2013

Proceso Digital de Imágenes

David Guillermo Morales Sáez

# [ MODELO DE ATENCIÓN VISUAL DINÁMICO EN UNA SECUENCIA DE IMÁGENES]

# Índice

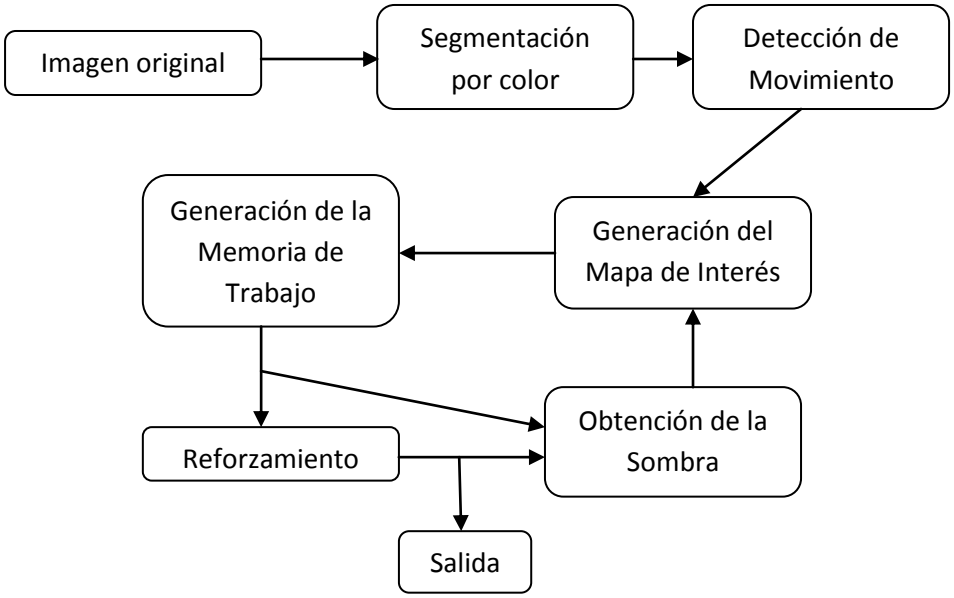
# Introducción

El reconocimiento de formas en secuencias de imágenes es un elemento clave en múltiples entornos, como en sistemas de seguridad. Tristemente, no existe un modelo universal a día de hoy, es más, no existe un modelo con una tasa de acierto superior al 90%. Aún así, se está investigando y desarrollando algoritmos que permitan localizar y seguir objetos en secuencias de imágenes.

Uno de estos algoritmos es el explicado en el documento *Dynamic visual attention model in image sequences*, por María T. López, Miguel A. Fernández, Antonio Gernández-Caballero, José Mira y Ana E. Delgado, de la Universidades de Castilla-La Mancha y de la Universidad Nacional de Educación a Distancia. En este documento, explicaré el algorimo y mostraré una implementación parcial del mismo en Matlab.

# Modelo

Este modelo busca diferenciar elementos en secuencias de imágenes que se muevan, pero no cualquier elemento, sino una serie de elementos que estén dentro de un grupo escogido. Para ello, utiliza un sistema de reforzamiento que explicaremos más adelante. Primero, veamos el diagrama del modelo:



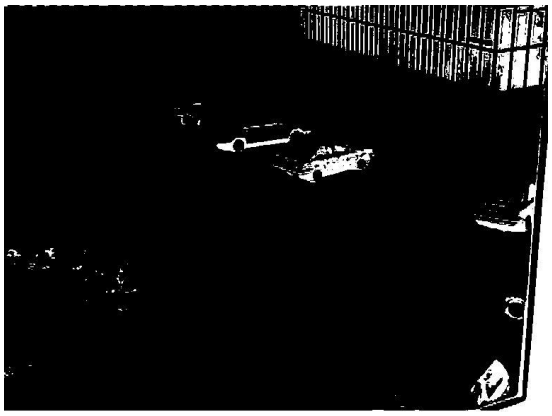

## Segmentación por Color

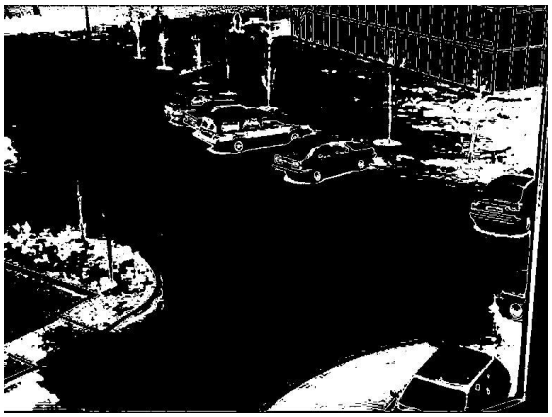
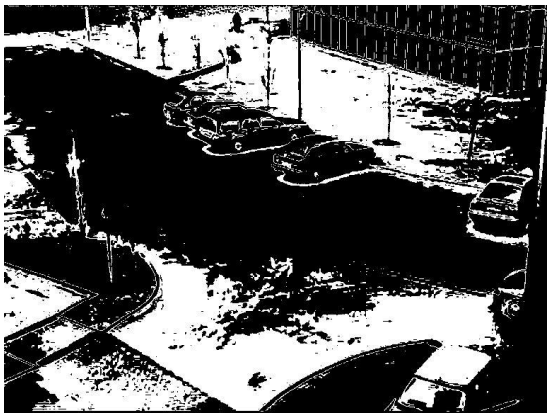
Primero, realizamos una segmentación en bandas por color, es decir, convertimos una imagen de 256 niveles de gris en una de menos niveles, para facilitar el trabajo. Un número aceptable y cuyos resultados suelen ser bastante buenos son los 8 niveles. A la imagen convertida se le llama una imagen de bandas de (en este caso) 8 niveles de grises (*eight grey-level band (GLBs)* ). Si suponemos que $GL_{max}$ es el valor máximo de la imagen inicial (generalmente 255), $GL_{min}$ el valor mínimo (suele ser 0) y S es la diferencia mínima para que se detecte un cambio entre dos niveles:

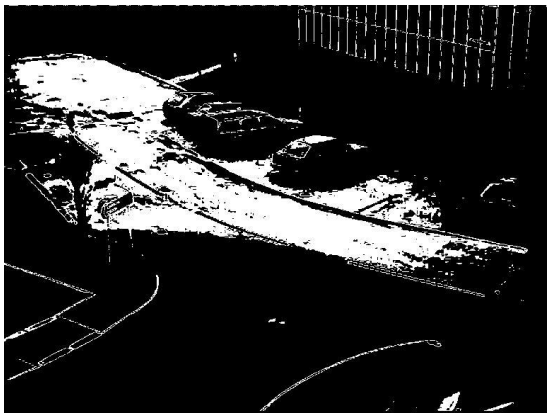
$$GL_{diff} = GL_{max} - GL_{min} + 1$$

$$GLB[x,y,t] = \begin{cases} GLB[x,y,t] \text{ si } max\left(\frac{(GLB[x,y,t-1]-1)*GL_{diff}}{n} - S, GL_{min}\right) \\ \left\lfloor \frac{GL[x,y,t]*n}{GL_{diff}} \right\rfloor + 1 \text{ en cualquier otro caso} \end{cases}$$

Con esta función, obtenemos las siguientes salidas, segmentadas por nivel:



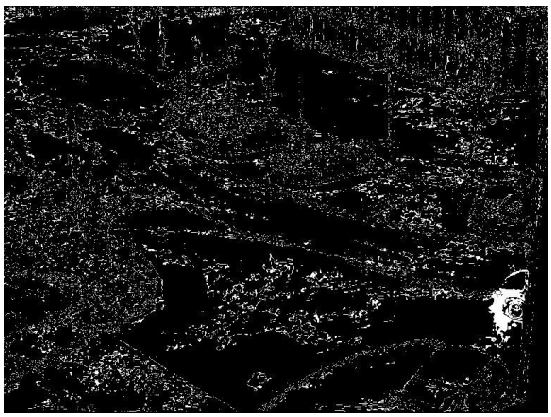


## Detección de Movimiento

Una vez segmentadas las imágenes, procedemos a extraer las componentes de movimiento. Para ello, en primer lugar debemos comprobar si los píxeles entre dos imágenes consecutivas han variado:

$$Mov[x,y,t] = \begin{cases} 0\ si\ GLB[x,y,t] = GLB[x,y,t-1] \\ 1\ si\ GLB[x,y,t]\ \neq GLB[x,y,t-1] \end{cases}$$

Con esta función comprobaremos si ha habido algún movimiento entre dos imágenes:



Por otro lado, tenemos la carga de cada píxel, que no es más que la información del movimiento como un valor cuantificado. Su cálculo es el siguiente:

$$CH_{mov}[x,y,t] = \begin{cases} CH_{min} & si\ Mov[x,y,t] = 1 \\ \min(\mathrm{CH_{mov}[x,y,t-1] + C_{mov}, CH_{max}}) & \mathrm{si\ Mov[x,y,t] = 0} \end{cases}$$

donde $C_{mov}$ es la constante de carga que va incrementando el valor de la imagen hasta un valor máximo de saturación ($CH_{max}$) siempre que no se detecte movimiento. En cambio, la carga de la imagen cae a un valor mínimo ($CH_{min}$) siempre que se detecte movimiento. Para ejemplarizarlo, veamos la salida de esta función:

## Generación del Mapa de Interés

La siguiente etapa en este viaje es la generación del Mapa de Interés. En primer lugar, habría que explicar qué es el Mapa de Interés. El mapa de Interés almacena e integra los valores obtenidos en la etapa de Detección de Movimiento y en la de Obtención de la Sombra. Con estos elementos, subdivide la imagen en tres tipos de clases:

- activos
- inhibidos
- neutrales

donde los activos son los elementos en movimiento que buscamos, los inhibidos son elementos en movimiento pero que no cumplen los requisitos solicitados y los neutrales son los elementos que no están en movimiento. El problema de esta capa es la complejidad de su implementación, ya que depende de muchos factores, desde el perfil de la sombra hasta los intereses del usuario.

## Generación de la Memoria de Trabajo

La Memoria de Trabajo unifica el Mapa de Interés con las bandas de niveles de grises, resaltando los elementos potencialmente interesantes. Para esto, comprueba cada píxel de cada banda de nivel y comprueba su valor y el del Mapa de Interés:

$$v_i[x,y] = \begin{cases} (x * NC + y) + 1 & si\ BNG[x,y,t] = i\ \wedge IM[x,y,t] = activo \\ v_{max} & si\ BNG[x,y,t] = i\ \wedge IM[x,y,t] = neutral \\ v_{min} & en\ cualquier\ otro\ caso \end{cases}$$

Una vez definidos los valores iniciales, procedemos a compararlos con sus vecinos:

$$v_i[x,y] = \begin{cases} v_{min} & si\ \ v_i[x,y] = \min(\mathrm{v_i}[\alpha,\beta]) = \mathrm{v_{min}} \\ \min(\mathrm{v_i}[\alpha,\beta]) & \mathrm{si}\ v_{min} < \min(\mathrm{v_i}[\alpha,\beta]) < v_i[x,y] \leq v_{max} \\ v_i[x,y] & si\ v_{min} < v_i[x,y] < \min(\mathrm{v_i}[\alpha,\beta]) \leq v_{max} \\ v_{max} & si\ v_i[x,y] = \min(\mathrm{v_i}[\alpha,\beta]) = v_{max} \\ \forall[\alpha,\beta]\ \mathrm{E}\ [\mathrm{x} \pm 1, \mathrm{y} \pm 1] | 0 < \mathrm{v_i}[\alpha,\beta] \leq v_{max} \end{cases}$$

Con la matriz $v_i$ definida ya con el consenso de sus vecinos, procedemos a generar la Memoria de trabajo de cada banda de nivel:

$$WM_i[x,y,t] = \begin{cases} 0 & si\ (v_i[x,y] = v_{min})\ \vee (v_i[x,y] = v_{max}) \\ v_i[x,y] & en\ cualquier\ otro\ caso \end{cases}$$

Y, finalmente, obtenemos el valor máximo para cada elemento, generando la Memoria de Trabajo final:

$$WM[x,y,t] = \arg\ \max_i WM_i[x,y,t] \qquad \forall i\ \mathrm{E}\ [1..\mathrm{n}]$$

## Reforzamiento de Atención

Cuando se genera la Memoria de Trabajo, no siempre se resaltan los elementos que más nos interesan, viéndose los elementos inhibidos. Dado que en la siguiente iteración desaparecerán, se considera a la Memoria de Trabajo como una memoria con ruido. Para reducirlo y estabilizar la imagen, utilizamos lo que se llama Reforzamiento de Atención, que no es más que un mecanismo acumulativo seguido de un umbral. Para ello, creamos una memoria llamada Mapa de Atención, cuya definición es la siguiente:

$$AM[x,y,t] = \begin{cases} max(AM[x,y,t-1] - D_{AM}, CH_{min}) & si\ WM[x,y,t] = 0 \\ min(AM[x,y,t-1] + C_{AM}, CH_{max}) & si\ WM[x,y,t] > 0 \end{cases}$$

donde $D_{AM}$ y $C_{AM}$ son constantes de descarga y recarga del mapa, respectivamente. Partiendo de este mapa, podemos empezar a definir el Foco de Atención, que resalta los elementos en los que debemos centrarnos. Para ello, empezamos definiendo los valores iniciales:

$$v[x,y] = \begin{cases} (x * NC + y) + 1 & si\ AM[x,y,t] > 0 \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

Una vez definidos los valores iniciales, vamos comparando con sus vecinos:

$$v[x,y] = \begin{cases} 0 & si\ v[x,y] = \min(\mathrm{v}[\alpha,\beta]) = 0 \\ \min(\mathrm{v}[\alpha,\beta]) & si\ 0 < \min(\mathrm{v}[\alpha,\beta]) < v[x,y] \\ v[x,y] & si\ 0 < v[x,y] < \min(\mathrm{v}[\alpha,\beta]) \\ \forall[\alpha,\beta]\ \mathcal{E}\ [\mathrm{x} \pm 1, \mathrm{y} \pm 1] | 0 < v[\alpha,\beta] \end{cases}$$

Por último, asignamos el valor al Foco de Atención:

$$AF[x,y,t] = v[x,y]$$

## Obtención de la Sombra

En la obtención de la sombra, extraemos las distintas características de los elementos almacenados en la Memoria de Trabajo. A esto se le llama Extracción de la Zona Característica de la Sombra (*Spot Shape Feature Extraction*).

$$s\_WM[v_{label}] = count(WM[x,y,t]) | WM[x,y,t] = v_{label}$$
$$w\_WM[v_{label}] = \max(y) - \min(y) \,| WM[x,y,t] = v_{label}$$
$$h\_WM[v_{label}] = \max(x) - \min(x) \,| WM[x,y,t] = v_{label}$$

De la misma forma, extraemos los elementos almacenados en el Foco de Atención:

$$s\_AF[v_{label}] = count(AF[x,y,t]) | AF[x,y,t] = v_{label}$$
$$w\_AF[v_{label}] = \max(y) - \min(y) \,| AF[x,y,t] = v_{label}$$
$$h\_AF[v_{label}] = \max(x) - \min(x) \,| AF[x,y,t] = v_{label}$$
$$hw\_AF_{[v_{label}]} = \frac{h\_AF[v_{label}]}{w\_AF[v_{label}]}$$
$$c\_AF_{[v_{label}]} = \frac{s\_AF[v_{label}]}{h_{AF[v_{label}]} * w\_AF[v_{label}]}$$

# Conclusiones

Debido a la dificultad de la definición de las capas de la Generación del Mapa de Interés, el Reforzamiento de Atención y la Obtención de la Sombra, no se ha podido comprobar completamente la funcionalidad de este sistema. Teóricamente, debería ser capaz de distinguir y mostrar aquellos elementos que deseamos ver (por ejemplo, los coches) frente a aquellos elementos que no nos resulten interesantes (véase, los peatones).