

Detección de Rostros por medio de las Wavelets de Morlet

Jorge Guevara-Díaz¹

¹ Maestría en Ciencia de la Computación, Universidad Nacional de Trujillo

jorge.jorjasso@gmail.com

Resumen

En el presente trabajo, se aplica la Transformada Wavelet de Morlet, en la detección de rostros, además se describe una algunas técnicas comunes utilizadas para la detección de rostros en una escena, también se describe un algoritmo para el proceso de detección de rostros utilizando las wavelet de Morlet.

1. Introducción

¿Como es que podemos reconocer una cara conocida en una fotografía?. ¿Como haríamos que las máquinas puedan detectar rostros en una escena?. La solución de diversos problemas de la vida diaria y de la industria a veces requieren que un computador tenga muchas de estas capacidades.

Una solución sería inspirarse en el modelo biológico, es decir en nuestro caso en el modelo visual, las Wavelets de Morlet ofrecen una aproximación a este modelo, es por eso que se utilizarán estos para el presente trabajo.

El documento esta organizado de la siguiente manera: En la sección 2 se describe una breve introducción al problema de la detección de rostros, la sección 3 describe los métodos basados en el conocimiento para la detección de rostros, en la sección 4. Se describen los métodos basados en características invariantes, en la sección 5 se describen los métodos basados en plantillas, en la sección 6 los métodos basados en la apariencia, en la sección 7 se describe la Transformada Wavelet de Morlet y su aplicación en la detección de rostros, en la sección 8 se muestra un algoritmo que utiliza las Wavelets de Morlet y redes neuronales y finalmente en la sección 9 se muestran algunos resultados, así como la discusión de los mismos, también se detallan las conclusiones a las que se llegó en el presente trabajo.

2. Detección de Rostros

La detección de rostros por parte de una computadora es el proceso por el cual la computadora ubica los rostros presentes en una imagen o en un video, el proceso no es tan sencillo como lo haríamos los humanos, pues el reconocimiento es casi instantáneo y esto se debe al gran paralelismo existente en las redes neuronales de nuestro cerebro, esto no es tan rápido ni sencillo en una computadora secuencial, pero existen diversos métodos algunos ya tradicionales para poder realizar detección de rostros por medio de una computadora

Las técnicas existentes pueden ser agrupadas de la siguiente manera:

1. *Métodos Basados en Conocimiento* Codifican de alguna manera el conocimiento de lo que es un rostro, generalmente son atributos geometricos codificados en forma de reglas.

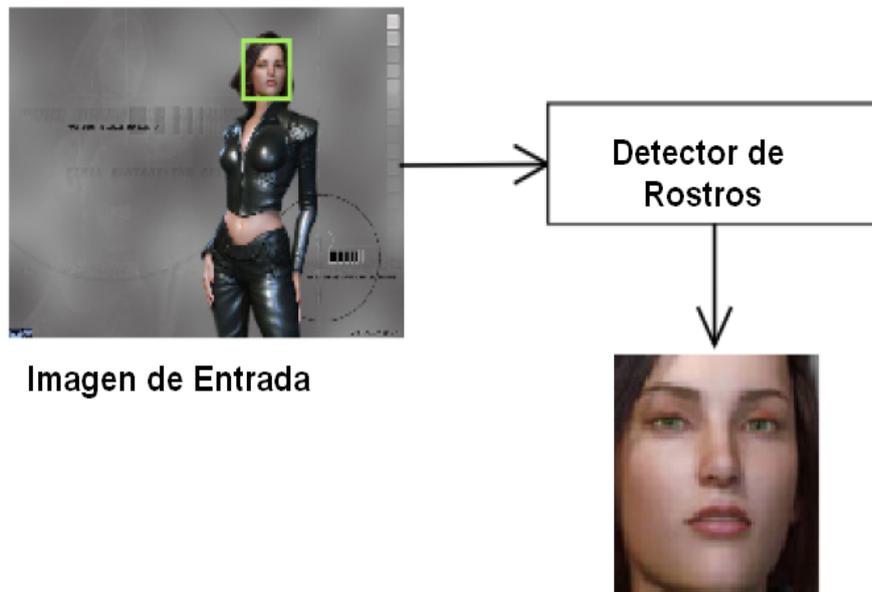


Figura 1: Proceso de detección de Rostros en un computador

2. *Métodos Basados en Características Invariantes* Son técnicas que representan los rostros independientemente de su orientación generalmente utilizan información del color y la textura de la imagen.
3. *Métodos Basados en Plantillas* Es una técnica general empleada para detectar objetos de una escena, donde el mismo es representado por una familia de curvas que representan el objeto, puede ser considerada una técnica basada en el conocimiento.
4. *Métodos Basados en la Apariencia* No necesitan conocimiento previo de la característica a ser detectada, generalmente las técnicas que pertenecen a esta clasificación necesitan varias imágenes, a partir de ellas aprenden y codifican solamente lo necesario para hacer la detección de las características de interés en este caso un rostro humano, y pertenecen a este grupo las redes neuronales, Eigenfaces y Modelos Ocultos de Markov.

Existen algunos problemas más comunes en la detección de rostros que se dan en el proceso de adquisición de las imágenes, condiciones de luminosidad, bajo contraste, objetos extraños, rostros parcialmente capturados, distorsión de la imagen, etc. A continuación algunos de los problemas más comunes:

- *Posición* Las imágenes varían de posición y orientación y esto se debe al medio de captura, o a la posición de las personas en el momento de capturar la imagen.
- *Expresión Facial* La expresión facial, por ejemplo la risa, el enojo, gestos, etc.
- *Presencia de estructuras* Elementos como barba, bigote, lentes, pueden constituir un problema.
- *Ocultación* Las imágenes pueden aparecer superpuestas por otras imágenes de rostros

3. Métodos basados en Conocimiento

Los métodos basados en el conocimiento representan las técnicas de detección de rostros basados en alguna base de reglas previas, que ha sido definida por el investigador, por ejemplo, se puede decir que el rostro tiene una forma ovalada y tiene dos ojos una nariz y una boca, además que estos elementos no se encuentran de una manera aleatoria en dentro del rostro si no más bien siguiendo una dsitribución común; éste conocimiento hace posible establecer reglas que identifiquen un rostro humano.

Las desventajas de éste método es que las reglas utilizadas si son muy generales entonces la tasa de detección de rostros de regiones de imágenes que no son rostros en realidad (falsos positivos), es muy alta, y lo inverso también ocurre, que si las reglas son muy específicas, la tasa de aceptación de rostros es muy baja.

Una técnica propuesta por [MHY02] utiliza un conjunto de reglas jerárquicas, En primer nivel este conjunto de reglas tratan de capturar en la imagen las posibles áreas candidatas a ser identificadas como rostros por medio de un algortimo que dezliza una ventana en busca de los candidatos a ser rostro , luego en esas áreas hace la búsqueda de los ojos, nariz y boca, para validar que el área candidata es un rotro.

Las reglas son aplicadas por el grado de resolución de la imagen



Figura 2: Proceso de detección de Rostros basado en el conocimiento, la imagen es vista en varios niveles de resolución, y para cada nivel se aplica cierta regla en particular

Otro método propuesto por [CK97] es encontrar sus proyecciones verticales y horizontales y basndonos en esta información y también en un conjunto de reglas detectar la ubicación de los rostros.

$$HI(x) = \sum_{y=1}^n I(x, y) \quad (1)$$

$$VI(x) = \sum_{x=1}^m I(x, y) \quad (2)$$

donde HI y VI son las proyecciones horizontales y verticales respectivamente. y considerando que la imagen tiene tamaño $n \times m$, las proyecciones verticales son utilizadas para hallar características faciales como los ojos por medio de la detección de mínimos locales de la proyección, de la misma manera se hace para la proyección horizontal.

Una desventaja de esta técnica es que la detección del rostro se hace difícil cuando existe un fondo complejo

4. Métodos basados en Características Invariantes

El color de la piel y la textura de la piel son utilizadas por éstos métodos. la principal desventaja es que las condiciones de iluminación y el ruido comprometen el desempeño de éstos algoritmos.

Cuando se utiliza la información del color de la piel, los algoritmos que utilizan toda la gamma de colores posibles tienen mejor desempeño que los que utilizan solamente escala de grises, éstos métodos tienden a formar clusters en el espacio de colores pudiendo ser modelados por medio de distribuciones gaussianas, por consiguiente esta característica puede ser utilizada para encontrar rostros en una escena

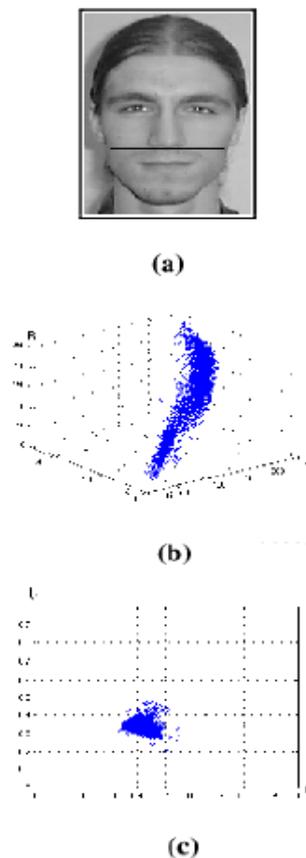


Figura 3: a) región de rostro seleccionado de la escena, b) cluster de colores en el espacio de colores, c) cluster en el espacio cromático

Un método propuesto por [MAB03], propone el siguiente algoritmo:

en primer lugar las imágenes que están en RGB, pasarlas al espacio de colores YIQ que es un sistema ampliamente utilizado en las transmisiones de televisión, donde cada componente indica matiz, luminosidad y saturación, la matriz de transformación está dada de la siguiente manera:

$$\begin{pmatrix} Y \\ I \\ Q \end{pmatrix} = \begin{pmatrix} 0,299 & 0,587 & 0,114 \\ 0,596 & -0,275 & -0,320 \\ 0,212 & -0,523 & 0,311 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (3)$$

Luego un Umbral es utilizado para detectar los pixeles que corresponden a la piel, según su experimento, Buhiyan utiliza el siguiente umbral determinado experimentalmente

$$(60 < Y < 200) \text{AND} (20 < I < 50) \quad (4)$$

luego de aplicar éste criterio se tiene todos los pixeles que corresponden a la piel humana, para ubicar de todos estos pixeles a la región del rostro , se toma la región que tenga la mayor area conectada de pixeles.

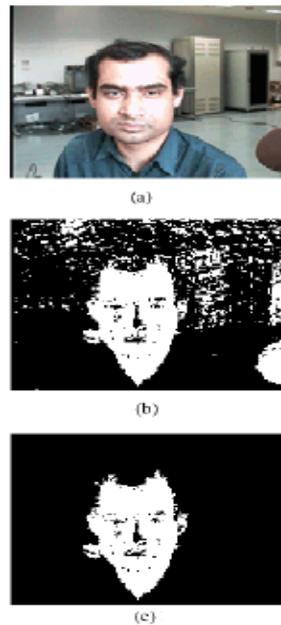


Figura 4: Detección de los rostros por medio del color de la piel a) Imagen, b) Segmentación por el color de la piel, c) Mayor región conectada de pixeles en al piel

5. Métodos basados en Plantillas

Se modela geométricamente la forma del objeto por medio de componentes básicos como círculos, elipses, etc. éstas vienen a ser las plantillas, luego se evalúa la correspondencia entre el objeto y la plantilla por medio de una función de energía.

Las principales técnicas son las plantillas deformables y los contornos activos

6. Métodos basados en Apariencia

Los métodos basados en la apariencia es una técnica que no necesita conocimiento apriori sobre las características del objeto a ser buscado en la imagen, En esta clase de algoritmos

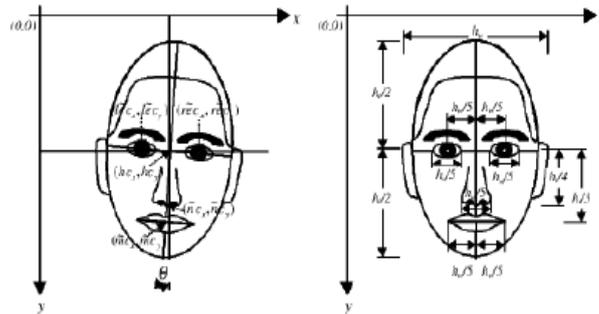


Figura 5: Modelo utilizado para detectar el rostro

surgen los conceptos de entrenamiento y de aprendizaje, una vez que la información necesaria es extraída de las imágenes sin intervención externa.



Figura 6: Imágenes originales y su proyección en el espacio de eigenfaces

Entre estas técnicas encontramos las Eigenfaces, Redes Neuronales y los Modelos Ocultos de Markov.

7. Aplicación de los wavelets en el reconocimiento de rostros

La Transformada Wavelet es una herramienta matemática que “corta” los datos, funciones o operadores en diferentes componentes de frecuencia [Dau92] y estudia cada componente a una resolución ubicada a esa escala, tienen la ventaja sobre la transformada de Fourier en que analizan las señales con una mejor localización tiempo-frecuencia.

El término wavelet proviene del campo de la sismología, donde fue bautizado por Ricker en 1940 para describir el disturbio resultante de un impulso sísmico agudo o una carga explosiva. En 1982, Morlet mostró como estos wavelets sísmicos podían ser modelados eficientemente con las funciones matemáticas que Gabor definió.

7.1. Transformada Wavelet

La transformada Wavelet está dada de la siguiente manera

$$(T^{wav} f)(a, b) = |a|^{-\frac{1}{2}} \int \delta t f(t) \psi\left(\frac{t-b}{a}\right) \quad (5)$$

la familia de wavelets se puede construir dilatando y trasladando

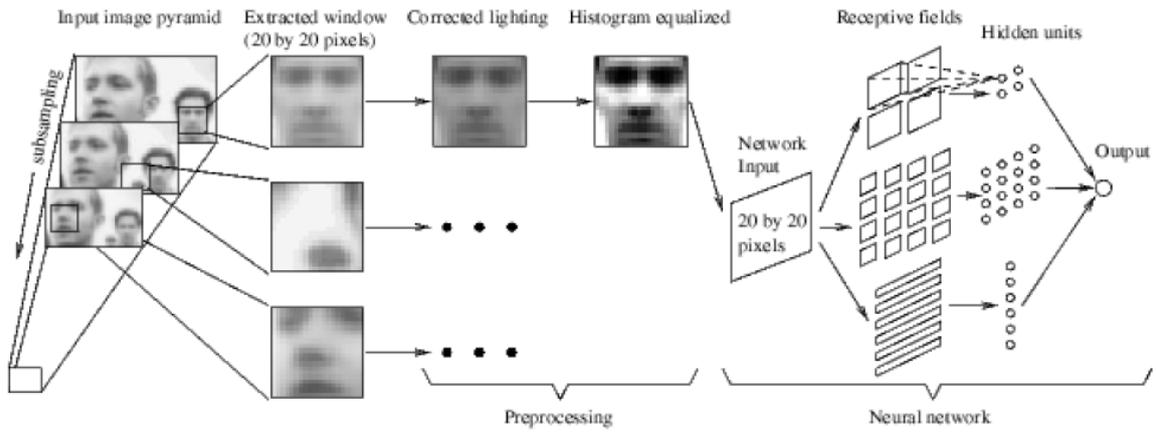


Figura 7: Red neuronal utilizada para reconocer rostros

$$\psi^{a,b}(x) = |a|^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right) \quad (6)$$

entonces la transformada wavelet continua respecto a esta familia de wavelets es:

$$(T^{wav} f)(a, b) = \langle f, \psi^{a,b} \rangle \quad (7)$$

$$(T^{wav} f)(a, b) = \int \delta x f(x) |a|^{-\frac{1}{2}} \psi \quad (8)$$

la función f puede ser recuperada de su transformada wavelet como sigue:

$$f = C_{\psi}^{-1} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\delta a \delta b}{a^2} \psi(T^{wav} f)(a, b) \psi^{a,b} \quad (9)$$

7.2. Wavelet de Morlet

LLamado también Wavelet Gaussiano Modulado o Wavelet de Gabor, está función propuesta por Gabor y ampliamente usada por dos científicos : Morlet y Kronland. está definida como la trasformada de Fourier de una gaussiana trasladada de tal manera que en promedio la función sea cero

$$\psi(x) = \pi^{-\frac{1}{4}}(e^{-i\omega x} - e^{-\frac{\omega^2}{2}})e^{-\frac{x^2}{2}} \quad (10)$$

y una respuesta en el dominio de la frecuencia

$$\psi(\omega) = \pi^{-\frac{1}{4}}[e^{-(\omega-\omega_0)} - e^{-\frac{\omega^2}{2}}e^{-\frac{\omega_0^2}{2}}] \quad (11)$$

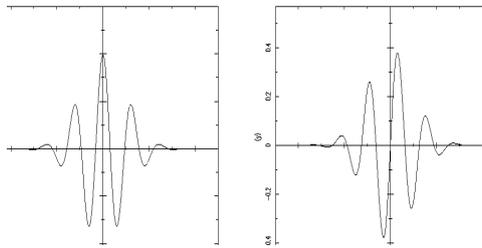


Figura 8: Parte real e imaginaria de las Wavelets de Morlet

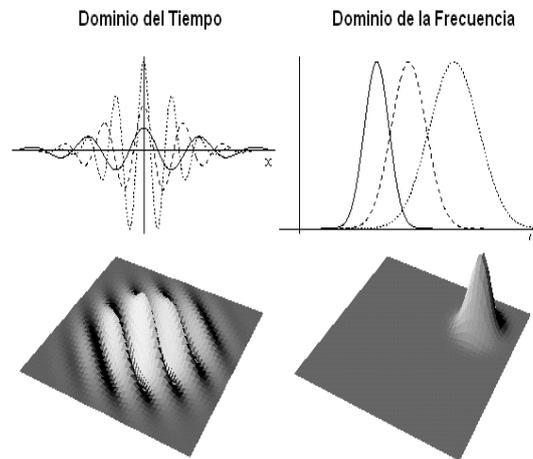


Figura 9: Parte real de las Wavelets de Morlet en el dominio del tiempo y en el dominio de la frecuencia en 1D y 2D respectivamente

7.3. Justificación

La justificación del porqué se utilizan los wavelets de Morlet para la detección de rostros esta en el hecho de que las células primarias en la corteza visual tienen campos receptivos, las cuales son pequeñas regiones altamente estructuradas; investigaciones por parte de los científicos Hubel y Wiesel describieron a esas células como detectores de bordes, algunas investigaciones más recientes como la hecha por Jhones y Palmer [JP87] mostraron que el comportamiento de las respuestas de estas células en los gatos, correspondían a medidas locales en la frecuencia. También se notó que la respuesta dependía de la frecuencia y orientación del campo visual.

En los experimentos hechos por Jhones y Palmers, la respuesta de estas células primarias fueron medidas con electrodos, los campos receptivos de estas células primarias fueron medidos por proyectar estímulos parecidos a un punto en una ventana homogénea.

En otro trabajo dos científicos Pollen y Ronner [PR81], examinan la relación de fase entre células adyacentes de la corteza visual de los gatos, ellos concluyen que las células de un par

Funciones Gabor y Células Corticales

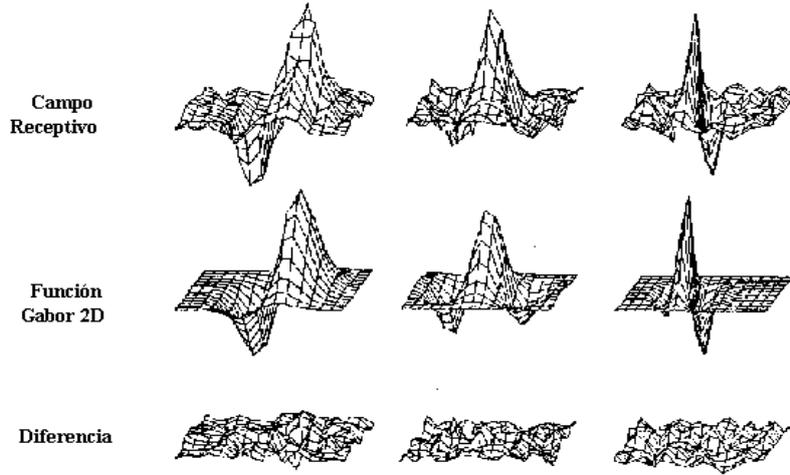


Figura 10: Adaptación de filtro de Gabor a los datos medidos por el comportamiento de las células primarias visuales, se muestra en la parte superior el experimento, en la parte central un filtro de gabor adaptativo, y en la parte inferior la diferencia entre la parte experimental y el filtro de gabor adaptativo.

de células adyacentes tienen ciertas simetrías definidas por frecuencias similares, magnitudes similares, y especificaciones similares de las direcciones: Una de ambas tiene simetría par y otra simetría impar. Esto permite modelar ambos campos receptivos de los pares de células por funciones de valor complejo.

Así la respuesta del campo de células α puede estar dada por como la correlación entre los datos de entrada, es decir la imagen $I(x)$, con el campo receptivo visual $\rho(x)$

$$\alpha_k(x_0) = \int I(x) \rho_k(x - x_0) \delta x \quad (12)$$

Una manera de modelar el campo receptivo es utilizando filtros de Gabor en 2 Dimensiones, como son funciones complejas, la parte real corresponderá a células con simetría par y la parte imaginaria a células con simetría impar.

Así biológicamente inspirado el filtro $\rho_k(x)$ es formulado [Dau88] como sigue:

$$\rho_k(x) = \frac{k^2}{\sigma^2} e^{-\frac{k^2}{2\sigma^2} x^2} (e^{ikx} - e^{-\frac{\sigma^2}{2}}) \quad (13)$$

Luego para k se tiene que es un vector que define la frecuencia central del filtro:

$$k(\nu, \mu) = 2^{-\frac{\nu+2}{2}} \pi e^{j\mu\eta} \quad (14)$$

Luego construimos la Wavelet con los filtros de Gabor, para poder analizar en distintos espacios de resolución:

$$\psi_{\nu,\mu}(x) = \frac{\overrightarrow{k}_{\nu,\mu}^2}{\sigma^2} e^{-\frac{\overrightarrow{k}_{\nu,\mu}^2}{2\sigma^2} x^2} (e^{i\overrightarrow{k}_{\nu,\mu}x} - e^{-\frac{\sigma^2}{2}}) \quad (15)$$

que es la wavelet de Morlet en 2 Dimensiones. η es el factor de escalamiento y por tanto controla la resolución de frecuencia en cada escala, μ es la orientación, y σ da el soporte del wavelet.

8. Algoritmo de Detección de Rostros mediante la Trasformada Wavelet de Morlet

Para el desarrollo del presente trabajo, se consideró usar una red neuronal del tipo Back-propagation, El algoritmo para la detección de rostros se detalla a continuación:

primero crear 40 wavelets de Morlet en 5 escalas y 8 orientaciones:

DETECCION-ROSTROS-WAVELET-MORLET(*Patrones*)

- 1 Crear 40 wavelets de Morlet en 5 escalas y 8 orientaciones
- 2 Crear red neuronal backpropagation
- 3 $I_{ent}(x) \leftarrow$ cargar las imagenes de entrenamiento
- 4 ecualizar el histograma para cada $I_{ent}(x)$
- 5 pasar al dominio de las frecuencias cada $I_{ent}(x)$ como sigue: $I_{ent}(\omega) = \int_t I_{ent}(x)e^{-j\omega t}$
- 6 pasar al dominio de la frecuencia los wavelets $\psi(\omega) = \int_t \psi(x)e^{-j\omega t}$
- 7 multiplicar $H(I_{ent}, \psi) = I_{ent}(\omega)\psi(\omega)$
- 8 sacar la inversa de fourier al resultado $Caracteristica_{ent} = \int_\omega H(I_{ent}, \psi)e^{j\omega t}$
- 9 entrenar a la red neuronal con las $Caracteristica_{ent}$ entre -1 y 1
- 10 cargar imagen para analizar
- 11 convolucionar la imagen con una plantilla de rostro $Co = I(x) * Plantilla(x)$
- 12 sacar los maximos de la imagen que son candidatos para ser rostros
- 13 recorrer con una ventana esa area en forma de espiral $I_{reg}(x) \leftarrow$ extrayendo regiones
- 14 para cada region ecualizar el histograma para cada $I_{reg}(x)$
- 15 pasar al dominio de las frecuencias cada $I_{reg}(x)$ como sigue: $I_{reg}(\omega) = \int_t I_{reg}(x)e^{-j\omega t}$
- 16 pasar al dominio de la frecuencia los wavelets $\psi(\omega) = \int_t \psi(x)e^{-j\omega t}$
- 17 multiplicar $H(I_{reg}, \psi) = I_{reg}(\omega)\psi(\omega)$
- 18 sacar la inversa de fourier al resultado $Caracteristica_{reg} = \int_\omega H(I_{reg}, \psi)e^{j\omega t}$
- 19 predecir en la red neuronal cada $Caracteristica_{reg}$
- 20 devolver todos los posibles rostros.

9. Resultados y Conclusiones

Las Wavelets de Morlet son buenos extractores de características de las imágenes, forman como una especie de firma del rostro, ya que se basan en la respuesta de como los pares de células de la corteza visual del ojo...

El algoritmo es muy limitado, pues solo trabaja con imágenes de tamaño pequeñas, y busca rostros con una medida estandar, tendra mal desempeño en imágenes grandes, también tendrá mal desempeño si se buscan localizar imágenes con rostros muy grandes, o imágenes con rostros muy pequeños. Se puede seguir investigando para proponer un algoritmo más robusto que busque las áreas candidatas a ser rostro, pero en imágenes a diferentes niveles de resolución para poder reconocer rostros pequeños y rostros muy grandes.

Los resultados obtenidos se muestran a manera de imágenes al final de documento.

Referencias

- [CK97] I. Pitas C. Kotropoulos, *Rule-based face detection in frontal views*, Em Proc. Intl Acoustics, Speech and Signal Processing **4** (1997), 2537–2540.
- [Dau88] D. Daugman, *Complete discrete 2-d gabor transforms by neural networks for image analysis and compression*, IEEE Trans. Acoustics, Speech, and Signal Processing (1988), no. 36, 1169–1179.
- [Dau92] Daubechies, *Ten lectures on wavelets*, 8 ed., Society for Industrial and Applied Mathematics, 1992.
- [JP87] J.P. Jones and L.A. Palmer, *An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex.*, J. Neurophysiol. **6** (1987), no. 58, 1233–1258.
- [MAB03] S.Yo H. Ueno Md. A.A. Buihyan, V.Ampornaramveth, *Face detection and facial feature localization for human-machine interface*, no. 5.
- [MHY02] Narendra Ahuja Ming-Hsuan Yang, David J. Kriegman, *Detecting faces in images: A survey*, IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (2002), no. 1.
- [PR81] Daniel A. Pollen and Steven F. Ronner, *Phase relationship between adjacent simple cells in the visual cortex*, Science **212** (1981), 1409–1411.
- [Sak] Omid Sakhi, *Face detection using gabor-feature extraction and neural networks*, Version : 4.1.

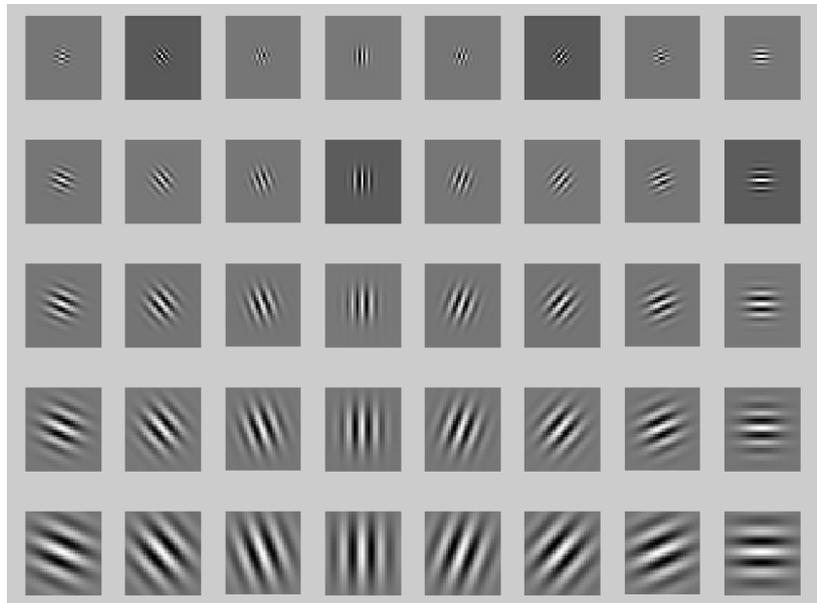


Figura 11: wavelets de gabor a 5 escalas y 8 orientaciones parte real.

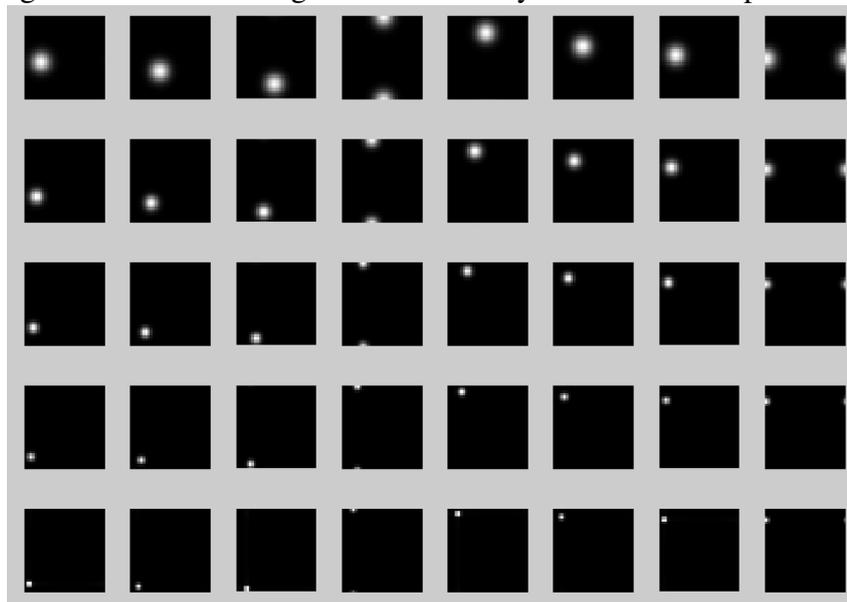


Figura 12: modulo de las wavelets de gabor a 5 escalas y 8 orientaciones en el dominio de la frecuencia.

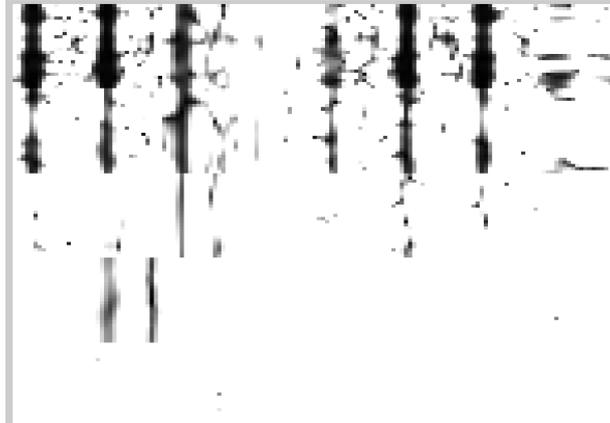


Figura 13: características obtenidas.

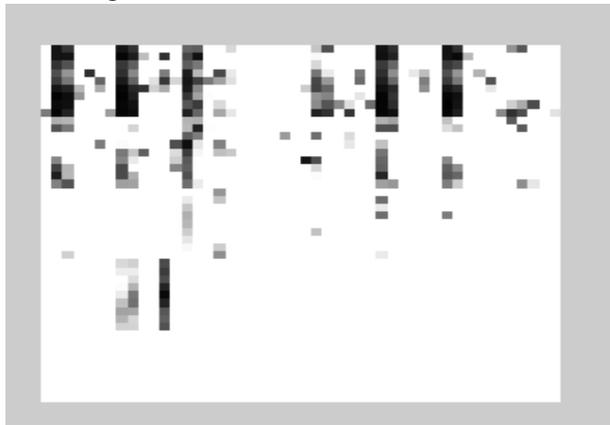


Figura 14: características recortadas a un tamaño de menor dimension.

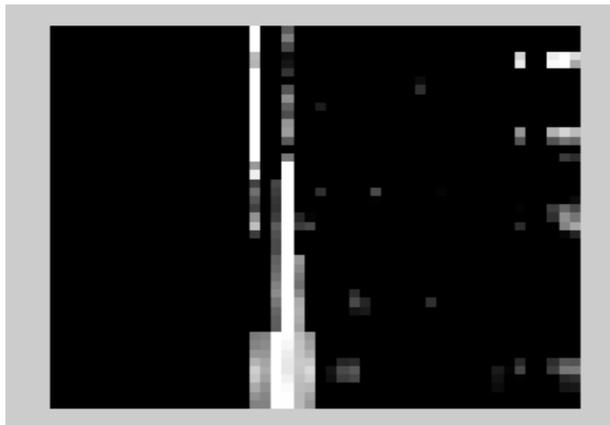


Figura 15: características presentadas a la red neuronal.

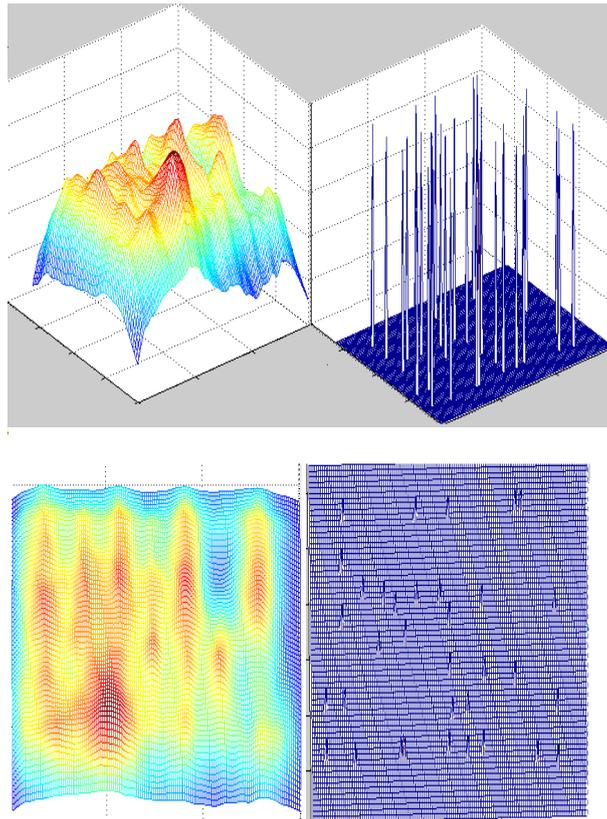


Figura 16: comparacion del resultado de la convolucion de una imagen con una plantilla de imagen y las regiones maximas extraidas.



Figura 17: resultado del algoritmo.