



INTERFAZ HOMBRE-MÁQUINA BASADO EN EL ANÁLISIS DE GESTOS FACIALES MEDIANTE VISIÓN ARTIFICIAL



Jose F. Velasco, Daniel Pizarro, Javier Macías-Guarasa y Cristina Losada
Grupo GEINTRA - Departamento de Electrónica - Universidad de Alcalá
Email: {jose.velasco, pizarro, macias y losada}@depeca.uah.es

RESUMEN

- Aplicación orientada a la ayuda a la discapacidad.
- Análisis de los gestos faciales mediante visión artificial.
 - Modelado paramétrico de los gestos faciales mediante Modelos de Apariencia Activa (AAM).
- Uso del espacio invariante de iluminación para hacer robusto el algoritmo ante cambios de iluminación
 - Modelado paramétrico de las propiedades espectrales de las cámaras.

ESPACIO INVARIANTE DE ILUMINACIÓN

Basado en las siguientes suposiciones:

- Superficies Lambertianas:** el brillo aparente es igual en todas las direcciones



Reflectividad: $\sigma(A_k)$

- La luz sigue el modelo de Plank:

$$E(\lambda, T) = I c_1 \lambda^{-5} e^{\left(\frac{c_2}{\lambda T}\right)} \left[\frac{W}{m^2} \right]$$

- Sensor de ancho de banda estrecho para cada componente.
- $Q_k \delta(\lambda - \lambda_k)$

Con lo anterior, el color obtenido en cada píxel RGB es:

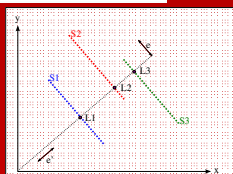
$$\rho_k = \sigma(A_k) E(\lambda, T) Q_k \delta(\lambda - \lambda_k) \quad \forall k \in \{R, G, B\}$$

Representación invariante:

$$\chi_1 = \log(\rho_1/\rho_3) = \log(s_1/s_3) + \frac{(\rho_1 - \rho_3)}{T}$$

$$\chi_2 = \log(\rho_2/\rho_3) = \log(s_2/s_3) + \frac{(\rho_2 - \rho_3)}{T}$$

Superficie Cámara y T



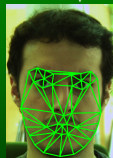
Se obtiene una representación invariante a T proyectando sobre el vector perpendicular a $\vec{e} = (e_1 - e_3, e_2 - e_3) \rightarrow \vec{e}^\perp = (\cos \theta, \sin \theta)$

$$\mathcal{L}(\rho, \theta) = \chi_1(\rho) \cos \theta + \chi_2 \sin \theta$$

MODELOS DE APARIENCIA ACTIVA (AAM)

Parámetros de forma

Describen N puntos que conforman una malla:



$$s = (u_1, v_1, \dots, u_N, v_N)$$

Obtenidos mediante entrenamiento con PCA:

$$s(p) = s_0 + \sum_{i=1}^n p_i s_i$$

Parámetros de apariencia

Describen el mapa de bits definido en el interior de la malla s_0

Obtenidos mediante entrenamiento con PCA:

$$A(x; \lambda) = A_0(x) + \sum_{i=1}^m \lambda_i A_i(x)$$

Ambos parámetros relacionados mediante:

FUNCIÓN WARP

$$x' = W(x; p)$$

Función afín definida a trozos que relaciona:

- x : Coordenadas de los píxeles dentro de s_0
- x' : Coordenadas de los píxeles dentro de $s(p)$

Ajuste del Modelo

Dada una imagen de entrada $I(x')$, ajustar el AAM es equivalente a encontrar los parámetros p y λ que minimizan la siguiente función de error:

$$\sum_{x \in s_0} [A(x; \lambda) - I(W(x; p))]^2$$

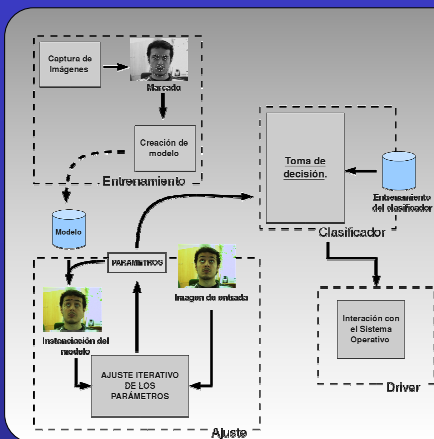
Algoritmo de optimización: Lucas-Kande

AAM + ESPACIO INVARIANTE DE ILUMINACIÓN

$$\sum_{x \in s_0} [\mathcal{L}(A(x; \lambda); \theta_1) - \mathcal{L}(I(W(x; p)); \theta_2)]^2$$

- θ_1 : Caracteriza el entrenamiento
- θ_2 : Caracteriza imagen de entrada
- θ_1 y θ_2 : Nuevos parámetros para optimización

APLICACIÓN



Entrenamiento:

- Necesita de un etiquetado manual.

Ajuste:

- Inicializado mediante Algoritmo de detección de caras de Viola & Jones.
- Ajuste iterativo hasta que Δ parametros $< \epsilon$.

Clasificador:

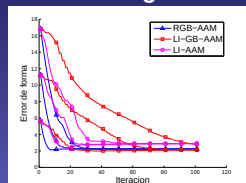
- Relaciona los parámetros de forma y apariencia con gestos.
- Es necesario un entrenamiento previo.

Driver:

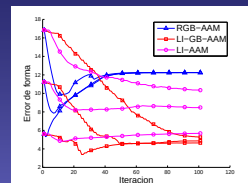
- Se comunica con el Sistema Operativo.

RESULTADOS

Ilum. Original



Con cambios de Ilum.



RGB



Light Invariant



RESULTADO APLICACIÓN



El usuario es capaz de mover el cursor con movimientos de la cara
Abrir la boca se interpreta como hacer 'click'

CONCLUSIONES

- Los AAM presentan ventajas como la identificación del usuario a través del modelo
- El Espacio invariante de iluminación hace el algoritmo más robusto ante cambios de luz.
- Con la incorporación de parámetros de offset y ganancia se mejora el ajuste y se compensan las no linealidades de la cámara.