

**Nombre:** Vergara Kenia.

**C. I. N°:** V- 16.443.850.

## Reconociendo las emociones expresadas por la pose del cuerpo: un modelo neuronal inspirado biológicamente

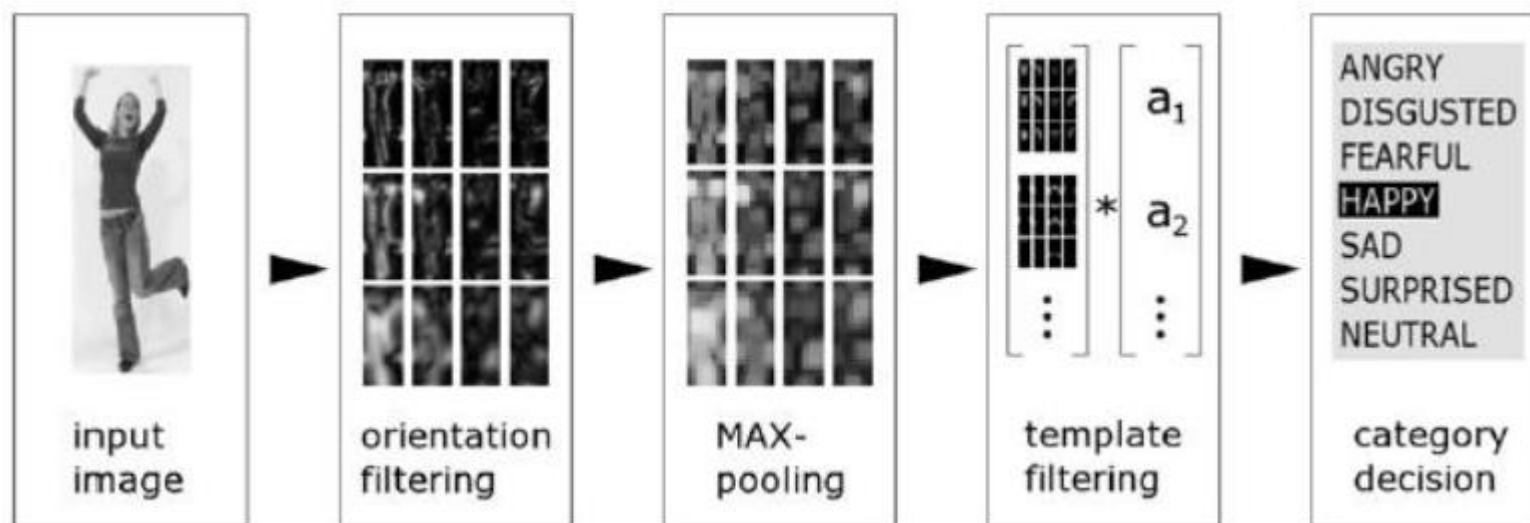


Fig. 1. Ilustración del modelo neuronal. A partir de la imagen cruda en la retina, se extrae la orientación local (área V1), se agrupa en vecindarios espaciales (V2 / V4) y se filtra con características complejas aprendidas (V4 / I). La respuesta del filtro sirve como entrada en un clasificador discriminativo (ESO). Los parámetros se eligieron con fines ilustrativos y son diferentes de la implementación real.

El primer nivel de la jerarquía consiste en un conjunto de filtros log-Gabor con parámetros fisiológicamente plausibles, para extraer la orientación local en múltiples escalas desde la imagen de entrada. El filtrado tipo Gabor es una forma estándar de aproximar las respuestas de las células simples de Hubel y Wiesel (1962) en el área V1 de la corteza visual primaria. Específicamente, utilizamos filtros log-Gabor, que permiten una mejor cobertura del espectro que la versión estándar (lineal) con menos frecuencias preferidas, y son consistentes con mediciones electrofisiológicas.

El siguiente nivel consiste en neuronas con campos receptivos más grandes, que agrupan las respuestas del filtro de células espacialmente adyacentes, produciendo una invarianza de posición más alta, como se observa para las células en el área V2. La agrupación se realiza con el operador MAX, lo que significa que la respuesta más fuerte determina la salida de la neurona de agrupamiento, por separado para cada orientación y escala.

El tercer nivel consiste en detectores de características más específicas sensibles a las estructuras más complejas (por ejemplo, las “unidades ajustadas por componentes” de Riesenhuber y Poggio (1999), atribuidas a las áreas V4 / IT). En nuestro modelo, estas estructuras se aprenden a través del análisis de componentes principales (PCA) de un gran conjunto de respuestas del nivel anterior. Una vez aprendido, cada vector de base de la PCA reducida representa un detector. Por proyección en PCA, el conjunto de detectores se aplica a la señal entrante del nivel anterior, y los coeficientes resultantes forman la salida de la capa.

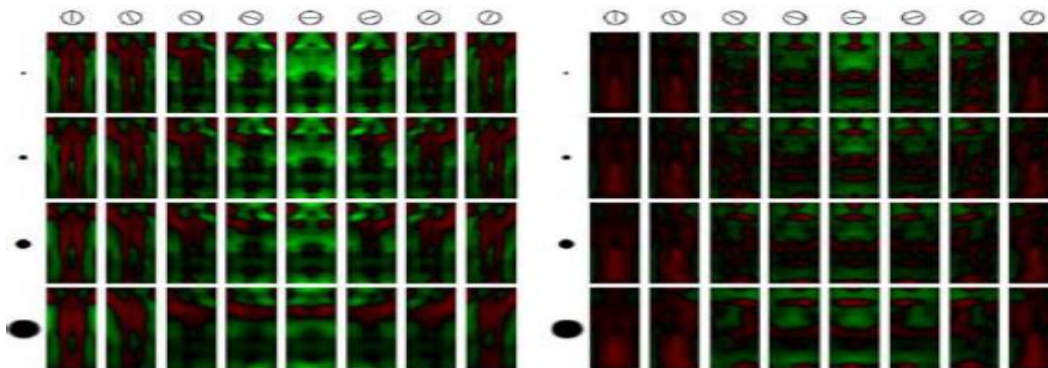


Fig. 2. Los vectores de base PCA se utilizan como plantillas para funciones complejas que codifican componentes de la pose humana. Se muestran el primer y segundo vector de base para el banco de filtros de 32 canales. Los valores positivos se imprimen en rojo, los valores negativos en verde, más claro significa valores más grandes. Los símbolos a la izquierda y en la parte superior de las plantillas indican la escala y la orientación del filtro Log-Gabor correspondiente.

En el nivel superior, se trabaja con un clasificador de máquina de vector de soporte (SVM), tienen sus raíces en la teoría del aprendizaje estadístico y estiman el hiperplano de separación entre dos categorías como la solución de un problema de optimización de forma cerrada. Esta propiedad tiene ventajas para nuestra simulación, debido a la cantidad limitada de datos de entrenamiento disponibles, pero también se puede aprender un clasificador equivalente con mecanismos neuronales, por ejemplo, Anguita y Boni (2002). Para extender el concepto SVM binario a  $N > 2$  categorías, usamos el método "uno contra todos", es decir, para cada categoría se aprende un clasificador binario, que lo discrimina de todos los demás. La señal de entrada se alimenta a todos los clasificadores  $N$ , y se selecciona la categoría con la salida más fuerte (correspondiente al margen de decisión más grande). Permitimos la asignación de varias poses diferentes a la misma emoción. Un ejemplo ilustrativo se muestra en la Fig. 3.

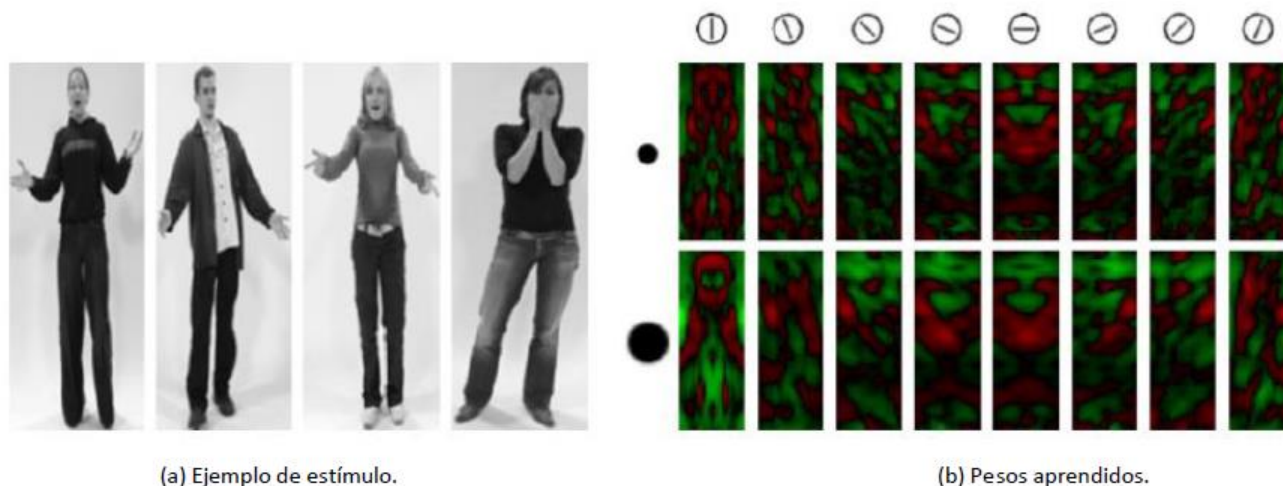
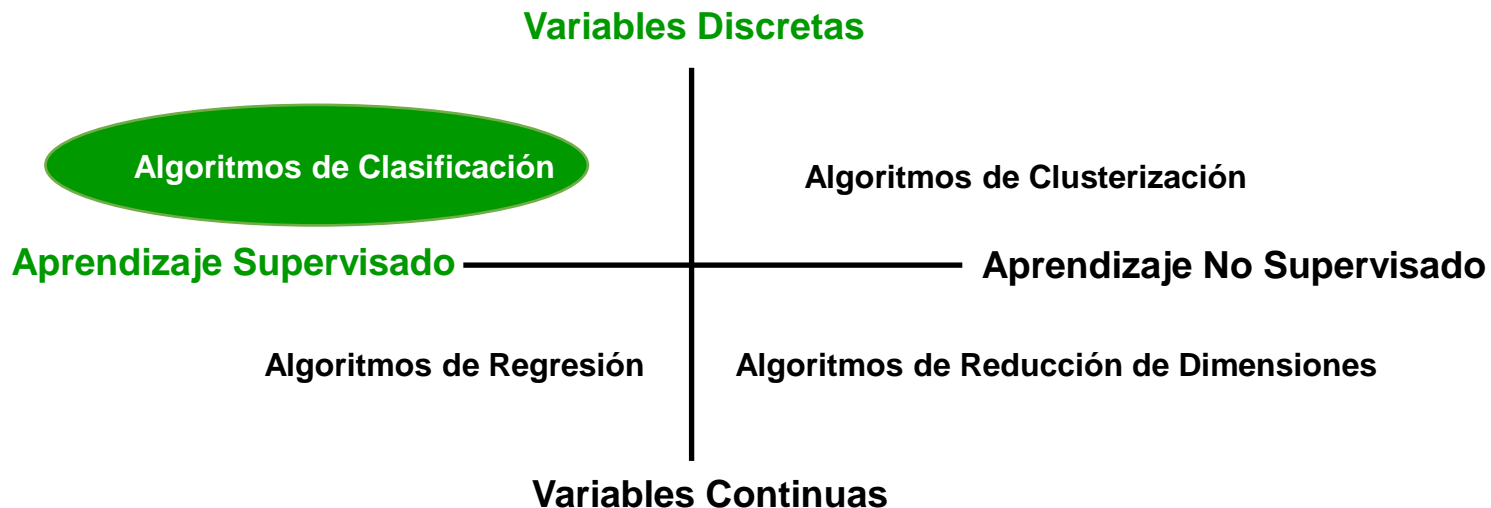


Fig. 3. Pose sorprendida de diferentes actores, y pesos asignados a salidas de nivel 3 para clasificar sorprendidos (solo se muestran 2 escalas). El peso positivo (rojo) se asigna a las configuraciones locales típicas de la categoría, peso negativo (verde) a las atípicas. Observe cómo el clasificador enfatiza las configuraciones de brazo y mano, que son características de sorpresa.



**Propuesta: Reconocer si la postura es positiva, negativa o neutral.**



**Algoritmo de clasificación a utilizar:**

✓ SVM (Support Vector Machine).

En general, para llevar a cabo el proceso de aprendizaje con el SVM, se debe realizar una serie de pasos:

1. Se reúnen los datos con los que se va a trabajar, para el caso de algoritmos utilizados en aprendizaje supervisado, es importante que estos datos tengan asociada una etiqueta indicando a que categoría pertenecen, esta etiqueta será utilizada por el algoritmo para determinar si está obteniendo los resultados esperados o si necesita ajustarse.





2. Los datos deben ser preparados para ser procesados por el algoritmo, esto es, que se encuentren en un formato apropiado, que no existan duplicados, eliminar aquellas características que no sean útiles para el algoritmo, verificar que los datos se encuentren balanceados (que se cuente con un numero de ejemplos similar para las diferentes categorías).
3. Dividir los datos en dos conjuntos, uno de entrenamiento y otro para pruebas, el conjunto de datos de entrenamiento se utilizará para entrenar el algoritmo, mientras que el conjunto de datos de prueba será utilizado para verificar que el algoritmo es capaz de generalizar y producir buenos resultados en datos con los que no ha estado en contacto previamente. Los porcentajes pueden variar, pero una buena práctica es asignar el 70% del total de datos para entrenamiento y el otro 30% para pruebas.
4. Ejecutar el algoritmo, el cual utilizará el conjunto de datos de entrenamiento para ajustar sus parámetros internos hasta lograr producir resultados satisfactorios según la tarea de que se trate.
5. Finalmente, utilizar el conjunto de datos de prueba para verificar que el algoritmo es capaz de generalizar y producir buenos resultados en datos nuevos.



## ¿Cómo se relaciona con la robótica social?

El lenguaje corporal es una de las formas básicas para la comunicación no verbal.

El psicólogo Albert Mehrabian, conocido por su trabajo pionero en el campo de la comunicación no verbal. Llevó a cabo experimentos sobre actitudes y sentimientos y encontró que en ciertas situaciones en que la comunicación verbal es altamente ambigua, solo el 7% de la información se atribuye a las palabras, mientras que el 38% se atribuye a la voz (entonación, proyección, resonancia, tono) y el 55% al lenguaje corporal (gestos, posturas, movimiento de los ojos, respiración, entre otras)

El lenguaje corporal ha sido implementado en los robots, por ejemplo: Pepper posee sensores capaces de identificar nuestro estado de ánimo, a través de la interpretación de nuestro lenguaje corporal y tono de voz. Del mismo modo, Erica imita el habla y lenguaje corporal de los humanos.

Finalmente, la comprensión del lenguaje natural humano permitirá a las máquinas entender lo que quieren decir las personas, independientemente de la forma en que lo expresen.