

Diseño, Implementación y Evaluación de algoritmos de procesamiento de imagen

Con la afluencia masiva de los sistemas informáticos en la sociedad, la interacción humano-ordenador (Human-computer interaction: HCI) se está convirtiendo en un aspecto cada vez más importante de nuestra vida cotidiana. A medida que los sistemas informáticos y de comunicación, progresan al nivel tecnológico, se tienen que desarrollar técnicas de HCI para permitir la utilización eficaz de los sistemas y de los flujos de información disponible. Por ejemplo, los HCI más populares hasta ahora son teclados y ratones. Estos dispositivos son muy comunes pero limitan la velocidad y la naturalidad con la que podemos interactuar con los sistemas. En los últimos años, se está produciendo un enorme impulso de investigación de nuevos dispositivos y técnicas HCI.

Una de las técnicas de HCI más maduras se basa en el reconocimiento automático del habla. Se han hecho progresos enormes en este campo y varias interfaces de voz han sido desplegadas con éxito comercial. Sin embargo, existe un interés creciente en tratar de introducir otras modalidades de HCI como técnicas basadas en el movimiento del brazo humano o gestos de las manos. Los gestos de las manos son un potente medio de interacción no verbal entre personas. Van desde simples acciones para señalar y mover objetos a comunicaciones más complejas que expresan nuestros sentimientos o incluso cualquier semántica mediante lenguaje de signos.



El reconocimiento de gestos tiene potencialmente numerosas aplicaciones incluyendo el control de ordenadores, de dispositivos móviles, de equipos de sonido y televisión, de video juegos.

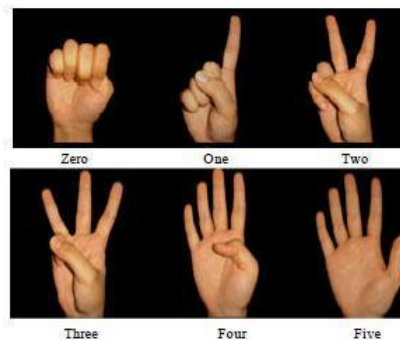




Ofrece también la posibilidad de interactuar con sistemas en entornos médicos o domésticos sin tener que preocuparse por la contaminación o la limpieza.



En este proyecto de programación vamos a concentrarnos sobre un aspecto específico de reconocimiento de gestos relacionado con el lenguaje de signos. En concreto, el objetivo final del algoritmo es el reconocimiento de números desde 0 a 5 mediante el análisis de imágenes de manos:



El trabajo de programación se va a desarrollar en dos fases. La primera fase consta de tres sesiones de trabajo en las cuales queremos definir, programar y evaluar **un sistema que permita la identificación en imágenes de zonas con piel representando potencialmente manos, dedos, brazos, etc.** En la segunda fase del proyecto (también con tres sesiones), utilizaremos el resultado de la primera fase para definir un **nuevo sistema que tiene que contar el número de dedos enseñado en las imágenes** y así reconocer el mensaje del usuario.

Para poder evaluar objetivamente y optimizar los sistemas, vamos a trabajar con dos bases de datos: Dataset I y II. Dataset I contiene imágenes simples con un fondo uniforme, mientras Dataset II contiene imágenes con fondos más complejos. En ambos casos, los resultados ideales están disponibles: Para el primer sistema, los resultados ideales se representan con máscaras binarias definiendo los píxeles de piel. Para el segundo sistema, es el número que aparece en primera posición del nombre del fichero que indica cuantos dedos se tiene que detectar en cada imagen.

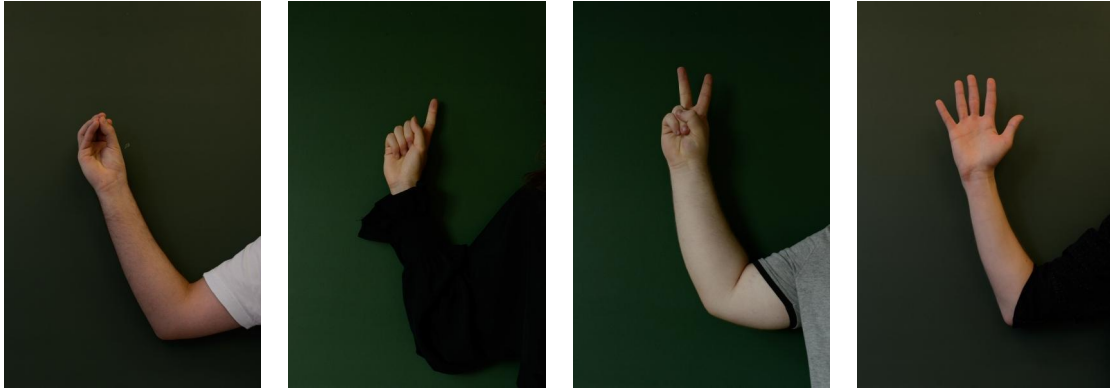


Figura 1: Ejemplo de imágenes del Dataset I

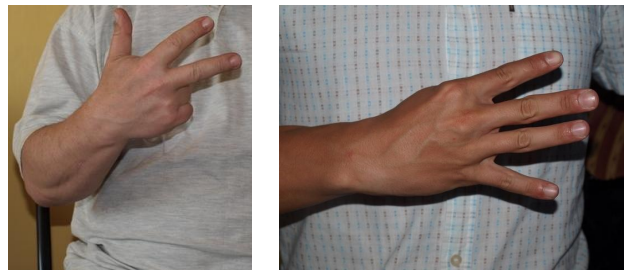


Figura 2: Ejemplo de imágenes del Dataset II

El objetivo del primer sistema es la obtención de un método simple que nos permita detectar las zonas de piel en una imagen. Por ello, nos basaremos en el color de la piel. La crominancia asociada a la piel humana es una característica bastante invariante, con poca varianza entre diferentes individuos, incluso de diferentes razas, siempre y cuando tengamos el mismo tipo de iluminación de la escena.

El sistema a desarrollar se basará en el histograma de las dos componentes de croma. Se generará un histograma modelo que deberá servir para realizar la búsqueda dentro de imágenes arbitrarias que se quieran analizar.

Breve descripción de la arquitectura del sistema

- Las imágenes de la base de datos nos van a servir primero para definir un modelo de piel y después para evaluar las prestaciones del sistema. Así que tenemos que separar las bases de datos en dos conjuntos independientes: un conjunto **de entrenamiento para la estimación del modelo** y un conjunto **de validación para la evaluación** del sistema. Podemos poner 50% de las imágenes en el conjunto de entrenamiento y el 50% restante en el conjunto de validación.
- Para caracterizar el histograma correspondiente a los píxeles de piel, el primer paso consiste en seleccionar zonas de las imágenes del conjunto de entrenamiento que contengan únicamente píxeles de piel (Fig. 3).



Figura 3. Regiones de interés (ROI) seleccionadas para la creación del modelo de piel.

- A continuación, extraemos las dos componentes de croma y calculamos el **histograma bidimensional** caracterizando los píxeles de piel. Es nuestro modelo principal del primer sistema. Una de las decisiones importantes para este primer sistema es definir el espacio de color

(RGB, YCrCv, HSV, Lab, etc.) para extraer las componentes de crominancia. Otro parámetro importante del modelo a optimizar es el número de bins del histograma.

- Una vez definido el modelo, la búsqueda de píxeles de piel se puede efectuar en las imágenes del conjunto de validación. Proponemos seguir los pasos siguientes:
 1. Transformamos las imágenes de conjunto de validación en el mismo espacio de color que el utilizado en el entrenamiento.
 2. Para cada posición (i,j), comparamos las componentes de crominancia del píxel con el histograma del modelo de piel mediante una cierta distancia. Si la distancia es suficientemente pequeña, se puede clasificar el píxel (i,j) como píxel de piel. Repetimos los dos últimos pasos para cualquier valor de (i,j). Aquí la decisión importante es definir una noción de distancia entre la crominancia del píxel y el histograma de la piel adecuada.

Evaluación de la calidad de los algoritmos:

Para poder evaluar la calidad del sistema, optimizar los parámetros (valor del umbral de decisión, número de bins de los histogramas, etc.) y tomar decisiones (sobre el mejor espacio de color, la mejor distancia) se necesita una evaluación objetiva de las prestaciones. Un enfoque clásico consiste considerar que estamos frente a un problema de clasificación con dos clases: cada píxel de las imágenes del conjunto de validación se tiene que clasificar en “Piel” o “No Piel”. Como tenemos acceso a las máscaras ideales podemos calcular los parámetros de **Precision** y **Recall** de la clasificación de los píxeles. Finalmente, mediremos la calidad final del sistema con la media armónica de Precision y Recall:

$$\text{F-Score} = 2 (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

SISTEMA 1: Sistema de detección de piel.

Para el desarrollo del primer sistema, se pide definir los siguientes algoritmos en Matlab:

- a. Implementar un script (algo1) que analice las imágenes del conjunto de entrenamiento y genere un histograma modelo de los píxeles de piel.
- b. Implementar una función (algo2) que analice una imagen arbitraria y genere una máscara binaria indicando los píxeles de piel.
- c. Implementar un script (algo3) que utilice la función algo2 para generar las máscaras de la detección de piel para un conjunto de imágenes. El conjunto de imágenes se define mediante un directorio “Images”. Durante el desarrollo del sistema se puede utilizar el directorio del conjunto de validación. Pero al final de la tercera sesión, este algoritmo se tendrá que utilizar con un nuevo directorio entregado para la evaluación final. Las máscaras se tienen que escribir en un directorio “Masks” con el mismo nombre que las imágenes procesadas.
- d. Implementar un script (algo4) que compare las máscaras obtenidas para un conjunto de imágenes arbitrarias (directorio “Masks”) y con las máscaras ideales correspondientes (directorio “Masks-Ideal”) y que calcule el F-score.

Finalmente, cada grupo deberá escribir un informe:

- a. Presentación del problema
- b. Razonamiento y justificación de las elecciones hechas
- c. Presentación de los resultados obtenidos:
 - a. F-Score en el conjunto de validación y en el conjunto entregado para la evaluación final (criterio 1)
 - b. Tiempo de cálculo^{1*} en obtener resultados para una imagen (criterio 2)
- d. Análisis de los resultados incluyendo ejemplos de buena detección y de errores. Estudio de la sensibilidad de los parámetros sobre los resultados.
- e. Finalmente, presentación de ideas para mejorar el sistema.

^{1*} Calculado en los ordenadores del laboratorio