

Propuesta de implementación de un sistema de reconocimiento de expresiones faciales basado en flujo óptico

Carlos Herrera Muñoz¹ Alfonso Phocco Diaz ¹ Filomen Incahuanaco Quispe ¹

¹ Universidad Nacional de San Agustín

cherrera@unsa.edu.pe, aphocco@unsa.edu.pe, fincahuanaco@unsa.edu.pe

Resumen

Este artículo describe la implementación de un sistema de reconocimiento de expresiones faciales, mediante el uso de la teoría de movimiento en secuencia de imágenes, llamada flujo óptico. El problema es parte de clasificación de patrones para lo cual se empleó K-SVM, maquina de vectores soporte multi clase, para la clasificación de cada gesto facial, dentro de los cuales incluimos estado normal o neutral, miedo, tristeza, enojo, sorpresa, asco, felicidad.

1. Introducción

La interacción inteligente de la Humano-computadora o Human-computer intelligent interaction (HCII) es el campo emergente de la ciencia que tuvo como objetivo el proporcionar de una manera natural para los seres humanos el uso de las computadoras como ayudas. Se discute que para que la computadora pueda obrar recíprocamente con los seres humanos, necesita tener las habilidades de la comunicación de seres humanos. Uno de éstas habilidades son la capacidad de entender el estado emocional de la persona. La forma mas expresiva de los seres humanos de mostrar sus emociones es a través de las expresiones faciales.

Paul Ekman, connotado doctor se encargo de estudiar la relación entre los músculos faciales y los cambios que provocan en la superficie facial, identificando hasta 46 elementos, a los cuales el denomina UAs(Unidades de Acción), con los cuales se puede determinar los estados emocionales con el simple hecho de observar estas unidades, así pues nuestro equipo se interesó por analizar estas unidades de acción iniciando por el área de flujo óptico, ya que este nos permite medir el grado de movimiento ocurrido en una imagen digital de algunos o todos los puntos componentes de la misma.

2. Trabajos previos

En el trabajo de Cohen et al., 2003 se puede notar varios avances que se han hecho en la construcción de un sistema para la clasificación de expresiones faciales en secuencias de video. Introducen y prueban diversas arquitecturas, centrándose en cambios en asunciones de la distribución y estructuras de la dependencia de la característica. Lo importante de este trabajo es que de los métodos existentes, proponen una nueva arquitectura de HMMs para automáticamente dividir y reconocer la expresión en segmentos facial humana de las secuencias video. La arquitectura realiza la segmentación y el reconocimiento de las expresiones faciales que usan automáticamente una arquitectura de múltiples niveles compuesta de una capa de HMM y de una capa modelo de Markov.

Se ha tratado también con las técnicas de IA (Inteligencia Artificial) así lo demuestra el citado trabajo Chang and Chen, , en el cual usan una red neuronal que les permite clasificar las expresiones. Primero usan unas rutinas para estimar los contornos más relevantes, y puntos que rodean las zonas de cejas, ojos y labios de la imagen del rostro. de esto Chang and Chen, define 30 puntos para describir tres características faciales, teniendo como fundamento las unidades de acción propuesta por Ekman and Friesen, 1978. Han llegado a obtener un ratio de reconocimiento superior alrededor de 92.1 %, sobre las expresiones neutral, enojado y feliz que según la literatura son las más resaltantes.

Uno de los primeros en proponer una plantilla de mapeo del rostro es Kobayashi and Hara, 1972 el cual usando una red neuronal logra un ratio del 70 %, pero eso debido a que fue una de las primeros intentos.

Todos los mencionados en los párrafos anteriores tienen como base el estudio del doctor Paul Ekman y Friesen en su trabajo Ekman and Friesen, 1978.

3. Planteamiento del problema

Los sistemas de reconocimiento de rostros ya tienen un lugar importante en los sistemas biométricos para la identificación de personas, pero además de reconocer a las personas en algunas áreas se necesita que se conozca además su estado de ánimo, existen trabajos que permiten determinar el estado de ánimo empleando diversas técnicas como la Kobayashi Kobayashi and Hara, 1972, clasificadores de Bayes y modelos ocultos de Markov, sin embargo no se ha explorado mucho en las técnicas propuestas por la teoría de movimiento en imágenes también conocido como flujo óptico.

4. Temas involucrados

4.1. Flujo óptico

Para conseguir la estimación del movimiento a partir del cálculo del flujo óptico de una secuencia de imágenes, únicamente contamos como dato de partida con la variación espacial y temporal del patrón de brillo de la imagen. Por lo tanto, a diferencia del procesamiento de imagen tradicional, es imprescindible introducir la variable tiempo. En este caso se utiliza un método basado en los gradientes espacio-temporales, que determina los cambios espaciales y temporales del patrón de grises de la imagen y a partir de ellos obtiene el flujo óptico. Se basa en que la información del movimiento de los objetos está contenida en los cambios de intensidad de la imagen. Este algoritmo genera un campo vectorial completo, con un vector de movimiento por cada píxel de la imagen, lo que resulta muy útil en algunas aplicaciones. Además, la complejidad del método no depende del tipo de secuencia escogida, sino únicamente del tamaño de las imágenes. Como contrapartida presenta numerosos problemas con imágenes con mucho ruido, y bajo circunstancias concretas los resultados obtenidos no son adecuados. El proceso de cálculo del flujo óptico que se ha implementado para este proyecto se basa en el método de los gradientes y la "ecuación de restricción del Flujo Óptico" formulada por Horn y Schunck Horn and Schunk, 1981:

$$E_x \cdot u + E_y \cdot v + E_t = 0$$

Esta ecuación es válida bajo condiciones de patrón de brillo de la imagen constante, y gradiente local de intensidad lineal. Desarrollando a partir de ella llegamos a una solución de tipo iterativa para cada uno de los componentes del flujo óptico Horn, 1986:

$$u = \bar{u} - \frac{E_x(E_x\bar{u} + E_y\bar{v} + E_t)}{\lambda^2 + E_x^2 + E_y^2} \quad v = \bar{v} - \frac{E_y(E_x\bar{u} + E_y\bar{v} + E_t)}{\lambda^2 + E_x^2 + E_y^2}$$

Para la obtención de resultados válidos este algoritmo requiere un mínimo de textura en los objetos de la imagen, desplazamientos pequeños entre imágenes y el movimiento continuo de los objetos Ballard and Brown, 1982.

4.2. SVM

El Modelo de clasificación elegido es el K-SVM (Maquina de vectores soporte multiclase, pero este mismo se basa en SVM el cual:

- Presentadas en 1992. Vapnik y Chervonenkis.
- Técnica para clasificación de datos.
- Existe aprendizaje.
- Más fácil de utilizar que las redes neuronales.
- Trabajan en un "espacio de características" inducido por el kernel.
- El kernel realiza la separación y traslado de las muestras al espacio de características producto escalar genérico

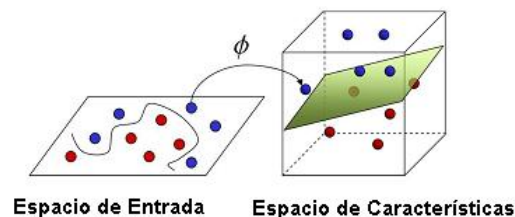


Figura 1: aaaa

- Función kernel genérica: $K(\bar{x}_i, \bar{x}_j) = (\phi(\bar{x}_i) \cdot \phi(\bar{x}_j))$
- Funciones kernel utilizadas:
 1. Lineal: $K(\bar{x}_i, \bar{x}_j) = \bar{x}_i \cdot \bar{x}_j$
 2. Sigmoide: $K(\bar{x}_i, \bar{x}_j) = \tanh(\gamma \bar{x}_i \cdot \bar{x}_j + coef)$
 3. Polinómica: $K(\bar{x}_i, \bar{x}_j) = (\gamma \bar{x}_i \cdot \bar{x}_j + coef)^{grado}$
 4. Gaussiana (RBF): $K(\bar{x}_i, \bar{x}_j) = \exp(-\gamma |\bar{x}_i - \bar{x}_j|^2)$
- tipos de SVM
 - Clasificación:
 1. C-SVM (ó SVM tipo 1)
 2. v-SVM (ó SVM Tipo 2), El parámetro $v \in (0, 1]$ y permite un control sobre el número de vectores soporte y los errores de entrenamiento
 - Regresión:
 1. ϵ -SVR (ó SVR Tipo 1)
 2. v-SVR (ó SVR Tipo 2)

- Clasificación Multiclase (Muestra con k clases) Uno-contra-resto:
 1. Se entrenan k clasificadores (una clase es la positiva y el resto la negativa).
 2. Se predice la clase para todos los clasificadores
 3. La clase asignada es aquélla con la que se consiguió mayor margen (en el caso en que se clasifique como positiva en más de un clasificador).

Uno-contra-uno

1. Se construyen $k(k-1) / 2$ clasificadores cada uno entrena datos de dos clases diferentes.
 2. Se usa la estrategia de votación para clasificar: cada clasificador binario se considera como un voto y se toma la clase con mayor número de votos.
- Estimación probabilística, extensión de SVM para estimar la probabilidad de que pertenezca a una clase u otra

4.3. OpenCV

OpenCV provee varios algoritmos de tratamiento de imágenes como son soporte de múltiples formatos de imagen, cambio de escala de imagen, cambio a escala de grises, segmentación, búsqueda de objetos circulares y rectangulares en una imagen incluso Haar cascade classifiers, que permite buscar objetos en una imagen, con lo cual se puede localizar rostros, vehículos, etc.

5. Arquitectura de la aplicación

Para iniciar como la idea principal era reconocer las expresiones en secuencias de video, primero se debía poder calcular el flujo óptico entre dos imágenes, para implementamos el algoritmo de Lucas-Kanade, seguidamente estos cambios de velocidad en módulo y dirección debíamos tomarlo de una secuencia de video, para ello empleamos OpenCV para agilizar el proceso. Estos datos obtenidos por cada expresion consta de 6-10 vectores de los

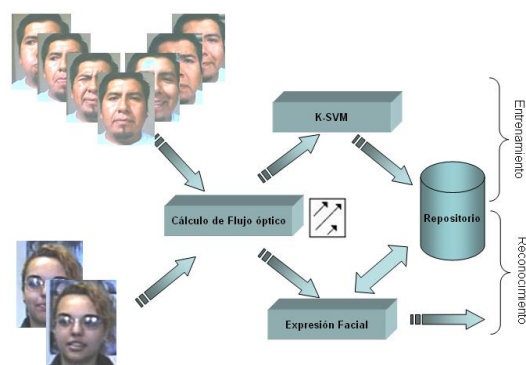


Figura 2: Esquema de trabajo del reconocedor de expresiones

módulos del flujo calculado, lo cual posteriormente se alimenta en un sistema de archivos, para posteriormente entrenar nuestra máquina de vectores soporte multi-clase.

Cuadro 1: Tiempos de demora en los entrenamientos

N°	Núcleo	Demora
1	Lineal	5min
2	Polinómica	1min
3	Gaussiana	50seg
4	Sigmoide	20seg

Para cada individuo se tomaron muestras de cada expresión, en si la toma de muestras consta de las transiciones del estado normal facial hacia los demás estados como son enojo, felicidad, etc.

Una vez conseguido las muestras suficientes, es que se implemente el renocedor de expresiones, tal como lo muestra el esquema de trabajo, los resultados de las pruebas se muestran en la tabla 1.

6. Pruebas y resultados

Para la etapa de entrenamiento se tomaron muestras de 17 individuos de los cuales 5 son del sexo femenino entre edades de 19 a 23 años, todos ellos estudiantes de pre-grado. las muestras se tomaron durante el día bajo condiciones normales de iluminación, las pruebas también se realizaron bajo las mismas condiciones, pero con individuos que no participaron en el entrenamiento.

7. Conclusiones

Este trabajo permitió comprobar que los sistemas de autenticación biométrica son cada vez mas necesarios y con ellos la especialización de los mismos, así pues es el de reconocimiento de expresiones que juega un papel muy importante en la interacción humano computador(HCI)

Las K-SVM, nos permitieron clasificar las expresiones con mucha facilidad, ya que una de sus propiedades es la de poder generalizar con pocas muestras.

Las pruebas realizadas en el desarrollo del trabajo nos permitieron observar que las emociones como son felicidad, sorpresa, enojo y asco, son las mas fáciles de detectar siendo las otras 2; tristeza y miedo son poco perceptibles debido a que sus señales demoran en su transición y la segunda por que es una mezcla de expresiones.

References

- Ballard, D. and Brown, C. (1982). *Computer Vision*. Prentice-Hall.
- Chang, J.-Y. and Chen, J.-L. Automated facial expression recognition system using neural networks.
- Cohen, I., Sebe, N., Chen, L., Garg, A., and Huang, T. (2003). Facial expression recognition from video sequences: Temporal and static modeling.
- Ekman, P. and Friesen, W. (1978). The facial action coding system: A technique for measurement of facial movement.

- Horn, B. (1986). *Robot Vision*. Ed Mc Graw-Hill.
- Horn, B. and Schunk, B. (1981). Determining optical flow. *Artificial Intelligence*, 17:185–203.
- Kobayashi, H. and Hara, F. (1972). Recognition of mixed facial expressions by neural network. *IEEE International workshop on Robot and Human Communication*, pages 381–386.