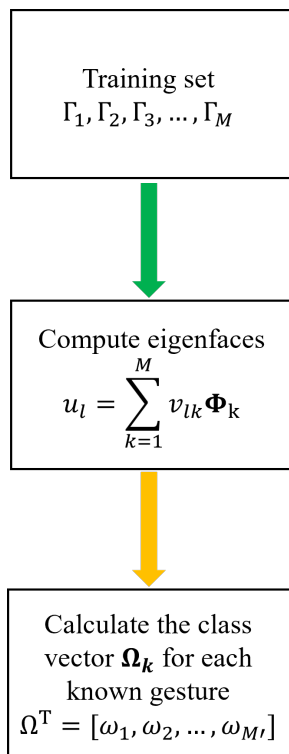


Figure 2: Operaciones de inicialización.

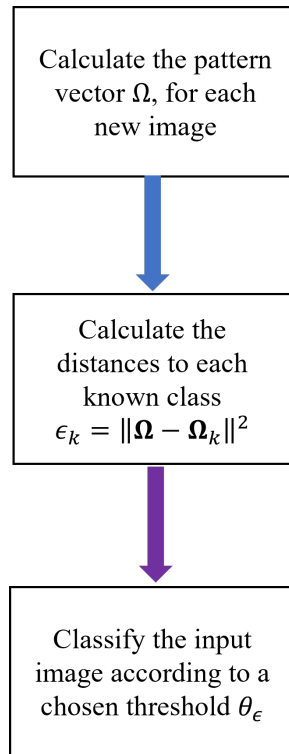


Figure 3: Operaciones para clasificación.

3.1 Adquisición de datos

La base de datos utilizada para este proyecto se creo a partir de la captura de imágenes en infrarrojo, utilizando el Sensor mostrado en la figura 4.



Figure 4: Structure sensor 3D [1].

3.2 Training set

Se utilizaron $M = 70$ imágenes para el conjunto de entrenamiento, donde se incluyeron los 7 gestos de mostrados en la figura 5.

Gestos para el set de entrenamiento

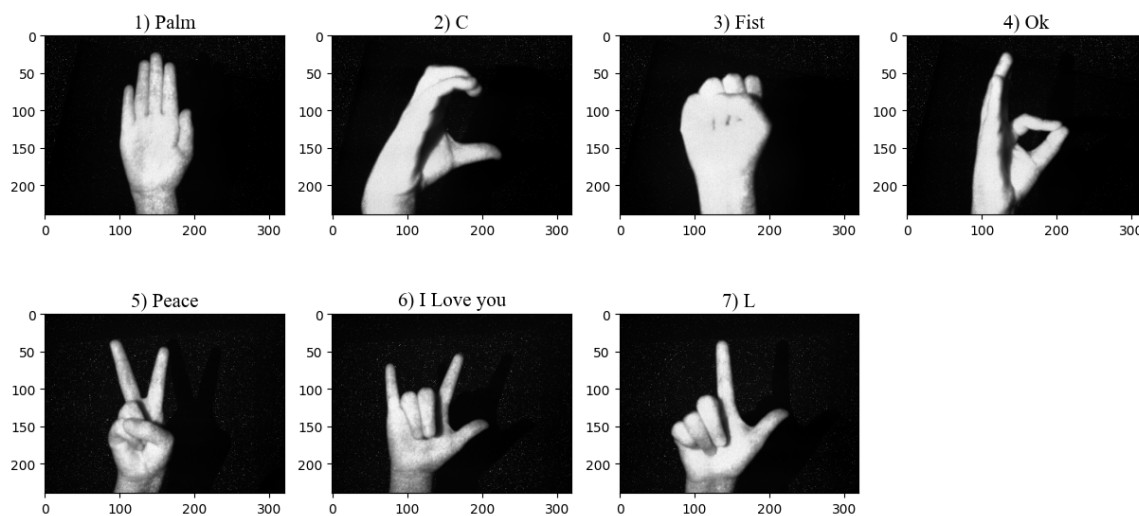


Figure 5: Gestos para de entrenamiento.

Para seguir con el proceso es necesario realizar un preprocesamiento de las imágenes de la base de datos de la figura 5. Este consiste en realizar una binarización de las imágenes, para encontrar el contorno de la mano. Posteriormente dibujar el bounding box sobre la imagen original para recortarla y finalmente realizar un reescalamiento de la imagen a 200x200 píxeles.

Denotando al conjunto de entrenamiento como

$$\Gamma_1, \Gamma_1, \dots, \Gamma_M \quad (2)$$

Cada imagen de dimensiones $N \times N$ se convierte en un vector Γ de dimensión N^2

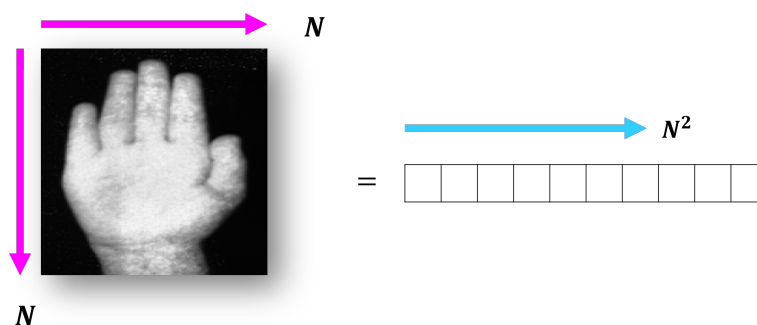


Figure 6: Vector imagen.

Estos vectores forman el conjunto de entrenamiento (2),

$$\begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ \vdots \\ \Gamma_M \end{bmatrix} = \begin{pmatrix} \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} \\ \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} \\ \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} & \boxed{} \end{pmatrix}_{M \times N^2}$$

Figure 7: Conjunto de entrenamiento.

3.3 Eigenhands con PCA

Teniendo los vectores del conjunto (2), se obtiene la mano promedio (figura 8) mediante la siguiente operación

$$\Psi = \frac{1}{M} \sum_{n=1}^M \Gamma_n \quad (3)$$

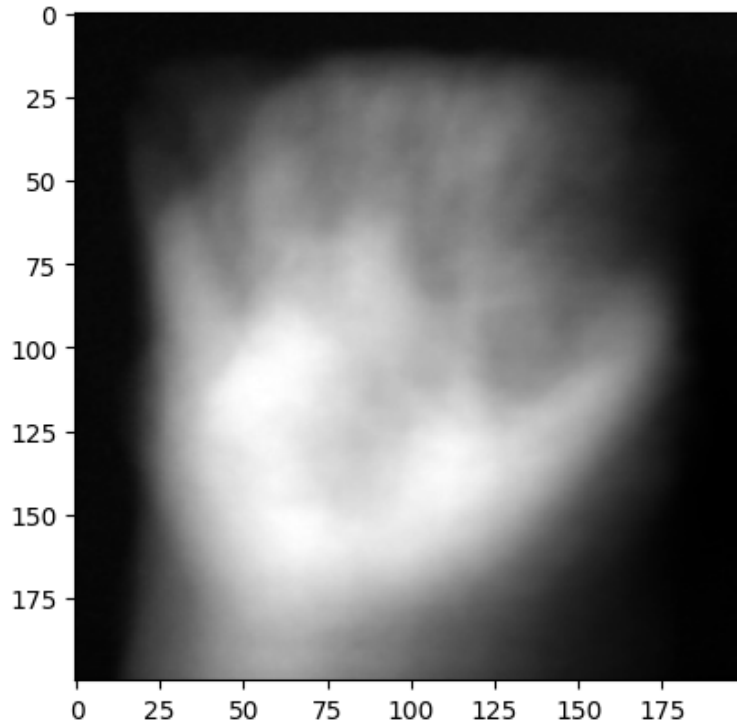


Figure 8: Mano promedio.

Posteriormente se calcula la desviación Φ que tiene cada Γ con respecto al promedio Ψ

$$\Phi = \Gamma_i - \Psi \quad (4)$$

Este conjunto de vectores Φ se somete a un análisis de componentes principales (PCA), que busca un conjunto de M vectores ortonormales, u_n , que mejor describen la distribución de los datos.

Para esto definimos una matriz A que contiene a los vectores de las desviaciones Φ

$$A = [\Phi_1, \Phi_2, \dots, \Phi_M] \quad (5)$$

Para construir una matriz L de dimensiones $M \times M$

$$L = A^T A \quad (6)$$

Y encontrar los eigenvalores μ_i y eigenvectores v_i de la matriz L , tal que

$$Av_i = \mu_i v_i \quad (7)$$

Se deben encontrar los M eigenvectores v_i asociados a los eigenvalores μ_i más representativos. Para esto obtenemos la suma acumulada y observamos que con 40 eigenvectores reconstruimos casi el 90% de la información (figura 9)

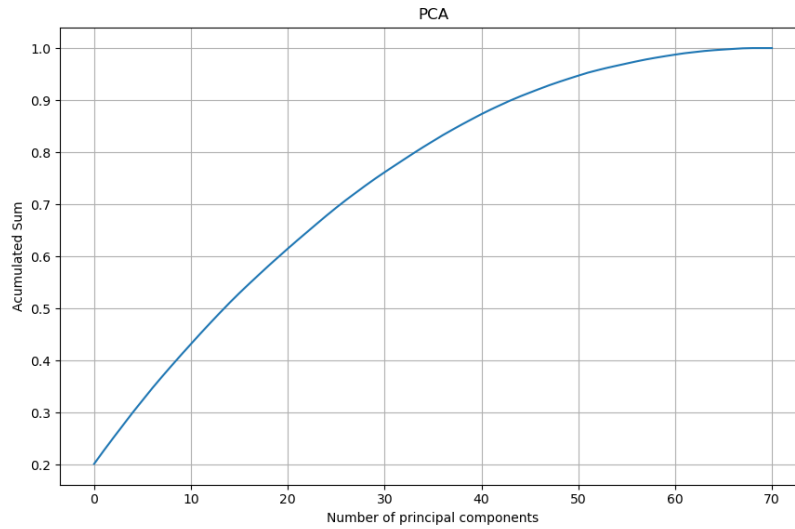


Figure 9: Suma acumulada de eigenvalores de L .

Estos vectores determinan combinaciones lineales de las imágenes del conjunto de entrenamiento M , y se utilizan para formar a las eigenhands u_i .

$$u_i = \sum_{k=1}^M v_{lk} \Phi_k, \quad l = 1, \dots, M \quad (8)$$

En la figura 10 se muestran 10 de las 40 eigenhands principales derivadas de las imágenes de entrenamiento.

Primeras 10 eigenhands obtenidas

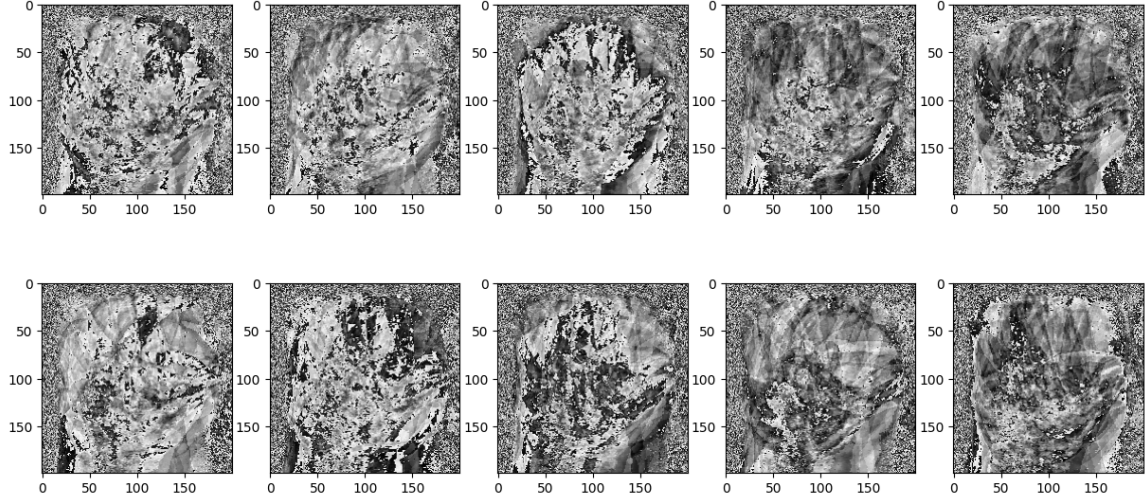


Figure 10: Primeras 10 eigenhands más representativas.

3.4 Clasificación

Las imágenes de eigenhands calculadas a partir de los vectores propios de L abarcan un conjunto básico (de dimensión $M \times M'$) con el que describir el conjunto de imágenes inicial.

Para realizar la clasificación es necesario encontrar los patrones Ω_K que describen a la k -ésima clase de gestos.

$$\Omega^T = [\omega_1, \omega_2, \dots, \omega_{M'}] \quad (9)$$

Los pesos que forman al vector Ω describen la contribución de cada eigenfaces en la representación de la imagen de la cara de entrada, y se calculan mediante la siguiente operación

$$\omega_k = u_k^T (\Gamma - \Psi) \quad (10)$$

Las clases de los gestos se calculan promediando los patrones de pesos Ω de un pequeño número de imágenes de cada gesto.

Para determinar a qué clase pertenece una nueva imagen de entrada se calcula su patrón de pesos Ω , para luego utilizar la distancia euclidiana

$$\epsilon_k = \|\Omega - \Omega_k\|^2 \quad (11)$$

La imagen se clasifica como perteneciente a la clase k cuando ϵ_k está por debajo de algún umbral elegido θ_k

4 Resultados

5 Conclusiones

References

- [1] Structure - the world's leading healthcare 3d scanning platform. Consultado el 10 de abril de 2023. [Online]. Available: <https://structure.io/>
- [2] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Josa a*, vol. 4, no. 3, pp. 519–524, 1987.
- [3] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.