

Sistema Avanzado de Asistencia a la Conducción para la Detección de la Somnolencia

Marco J. Flores^{a,*}, José M. Armingol M.^b, Arturo de la Escalera^b

^aUniversidad Carlos III de Madrid, C/ Butarque 15, 28911, Leganés-España.

^bDepartamento de Ingeniería de Sistemas y Automática, Universidad Carlos III de Madrid, C/ Butarque 15, 28911, Leganés-España.

Resumen

En este artículo se presenta un sistema avanzado de asistencia a la conducción (SAAC) diseñado para detectar automáticamente la somnolencia y la distracción del conductor. Este sistema se compone de dos partes: una para trabajar durante el día con iluminación natural, y otra para funcionar en la noche utilizando iluminación infrarroja. Los principales objetivos son localizar el rostro y los ojos del conductor para analizarlos a través del tiempo y generar un índice de somnolencia y uno de distracción. Para ello se han utilizado técnicas de Visión por Computador e Inteligencia Artificial. Finalmente, el sistema ha sido probado con varios conductores sobre un vehículo en condiciones reales de conducción, en el día y en la noche. Copyright © 2011 CEA.

Palabras Clave:

Inteligencia Artificial, Visión por Computador, somnolencia, distracción, conductor, accidentes de tráfico, iluminación infrarroja.

1. Introducción

La excesiva cantidad de accidentes de tráfico, a causa de la somnolencia, constituye el principal motivo que ha dado lugar al desarrollo de esta investigación. Este trabajo se enmarca en los denominados SAAC, que constituyen la nueva generación de sistemas de seguridad que involucran a las nuevas tecnologías, son parte de lo que actualmente se denomina *seguridad activa*, dentro del campo de los llamados *Sistemas Inteligentes de Transporte* (SIT) (Vlacić et al., 2001; Fletcher et al., 2003; Bloemkolk et al., 2007; Hilario, 2008; Küçükay and Bergholz, 2005).

El conductor se encuentra expuesto a cometer distintos errores voluntarios y/o involuntarios que finalizan en muchos casos en accidentes. Los errores involuntarios se producen por varios motivos, entre los principales están la somnolencia y la distracción.

La somnolencia produce la disminución de la capacidad de vigilia afectando negativamente a los sentidos, por ejemplo, mermando la capacidad de percepción, reconocimiento y control del vehículo (Ji and Yang, 2002), y a la vez, incrementado la probabilidad de sufrir algún accidente o un sobresalto en la carretera.

Por región geográfica se tiene la siguiente información: En Europa, la somnolencia genera entre el 10 % y el 20 % de

los accidentes de tráfico, que producen cuantías comprendidas entre 16.000 y 32.000 millones de euros. Para el caso francés se tiene que uno de cada tres accidentes de tráfico son producidos por la somnolencia (ASFA, 2008). En Inglaterra se estima que el 20 % de todos los accidentes de tráfico en autopistas son producto de la somnolencia y el 16 % en la carretera (Hagenmeyer, 2007). Por otra parte, el German Highways Department asegura que uno de cada cinco accidentes de tráfico es causado por conducir cansado. A su vez, la Deutsche Verkehrswacht informa que uno de cada cuatro accidentes en las autopistas alemanas son producto de dormirse al volante y las distracciones.

En Estados Unidos, a causa de la somnolencia se han producido alrededor de 100.000 accidentes al año, obviamente, ellos han dejado como saldo, la cantidad de 1500 muertos y 71.000 heridos. Los costes asociados alcanzan los 12.500 millones de dólares al año (Hanmi, 2005; Chang et al., 2007). Por otra parte, (Branzan et al., 2008) presentan información de lo que sucede en Canadá, donde el 20 % de los conductores (aproximadamente 4,1 millones) han sido víctimas de la somnolencia al menos una vez en los últimos 12 meses. En Australia, los accidentes por este percance implican USD 1500 millones (Longhurst, ???).

Las personas en estado de somnolencia producen varias características visuales que pueden ser detectadas sobre la cabeza y/o rostro: frecuencia de bostezo, frecuencia de parpadeo, dirección de la mirada, movimiento de la cabeza, expresiones faciales. Tomando en cuenta estas características visuales, la Visión por Computador se presenta como la

* Av. Gral. Rumiñahui s/n, Campus Politécnico ESPE, Sangolquí, Ecuador
Correos electrónicos: mjflores@ing.uc3m.es,
mjflores@espe.edu.ec (Marco J. Flores), armingol@ing.uc3m.es
(José M. Armingol M.), escalera@ing.uc3m.es (Arturo de la Escalera)

tecnología más prometedora para abordar este problema (Wang et al., 2006). En este sentido, este artículo presenta un sistema para detectar automáticamente la somnolencia del conductor y paralelamente su distracción, haciendo uso de las nuevas tecnologías.

Por lo tanto, en este trabajo se presenta un sistema embarcado en un vehículo para analizar al conductor en tiempo real, donde las principales aportaciones se centran en detectar los ojos y el rostro y seguirlos, tanto en el día y la noche, para seguidamente calcular los índices de somnolencia y distracción.

El artículo se encuentra organizado de la siguiente manera. La sección 2 presenta un extenso estado del arte, donde se revisan los trabajos más importantes en este campo del conocimiento. En la sección 3 se introduce el método propuesto en función de la clasificación derivada a partir del estado del arte. Luego, la sección 4 está dedicada al sistema para la conducción diurna, donde se presentan varios módulos destinados a detectar y seguir el rostro y los ojos, y calcular los índices de somnolencia y distracción. A continuación, en la sección 5 se describe el sistema para la conducción nocturna, donde adicionalmente se ha incorporado un sistema de iluminación infrarroja para generar imágenes donde la pupila se encuentre resaltada. Finalmente, las conclusiones y trabajos futuros son presentadas en la sección 6.

2. Estado del arte

El uso de la información visual para conocer el estado de somnolencia del conductor constituye la metodología más apropiada para tratar esta problemática (Kutila, 2006; Knipling and Wierwille, 1994; Chang et al., 2007), por ser fiable, no invasiva y no intrusiva, motivos por los cuales no interfiere en la conducción. En este sentido, los ojos generan valiosa información para determinar el nivel de somnolencia y uno de los índices más utilizados es el denominado PERCLOS que relaciona el porcentaje de cierre del párpado como una medida de fatiga en tiempo real (NHTSA, 1998; Wang et al., 2006).

Los sistemas que utilizan esta metodología se pueden clasificar de varias maneras; en este caso concreto se hará por el tipo de iluminación, porque parece la más adecuada dada la complejidad del problema, las condiciones de conducción (diurna y nocturna) y la bibliografía recopilada durante la investigación. Así se tienen:

1. Métodos en el espectro visible, destinados a la conducción diurna, aprovechando la iluminación natural.
2. Métodos en el infrarrojo cercano, para la conducción nocturna, utilizando una fuente artificial de iluminación.

2.1. Métodos en el espectro visible

Para analizar la somnolencia del conductor, (Horng et al., 2004) presentan un sistema que utiliza la información del color de la piel para localizar el rostro, trabajando con el espacio de color HSI , sobre el cual construyen un modelo para identificar el color de la piel humana. Posteriormente, con la información de los bordes localizan los ojos para obtener los modelos a seguir, y utilizando la técnica de correlación construyen un

método para su seguimiento (dynamical template matching, DTM). Además, usando la información del color de los ojos identifican su estado, abierto o cerrado, y calculan el índice de somnolencia, tomando como umbral cinco fotogramas consecutivos de ojos cerrados. En esta misma tendencia se encuentra el trabajo de (Tian and Qin, 2005) al utilizar la información del color de la piel como parte fundamental de su sistema. En este caso el espacio de color elegido es el $YCbCr$; el mismo que es ampliamente utilizado para identificar zonas de piel, donde se presta especial atención a las componentes Cb y Cr que son las que mejor capturan las características que representan al color de la piel humana. Las funciones de proyección vertical y horizontal son empleadas para localizar el rostro y los ojos, respectivamente. Luego, utilizando la función de complejidad sobre la imagen del contorno de los ojos el sistema calcula el estado del conductor basado la frecuencia de cierre de los ojos.

Otro trabajo de relevancia es el de (Dong and Wu, 2005) quienes también aprovechan las propiedades del espacio de color $YCbCr$; estos investigadores construyen un modelo de probabilidad del color de piel basado en la función de distribución normal bivalente sobre las componentes Cb y Cr . Una vez localizados los ojos proceden a calcular el índice de somnolencia utilizando la distancia entre los párpados para clasificar el estado del ojo entre abierto y cerrado; donde el índice de somnolencia está caracterizado por cinco fotogramas consecutivos de ojos cerrados. Por otra parte, (Rongben et al., 2003) trabajan con el espacio cromático rg sobre el cual construyen un modelo de probabilidad para modelar el color de la piel utilizando la función de distribución gaussiana bivalente. Basados principalmente en el color localizan el rostro y posteriormente los ojos utilizando restricciones sobre el tamaño y la posición. Para el seguimiento de los ojos utilizan el método de correlación, similar al trabajo de (Horng et al., 2004), junto con la distancia de Hausdorff (Rogers, 1998). Posteriormente, utilizando una red neuronal artificial (RNA) clasifican el estado del ojo, para ello extraen las características más sobresalientes de los mismos a través de los wavelets de Gabor.

Dos trabajos importantes en el campo del análisis del conductor han sido presentados por (Jiangwei et al., 2004a,b). El primero de ellos, constituye un método para analizar el estado de los ojos empleando una RNA que toma como entradas varias características geométricas del ojo. Estas características se obtienen directamente de la imagen del ojo y se representan a través del vector $Z_{ojo} = (H_{max}, W_{max}, H_{centro}, h_{L3}, h_{R3}, H_C)$, donde H_{max} es la distancia máxima entre el párpado superior e inferior, W_{max} es el ancho máximo del ojo, H_{centro} es la distancia entre el párpado superior y el centro del iris. Además, el ojo forma un arco cuando se encuentra en estado normal, una línea al estar adormecido y una concavidad en sueño. A fin de expresar estas características se definen tres valores h_{L3} , h_{R3} , H_C . En su segunda aportación, trabajan sobre el análisis del estado de la boca, construyendo el vector de características $Z_{boca} = (H_{max}, W_{max}, H_m)$, donde H_{max} es la distancia máxima entre el labio superior e inferior, W_{max} es el ancho máximo de la boca y H_m es la distancia entre el labio superior e inferior cuando la boca se encuentra en actividad. Con esta información

entrenan una RNA. Los estados en los que se puede encontrar la boca son tres: dormido, hablando y cerrado. Para localizar los ojos y la boca, en sus dos aportaciones utilizan el espacio de color dicromático rg y la función de distribución normal bivalente.

(Hayami et al., 2002) construyen un sistema para detectar somnolencia a través del análisis de los ojos, introduciendo un nuevo índice que se denomina *tasa de movimientos verticales del ojo*, que depende del índice PERCLOS, viene dado por $V = \frac{V_c}{1-P}$, donde V_c es el número de movimientos verticales del ojo y P es el valor de PERCLOS. V denota la frecuencia del movimiento vertical del ojo para 60, 30 y 15 segundos cuando el ojo está abierto. Este trabajo ha sido desarrollado sobre un simulador que consta de un casco y un soporte que alberga una cámara que adquiere directamente las imágenes de la zona de los ojos, con lo cual, no es necesario ningún método de detección ni de seguimiento.

2.2. Métodos en el infrarrojo cercano

Entre las principales investigaciones que se pueden citar en este caso se encuentran los trabajos de (Ji and Yang, 2001, 2002), (Ji et al., 2004), (Zhu et al., 2002a,b) donde los investigadores presentan un sistema de detección de la somnolencia basado en iluminación infrarroja y visión estéreo. Para localizar la posición de los ojos utilizan las propiedades físicas que presenta la pupila frente a la iluminación infrarroja. En el rango comprendido entre los 700 nm. y 900 nm. la pupila adquiere un brillo característico que la diferencia del resto de objetos, fenómeno que permite reducir el tiempo de cómputo para su localización. En todos los casos calculan los índices PERCLOS y AECS (average eye closure speed). Otro trabajo importante es el presentado por (Grace et al., 1998) para medir la velocidad de parpadeo, tomando como base la iluminación infrarroja. Los investigadores (Bergasa et al., 2004, 2006) también han desarrollado un sistema no invasivo que utiliza iluminación infrarroja, diseñado para medir el nivel de vigilancia usando un autómata de estado finito (FSM) (Brookshear, 1983) con seis estados para calcular varios índices, entre ellos, PERCLOS. Este sistema también es capaz de estimar la distracción utilizando un clasificador difuso que toma como entrada la posición de los ojos y el rostro.

2.3. Comentarios al estado del arte

- En el caso de los sistemas en el espectro visible se puede decir que no existe una estrategia definida y única, cada investigador propone una técnica distinta. A pesar de ello, se puede concluir que la mayoría utilizan técnicas de color para localizar y analizar al rostro y los ojos. Esta técnica es fácil de implementar, pero tiene el inconveniente de ser altamente dependiente de los cambios de iluminación, que la hace inadecuada para aplicaciones sobre vehículos reales. La somnolencia está caracterizada principalmente por el umbral de cinco fotogramas consecutivos de ojos cerrados y PERCLOS.
- Los sistemas diseñados para trabajar en la noche tienen un gran inconveniente en aplicaciones sobre

vehículos reales, donde los cambios de iluminación y las interferencias externas, etc., generan inestabilidad, debido a que el brillo de las pupilas puede desaparecer. Un caso típico ocurre cuando el conductor es iluminado por otro vehículo que venga en dirección contraria, haciendo que las pupilas se mimetizan con el resto del rostro. Para medir la somnolencia, en todos los sistemas se ha implementado el índice PERCLOS.

3. Diseño del sistema de detección de la somnolencia

En este artículo se presentan dos sistemas para detectar la somnolencia del conductor siguiendo la clasificación exhibida en la sección anterior, paralelamente, se incorpora en cada uno de ellos un módulo para analizar un tipo de distracción basado en la orientación del rostro. De esta manera, se plantea un sistema para trabajar en conducción diurna y uno para trabajar en conducción nocturna.

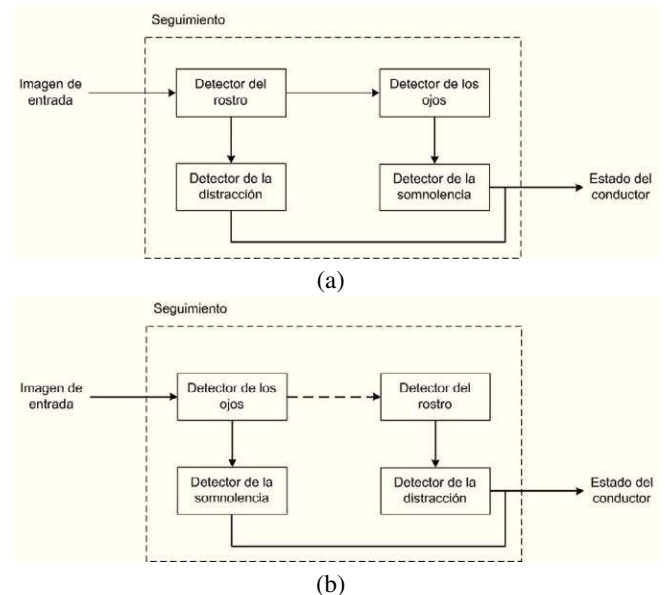


Figura 1: Apreciación global del sistema de detección de la somnolencia y la distracción, (a) sistema en conducción diurna, (b) sistema en conducción nocturna.

El esquema de los sistemas es presentado en la figura 1, donde se identifican los siguientes módulos: (i) detección del rostro, (ii) detección de los ojos, (iii) seguimiento del rostro, (iv) seguimiento de los ojos, (v) detección de la somnolencia y, (vi) detección de la distracción.

Cada una de estas partes serán descritas detalladamente en las próximas dos secciones. La diferencia entre estos sistemas radica en la incorporación de un sistema de iluminación infrarroja en el caso de la conducción nocturna. En los dos casos, se trabajará sobre imágenes monocromáticas tomadas con el sistema de percepción instalado en un vehículo real.

Este vehículo, denominado Ivvi (Intelligent Vehicle based on Visual Information), constituye una plataforma experimental

diseñada para desarrollar SAAC, figura 2 (a), alberga un sistema de procesamiento que consta de tres ordenadores instalados con MS Windows XP, figura 2(b), destinados a procesar la información visual proveniente del sistema de percepción.

En la actualidad Ivvl alberga cinco sistemas basados en Visión por Computador para analizar las líneas de la carretera, los peatones, los vehículos, las señales de tráfico, y el módulo presentado en éste trabajo para monitorizar al conductor.

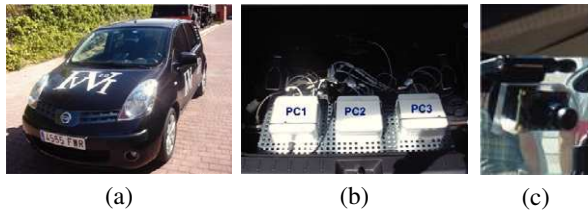


Figura 2: Ivvl, (a) vehículo experimental, (b) sistema de procesamiento y (c) cámara.

4. Sistema de detección de la somnolencia y la distracción en conducción diurna

En este caso se utilizará la información proveniente de una cámara CCD de uso común. Constituye una parte del sistema de percepción de Ivvl, ver figura 2 (c), el mismo que será descrito más adelante. Actualmente, se encuentra ubicado bajo el parasol del conductor, posición donde no obstruye al conductor.

4.1. Detección del rostro

Un método que ha generado buenos resultados para detectar el rostro es el algoritmo propuesto por Viola & Jones (VJ), que combina adecuadamente la imagen integral, el algoritmo de aprendizaje AdaBoost (Freund and Schapire, 1995) y la cascada de clasificadores, que es un conjunto de clasificadores débiles que permiten ir descartando zonas de la imagen que no son de interés (Viola and Jones, 2001, 2002a,b), cuyo resultado es un detector de objetos que trabaja en tiempo casi real.

Así, este algoritmo genera un rectángulo denominado región de interés (ROI) que localiza la posición del rostro a través de las coordenadas de la esquina superior izquierda (x_0, y_0) y sus dimensiones de ancho y alto (w, h) , ver figura 3 (a).

A pesar de las ventajas que presenta VJ, éste no es absolutamente fiable debido a que no es capaz de localizar el rostro cuando no se encuentra dentro de los rangos permitidos en la etapa de entrenamiento. Esto incluye giros laterales, giros hacia arriba o hacia abajo; todos ellos habituales durante la conducción y que se producen en situaciones como mirar el espejo retrovisor y/o los espejos laterales, etc. Esta dificultad será analizada y corregida más adelante, en la sección seguimiento.

4.2. Detección de los ojos

Localizar la posición de los ojos, sobre una imagen, es un trabajo altamente complicado porque involucra muchos parámetros que dificultan la construcción de un método robusto (Hansen and Ji, 2010; Guo and Guo, 2009; D'Orazio et al., 2004). Entre los principales inconvenientes se pueden citar:

- Las diferentes formas que adopta el ojo, debido principalmente a dos factores, la necesidad biológica de parpadear y el movimiento del rostro,
- Los cambios de iluminación, en especial en ambientes al aire libre, como es el caso de la conducción, y
- Una característica propia de los ojos, el color del iris (incluye a la pupila), que dependiendo de la raza, toma varios colores como el marrón, negro, azul, etc.

Todo esto ha llevado a plantear el siguiente procedimiento para localizarlos, utilizando la información geométrica del ojo y antropométrica del rostro, en conjunto con un algoritmo estadístico.

Las coordenadas de los rectángulos iniciales que contienen a los ojos vienen dadas por las ROI_L para el ojo izquierdo y ROI_R para el ojo derecho. Sus coordenadas se calculan a través de las ecuaciones (1), (2) y (3), respectivamente, ver figura 3 (b).

$$(u_{0L}, v_{0L}) = \left(x_0 + \frac{w}{6}, y_0 + \frac{h}{4} \right) \quad (1)$$

$$(u_{0R}, v_{0R}) = \left(x_0 + \frac{w}{2}, y_0 + \frac{h}{4} \right) \quad (2)$$

$$(w_L, h_L) = (w_R, h_R) = \left(\frac{w}{3}, \frac{h}{4} \right) \quad (3)$$

Para extraer una muestra aleatoria de los píxeles que forman el ojo se considera el siguiente algoritmo, cuyo objetivo será resaltar dichos píxeles. Esta muestra se formará con la información proveniente de tres imágenes que se denominarán B, G y L.

Sea I la imagen de entrada e $I(x, y)$ el valor del píxel en la escala de grises, es decir, $I(x, y) \in [0, 255]$ en el punto de coordenadas (x, y) , entonces a partir de aquí se tiene:

1. Generar la imagen J a través de (4):

$$J(x, y) = \frac{I(x, y) - m}{\sigma} \quad (4)$$

donde m es la media y σ es la desviación estándar, estos valores son calculados sobre los rectángulos descritos anteriormente.

2. Generar la imagen H usando (5):

$$H(x, y) = \begin{cases} J(x, y) - 256\delta_1 & \text{si } J(x, y) \geq 0 \\ 256\delta_2 + J(x, y) & \text{si } J(x, y) < 0 \end{cases} \quad (5)$$

donde $\delta_1 = \max(0, \text{ceil}(\frac{J(x, y)}{256}) - 1)$,

$\delta_2 = \max(1, \text{ceil}(\frac{|J(x, y)|}{256}))$ y $\text{ceil}(t)$ es la función que

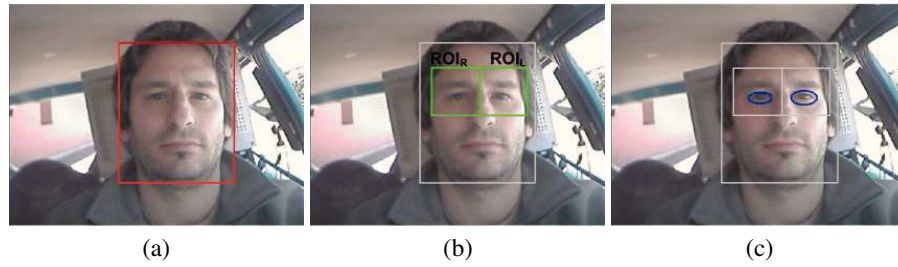


Figura 3: (a) Resultados del método de VJ para localizar el rostro, (b) rectángulos preliminares para detectar los ojos, (c) detección de los ojos.

retorna el menor entero mayor que t . La validez de las variables δ_1 y δ_2 queda justificada porque los valores de la imagen (matriz) $J(x, y)$ son números reales. El intervalo para δ_1 es $\{0, 1, \dots\}$ mientras que para δ_2 es $\{1, 2, \dots\}$.

3. Generar la imagen binarizada B , a partir de la imagen H , a través de la siguiente ecuación:

$$B(x, y) = \begin{cases} 255 & \text{si } H(x, y) \geq \kappa \\ 0 & \text{caso contrario} \end{cases} \quad (6)$$

donde el umbral κ se calcula automáticamente con el método de Otsu (Otsu, 1979), ver figura 4 (b).

4. Calcular la imagen gradiente, G , utilizando el operador de Sobel, seguido de un realce de bordes, figura 4 (c).

$$G = \sqrt{\|I * S_x\|^2 + \|I * S_y\|^2} \quad (7)$$

donde $\|\cdot\|$ es la norma euclídea y las máscaras son:

$$S_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, S_y = -S_x^T \quad (8)$$

donde T representa la transpuesta de una matriz.

5. Calcular la imagen logaritmo, L , con el objetivo de realzar los píxeles que corresponden al iris y que constituyen la parte central del ojo (Wu et al., 2004), figura 4 (d).

$$L(x, y) = \log(1 + I(x, y)) \quad (9)$$

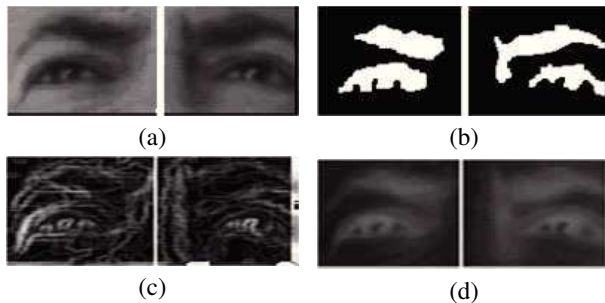


Figura 4: Imágenes utilizadas para localizar la posición de los ojos: (a) imagen original, (b) imagen B , (c) imagen G y (d) imagen L .

Sobre la muestra aleatoria se ajusta una elipse, que constituye un modelo paramétrico que representa al ojo. Para obtener sus parámetros se utiliza el algoritmo *EM* (Expectation Maximization) (Martinez and Martinez, 2002).

Para ejecutar el algoritmo *EM* se precisa de un punto inicial (parámetros de inicialización) para empezar el proceso de iteración, este punto inicial se calcula con el conocido método de mínimos cuadrados. Para cumplir las restricciones de tiempo real se permite un número acotado de iteraciones, en este caso particular se consigue con 10 iteraciones sobre una muestra en dos dimensiones y una gaussiana. La figura 5 presenta varios ejemplos que validan la propuesta, sobre ojos abiertos y ojos cerrados.



Figura 5: Varios ejemplos donde se localiza la posición de los ojos con el método propuesto.

4.3. Seguimiento del rostro y los ojos

Existen varias razones para desarrollar el seguimiento del rostro y los ojos, entre ellas se tienen:

- El método de VJ, a pesar su gran adaptación de trabajo en ambientes con iluminación variable, presenta ciertos inconvenientes que generan inestabilidad en el sistema y lo inducen a cometer errores.
- Otra razón es la necesidad de conocer la posición del rostro y los ojos continuamente en cada instante, es decir, en cada fotograma.
- El tercer motivo es satisfacer las condiciones de tiempo real, reduciendo el espacio de búsqueda y en consecuencia disminuyendo el tiempo de cómputo.

Para desarrollar algoritmos de seguimiento en Visión por Computador se han propuesto varias estrategias, entre ellas, el

algoritmo Condensation (Isard, 1998; Isard and Blake, 1998), que utiliza el filtro de partículas (Doucet et al., 2001; Ristic et al., 2004) para evaluar las complejas funciones de densidad de probabilidad generadas en el modelo probabilístico que evoluciona a través del tiempo.

En la estructura del algoritmo Condensation, la densidad a posteriori $p(x_t|z_{1:t})$ es representada por una muestra (partículas) $\{s_t^{(n)}, n = 1, \dots, N\}$ y la densidad de las observaciones dado el estado es aproximada por los pesos $\{\pi_t^{(n)}, n = 1, \dots, N\}$. Con esto se tiene la aproximación:

$$\pi_t^{(n)} \approx p(z_t|x_t = s_t^{(n)}) \quad \text{para } n = 1, \dots, N \quad (10)$$

que será utilizada en las siguientes secciones.

1. Seguimiento del rostro

El modelo dinámico del sistema para el seguimiento del rostro es un modelo autoregresivo de segundo orden $AR(2)$, que es representado por la ecuación (11).

$$x_t = A_2 x_{t-2} + A_1 x_{t-1} + \xi_t \quad (11)$$

donde A_2 y A_1 relacionan el estado presente con los estados anteriores y son similares a las descritas por (Isard, 1998; Doucet et al., 2001), ξ_t es un ruido blanco. El modelo de observación viene dado por (12).

$$z_t = H x_t + \eta_t \quad (12)$$

donde H relaciona el estado con las observaciones actuales, su estructura está dada en (Doucet et al., 2001), y η_t es un ruido blanco.

El vector de estado corresponde a la posición del rostro dentro de la imagen, su velocidad de desplazamiento y las dimensiones que ocupa el rectángulo que lo contiene, ecuación (13). El vector de medida se compone de la posición y las dimensiones del ancho y alto del rectángulo que contiene al rostro, ecuación (14).

$$x_t = (x, y, u, v, w, h) \quad (13)$$

$$z_t = (x, y, w, h) \quad (14)$$

Una vez planteado el modelo de seguimiento del rostro, resta introducir un método para evaluar la función de densidad de las observaciones. En este caso concreto su estimación dependerá del valor generado por una RNA (Looney, 1997; Evgeniou et al., 2000) al ser evaluada sobre cada una de las partículas. Con lo cual, la evaluación de la densidad de las observaciones viene dada por (15).

$$p(z_t|x_t) \propto \text{Cexp}(-\delta(R_1, R_2)) \quad (15)$$

donde $R_1 = R_1(x_{01}, y_{01}, w_1, h_1)$ es la región hipotética del rostro, $R_2 = R_2(x_{02}, y_{02}, w_2, h_2)$ es la región que contiene al rostro y que es indicada por la RNA, y $\delta(\cdot, \cdot)$ se define por (16).

$$\delta(R_1, R_2) = \frac{\Delta_x + \Delta_y}{\Delta_w + \Delta_h + 1} \quad (16)$$

donde $\Delta_x = |x_{01} - x_{02}|$, $\Delta_y = |y_{01} - y_{02}|$, $\Delta_w = \max(w_1, w_2)$ y $\Delta_h = \max(h_1, h_2)$.

Para inicializar el algoritmo Condensation es necesario conocer el número de partículas y la distribución inicial del estado, $p(x_0)$. En el primer caso se ha utilizado $N = 30$ partículas, obtenidas experimentalmente. En el segundo caso, se utiliza la distribución normal de parámetros μ_0, Σ_0 , es decir, $p(x_0) \equiv N(\mu_0, \Sigma_0)$, donde el estado inicial puede ser descrito por (17) cuando se conoce el estado en dos imágenes consecutivas t y $t + 1$, y la matriz de varianza se establece por (18).

$$(x_0, y_0, u_0, v_0, w_0, h_0) = (x_{t+1}, y_{t+1}, x_{t+1} - x_t, y_{t+1} - y_t, w_{t+1}, h_{t+1}) \quad (17)$$

$$\Sigma_0 = \text{diag}\{100, 100, 25, 25, 9, 9\} \quad (18)$$

Para comprobar la validez del método propuesto se han desarrollado varios experimentos, estos incluyen varias secuencias de imágenes tomadas sobre el vehículo IvvI, sus resultados se presentan en la tabla 1. Los errores presentados en ésta y la siguiente tabla se deben a falsas posiciones entregadas por el algoritmo de seguimiento. La figura 12 (a) ilustra este tipo de error.

2. Seguimiento de los ojos

El seguimiento se desarrolla sobre los rectángulos ROI_L y ROI_R detectados en conjunto con el rostro. El estado de cada ojo viene dado por la ecuación (19).

$$x_t = (x, y, u, v) \quad (19)$$

donde (x, y) es la posición del centro del ojo y (u, v) es su velocidad en las direcciones x e y , respectivamente.

El modelo de transición se describe por la ecuación (20), que representa un proceso autoregresivo de primer orden $AR(1)$.

$$x_t = A x_{t-1} + \xi_t \quad (20)$$

nuevamente, ξ_t representa la perturbación del sistema que viene dado por un ruido blanco y A es la matriz que genera la relación entre los dos instantes de tiempo, su estructura esta dada en (Isard, 1998; Koller-Meier and Ade, ???).

El modelo de medida es:

$$z_t = H x_{t-1} + \eta_t \quad (21)$$

con η_t un ruido blanco, H viene dada en (Doucet et al., 2001) y $z_t = (x, y)$ es la posición del centro el ojo.

Para la evaluación de la función de densidad a posteriori se ha utilizado la correlación normalizada junto con el modelo generado con el método propuesto para detectar los ojos. El modelo que ingresa en la correlación es sometido a un procesamiento previo a partir del cual

se obtienen dos modelos, uno alargado verticalmente y otro alargado horizontalmente. Los modelos alargados permiten captar mejor las variaciones que sufren los ojos debido a los múltiples cambios que pueden presentar, como pueden ser: las variaciones por parpadear, los movimientos del rostro, etc.

Los pesos $\pi_i^{(n)}$ son aproximados por el valor de la correlación dada por (22).

$$p(z_i|x_i = s_i^{(n)}) = \max(g_1(s_i^{(n)}), g_2(s_i^{(n)}), g_3(s_i^{(n)})) \quad (22)$$

en este caso, $g_i(\cdot)$ para $i = 1, 2, 3$, representa el valor de la correlación del modelo i en la posición predicha por el algoritmo Condensation.

La correlación consiste en calcular la similitud de una porción de imagen con respecto al modelo propuesto (de la Escalera, 2001; Pitas, 2000).

$$g(x, y) = \frac{\sum ((I(x, y) - I_m)(M(x, y) - M_m))}{\sqrt{\sum (I(x, y) - I_m)^2 \sum (M(x, y) - M_m)^2}} \quad (23)$$

los valores I_m y M_m representan el valor promedio de las imágenes I y M , respectivamente.

El modelo M (imagen utilizada de patrón) tiene dimensiones a y b para el ancho y el alto, respectivamente, por lo tanto, la suma de la ecuación (23) se realiza al amparo de estos dos parámetros sobre cada uno de los píxeles que corresponden a este dominio.

La inicialización del algoritmo Condensation se realiza con los siguientes parámetros: $(x_0, y_0, u_0, v_0) = (x_t, y_t, x_{t+1} - x_t, y_{t+1} - y_t)$, $\sum_0 = \text{diag}\{100, 100, 25, 25\}$ y el número de partículas es 50. La tabla 1 presenta los resultados obtenidos en el seguimiento de los ojos sobre cinco secuencias imágenes de conductores sobre IvvI. Los errores presentados en ésta y la siguiente tabla se deben a falsas posiciones entregadas por el algoritmo de seguimiento. Un error de este estilo se presenta en la figura 12 (b).

4.4. Detección de la somnolencia

PERCLOS es el índice más utilizado en la construcción de este tipo de sistemas, por tal motivo ha sido implementado en este trabajo. Existen otros índices como el AECS (Ji and Yang, 2002) y la duración de cierre del ojo durante cinco fotogramas consecutivos (Dong and Wu, 2005), que también son utilizados para éste propósito.

Para calcular los índices es imprescindible conocer el estado del ojo en cada instante. En este trabajo se manejan dos estados: abierto y cerrado. Para conocer estos estados se ha utilizado una técnica robusta de clasificación denominada Support Vector Machine (SVM) (Cristianini and Shawe-Taylor, 2006) en conjunto con el filtro de Gabor (Daugman, 1985), técnica ampliamente utilizada para contribuir en el realce de ciertas características (Liu, 2004; Zhou and Wei, 2006).

Matemáticamente SVM se plantea a través del siguiente problema de optimización cuadrática.

$$\begin{aligned} \min_{\alpha} J(\alpha) &= \alpha Q \alpha^T - e^T \alpha \\ \text{s.t.} \\ 0 &\leq \alpha_i \leq C \text{ para } i = 1, \dots, m \\ y^T \alpha &= 0 \end{aligned} \quad (24)$$

donde $e \in R^m$ es el vector de unos, C es una cota superior que penaliza el error, Q es una matriz de tamaño $m \times m$ con $Q_{ij} = y_i y_j K(x_i, x_j)$, $K(x_i, x_j)$ es la denominada función núcleo, que puede ser una función lineal o no-lineal, x_i es el vector de características y y es el vector que representa las clases, en este caso, $y = 1$ para el ojo abierto y $y = 2$ para el ojo cerrado.

Por su parte, el filtro de Gabor se forma a partir de los wavelets Gabor, también conocidos como núcleos o filtros (Liu, 2004; Zhou and Wei, 2006). La función que genera el wavelet Gabor viene definida por:

$$\Psi_{\mu, \nu}(z) = \frac{\|k_{\mu, \nu}\|^2}{\sigma^2} \exp\left(-\frac{\|k_{\mu, \nu}\|^2 \|z\|^2}{2\sigma^2}\right) \times \left[\exp(ik_{\mu, \nu} z) - \exp\left(-\frac{\sigma^2}{2}\right)\right] \quad (25)$$

donde $z = (x, y)$ es el punto de coordenadas x e y . Los parámetros μ y ν definen la orientación y la escala, $\|\cdot\|$ define la norma, y σ es la desviación estándar de la envolvente gaussiana. El vector complejo que representa la onda $k_{\mu, \nu}$ se define por (26):

$$k_{\mu, \nu} = k_{\nu} \exp(i\phi_{\mu}) \quad (26)$$

donde $i = \sqrt{-1}$ es el número complejo, $k_{\nu} = \frac{k_{\max}}{f^{\nu}}$, $\phi_{\mu} = \frac{\pi\mu}{8}$, k_{\max} es la frecuencia máxima y f es el espacio entre los núcleos en el dominio de la frecuencia.

La representación de una imagen con el filtro de Gabor es la convolución de dicha imagen con la familia de núcleos Gabor. Para este trabajo se han utilizado tres escalas ($\pi/8, \pi/4, \pi/2$) y cuatro orientaciones ($0, \pi/6, 2\pi/6, 3\pi/6$).

El vector de características, que consiste en dividir cada imagen, de tamaño 30 por 20 píxeles, en ventanas de tamaño 5 por 4 píxeles, y calcular el promedio de cada ventana. Con esto se tiene un vector de dimensión 360.

Para entrenar SVM se han utilizado los siguientes parámetros: $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \equiv RBF$, $C = 35$ y $\gamma = 0,0128$. Los resultados presentan una tasa de clasificación del 94 % sobre una base de datos de prueba de 500 ojos cerrados y 1000 ojos abiertos.

La figura 6 presenta los resultados obtenidos, donde se aprecian varios picos que se forman sobre los índices calculados a partir de un conductor que presenta alto grado de somnolencia. En ambos casos, el sistema es capaz de detectar la presencia de adormecimiento. En la tabla 1 se presentan más experimentos que validan el sistema propuesto. Los errores presentados en ésta y la siguiente tabla se deben a falsas clasificaciones entregadas por el método de clasificación entre ojos abiertos y cerrados. En la figura 12 (c) se presenta un ejemplo de este error, cuando el sistema clasifica al ojo como abierto cuando en realidad se encuentra cerrado.

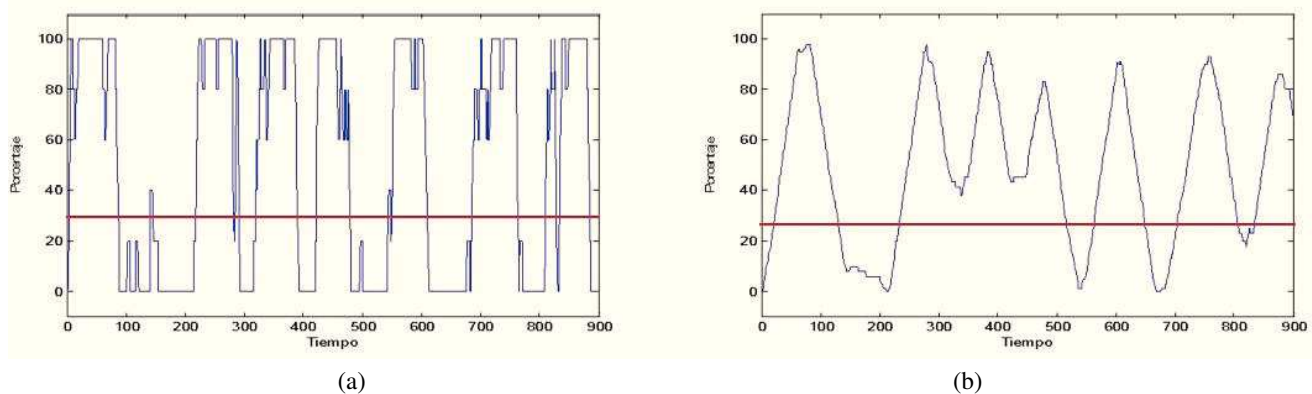


Figura 6: Evolución del índice de: (a) Horng *et al.* y Dong-Wu, (b) PERCLOS, en una secuencia de 900 imágenes de un conductor somnoliento. La línea roja representa el umbral de somnolencia

Tabla 1: Resultados del sistema en conducción diurna

Conductor	Número de imágenes	Seguimiento del rostro		Seguimiento de los ojos		Estado de los ojos		
		Número de fallos	Aciertos (%)	Número de fallos	Aciertos (%)	Ojos abiertos	Ojos cerrados	Aciertos (%)
D1	960	60	93.75	20	97.91	690/700	258/260	98.90
D2	900	22	97.55	30	96.60	520/560	339/340	96.27
D3	500	45	91.00	8	98.40	388/400	99/100	98.00
D4	330	15	95.45	14	95.75	150/170	152/160	91.61
D5	1400	50	96.42	90	93.57	891/980	401/420	93.19

4.5. Análisis de la distracción

En el caso de la distracción, se ha estimado que representa el 20 % de los accidentes de tráfico. Para detectarla es necesario analizar la cabeza/rostro del conductor, porque la posición y orientación del rostro/cabeza contienen información acerca del nivel de atención y fatiga (Ji and Yang, 2001; Lisheng et al., 2009). En este trabajo, para estimar la distracción del conductor se ha implementado el algoritmo, compuesto de dos partes:

1. Estimación de la orientación:

Para estimar la orientación del rostro a partir de la posición de los ojos se utiliza (27), que mide el ángulo que forma el rostro respecto al eje vertical, ver figura 7 (a).

$$\theta = \tan^{-1} \left(\frac{y_2 - y_1}{x_2 - x_1} \right) \quad (27)$$

Experimentalmente se ha comprobado que los límites fijados por (28) con $\theta_0 = 8^\circ$ dan una buena estimación para conocer la orientación del rostro, ver figura 7 (b).

$$\begin{cases} \text{Izquierda} & \text{si } \theta > \theta_0 \\ \text{Frente} & \text{si } |\theta| \leq \theta_0 \\ \text{Derecha} & \text{si } \theta < -\theta_0 \end{cases} \quad (28)$$

2. Estimación del giro del rostro:

El índice de distracción se calcula una vez que el sistema detecta que el rostro no está de frente, es decir, que se

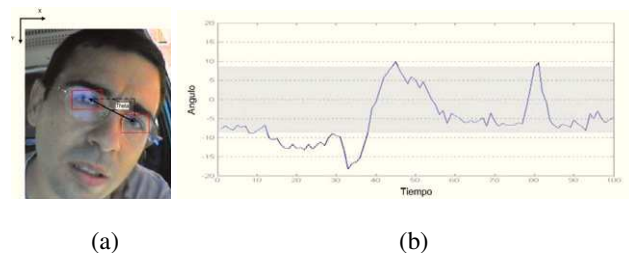


Figura 7: Orientación del rostro, (a) modelo, (b) evolución en una secuencia de 100 imágenes.

encuentra girado hacia la izquierda o derecha durante 5 fotogramas consecutivos. Para comprobar la orientación del rostro se ha utilizado una RNA entrenada con rostros en cinco estados: derecha, media-derecha, frente, media-izquierda e izquierda.

5. Sistema de detección de la somnolencia y la distracción en conducción nocturna

En este caso se utilizará la información proveniente del sistema de percepción excitado con iluminación infrarroja.

5.1. Sistema de percepción

Este sistema consta de tres partes: el sistema de visión, el sistema de sincronización y el sistema de iluminación.

La primera parte está compuesta por una cámara CCD (similar al caso anterior), en la que se ha removido el filtro IR. Esta cámara genera una señal de video compuesto (PAL, phase alternating line). La segunda parte toma la señal anterior y la divide en los campos par e impar, utilizando el sincronizador-separador de video LM1881N. Finalmente, el sistema de iluminación, basado en la iluminación infrarroja del tipo cercano (700-900 nm.), se encuentra compuesto por dos conjuntos de leds (uno interior y otro exterior) que producen las imágenes del campo par (I_P) e impar (I_I) sobre las cuales el efecto de la pupila es resaltado (Dong and Wu, 2005; Ji and Yang, 2002; Grace et al., 1998). El anillo interior se encuentra alrededor de la cámara, y los anillos exteriores están a una distancia simétrica respecto al centro de la cámara, uno a cada lado. La figura 8 (a), muestra este sistema, y la figura 8 (b) presenta un ejemplo de un conductor.

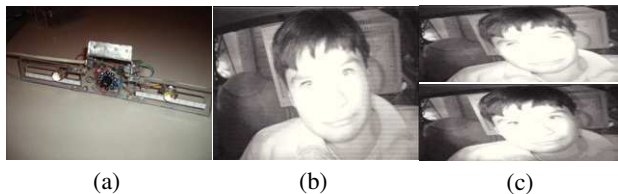


Figura 8: Sistema de percepción, (a) sistema de visión, (b) imagen del conductor con iluminación infrarroja, (c) imágenes del campo par e impar.

Luego, cada fotograma es dividido en los campos par e impar, que contienen la parte brillante y oscura de la imagen de la pupila, respectivamente. Esto se puede observar en la figura 8 (c), en la parte superior esta el campo par y en la inferior el campo impar.

5.2. Detección de los ojos

El brillo de las pupilas es el efecto utilizado para detectar los ojos. En este sentido, se han construido tres imágenes a partir de la imagen del conductor: la imagen diferencia (I_D), la imagen de aristas (I_G) y la imagen FRST (fast radial symmetry transform, (Loy and Zelinsky, 2003; Loy, 2003; Loy and Barnes, 2004)) (I_F).

La primera resulta de la diferencia en valor absoluto de las imágenes de los campos par e impar, figura 9 (a). Sobre esta la mayoría del fondo y la iluminación externa han sido removidos, y las pupilas resultan ser la parte más brillante. La segunda y la tercera imágenes se consiguen utilizando el operador de Sobel, figura 9 (b), dado por (29), y la parte positiva de FRST (Loy and Zelinsky, 2003) sobre la imagen diferencia, figura 9 (c).

$$I_G = \sqrt{\|I_D * S_X\|^2 + \|I_D * S_Y\|^2} \quad (29)$$

donde S_X y S_Y vienen dadas por (8).

En situaciones de conducción real, la imagen diferencia es deteriorada por la iluminación externa, vibraciones producto del movimiento del vehículo, etc. En estas circunstancias, es necesario incorporar mayores fuentes de información. Por lo tanto, se ha incluido las imágenes del gradiente y FRST, de la

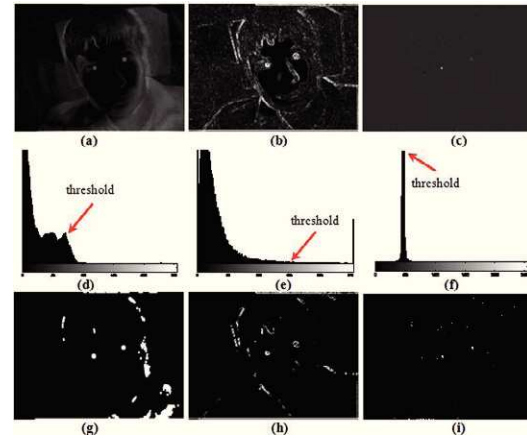


Figura 9: Imágenes I_D , I_G e I_F y sus respectivos histogramas e imágenes binarias.

imagen diferencia, ya que han demostrado experimentalmente ser más estables ante los factores citados.

Una vez identificadas las fuentes de información, el siguiente paso consiste en binarizar estas imágenes y obtener el conjunto de candidatos a pupilas, para ello se sigue el siguiente algoritmo:

- En el primer caso, el umbral se estima al ir eliminando sistemáticamente los píxeles de menor valor hasta alcanzar un número determinado de cuerpos, ver figuras 9 (d) y (g).
- En el segundo caso, el histograma se modeliza con la función de distribución de probabilidad Gamma (Durrett, 1991); al 90 % de confianza se obtiene el umbral, ver figuras 9 (e) y (h).
- En el último caso, el nivel de gris que corresponde al valor máximo de su histograma genera el umbral, ver figuras 9 (f) e (i).

Las pupilas son detectadas buscando sobre los cuerpos generados anteriormente y siguiendo una doble estrategia: eliminar los cuerpos que presentan mayor alto que ancho. Luego, utilizando el método de clasificación SVM (entrenado con el filtro de Gabor bajo una orientación 0 y cuatro escalas $0, \pi/4, \pi/2, 3\pi/4$). Los parámetros utilizados para su entrenamiento son: el núcleo RBF, $C = 100$ y $\gamma = 0,128$. El vector de características se compone del nivel de grises de estas imágenes. Con esto se ha obtenido una tasa de clasificación del 93 % sobre una base de datos de prueba que consta de 366 ojos abiertos y 595 no ojos.

5.3. Detección del rostro

Una vez que los ojos han sido detectados, el sistema localiza el rostro del conductor. Por lo cual, se construye un modelo del rostro a partir de la posición de los ojos y las medidas antropométricas del rostro (Gejgus and Sperka, 2003). Sean

$p_1(x, y)$ y $p_2(x, y)$ las posiciones del centro del ojo derecho e izquierdo, respectivamente, y d_{RL} es la distancia que los separa (en píxeles), viene dada por (30). La figura 10 muestra este modelo y su ajuste.

$$d_{RL} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (30)$$

$$\theta = \tan^{-1}\left(\frac{y_2 - y_1}{x_2 - x_1}\right) \quad (31)$$

$$R_1 = 1,5d_{RL} \quad (32)$$

$$R_2 = d_{RL} \quad (33)$$

$$p_c(x, y) = (x_1 + 0,5d_{RL}, y_1 + 0,3d_{RL}) \quad (34)$$

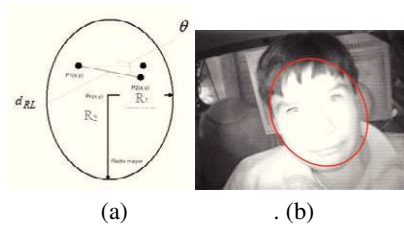


Figura 10: Modelo paramétrico del rostro a partir de la posición de los ojos.

5.4. Seguimiento del rostro y los ojos

El seguimiento es similar al caso del sistema en conducción diurna, por lo tanto, solo se procederá a describir la evaluación de la función de densidad de las observaciones, en los casos del rostro y los ojos.

1. *Seguimiento del rostro*: Para evaluar la función de densidad de las observaciones nuevamente se ha utilizado una RNA (Looney, 1997), entrenada con el filtro de Gabor (Daugman, 1985), con una escala y dos orientaciones. La tasa de clasificación sobre un conjunto de prueba que consta de 1000 rostros y 1500 no rostros, alcanza el 92 %.

El modelo dinámico y los parámetros de inicialización son similares al caso de la conducción diurna. Un ejemplo de los errores cometidos se presenta en la figura 12 (d), donde el algoritmo señala principalmente la mano. El resumen de más experimentos se encuentran en la tabla 2.

2. *Seguimiento de los ojos*: En este caso el estado del ojo es caracterizado por su posición y velocidad. Para evaluar la función de densidad de las observaciones se ha utilizado la función de densidad triangular, sobre la imagen diferencia, su rango es el intervalo [0, 255] que corresponde a la escala de grises.

El algoritmo es inicializado una vez que los ojos son detectados en la sección previa. El modelo dinámico y los

parámetros de inicialización son similares al caso de la conducción diurna. Un ejemplo de los errores cometidos por el algoritmo se presenta en la figura 12 (e), donde se señala la ceja en lugar del ojo. El resumen de más experimentos se encuentran en la tabla 2.

La tabla 2 muestra los resultados obtenidos en el seguimiento del rostro y los ojos sobre cinco conductores en el vehículo IvvI. Al igual que en caso anterior, para contrastar la información generada por el sistema ha sido necesario extraer la información manualmente de un conjunto de imágenes.

5.5. Análisis de los ojos e índice de somnolencia

Para conocer el estado del ojo, el sistema utiliza la información del área de la pupila detectada con el método de la sección anterior. Al cerrar el ojo, el efecto del brillo de la pupila desaparece, lo que implica que el ojo se encuentra cerrado.

En la tabla 2 se aprecian varios resultados generados en experimentos desarrollados sobre el vehículo IvvI para analizar la capacidad del sistema al clasificar entre ojos abiertos y cerrados, por otra parte, en la figura 11 (a) se observa la evolución del índice PERCLOS sobre una secuencia de 350 imágenes tomadas en IvvI en condiciones de conducción nocturna, donde la línea entrecortada marca el umbral de somnolencia; en la figura 11 (b) se observa su respuesta instantánea a un episodio de somnolencia. A la vez, en la figura 12 (f) se presenta el error típico de clasificación, que genera el algoritmo, al confundir cerrado con abierto.

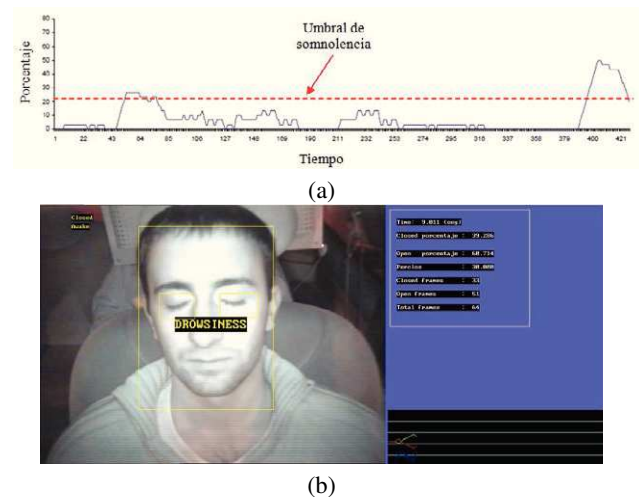


Figura 11: (a) Evolución de PERCLOS en una secuencia de 425 imágenes, (b) respuesta instantánea del sistema.

5.6. Análisis de la distracción

Este caso es similar al sistema para conducción diurna, donde también se ha utilizado la categorización frente, izquierda y derecha para conocer el estado de atención del conductor.

Tabla 2: Resultados del sistema en conducción nocturna

Conductor	Número de imágenes	Seguimiento del rostro		Seguimiento de los ojos		Estado de los ojos		
		Número de fallos	Aciertos (%)	Número de fallos	Aciertos (%)	Ojos abiertos	Ojos cerrados	Aciertos (%)
D1	800	40	95.00	14	98.25	690/700	97/100	97.78
D2	646	29	95.51	29	95.51	500/530	100/116	90.27
D3	600	20	96.67	24	96.00	345/374	206/226	91.69
D4	2500	65	97.40	95	96.20	2050/2200	297/300	95.75
D5	3500	103	97.05	150	95.71	3186/3250	230/250	95.01

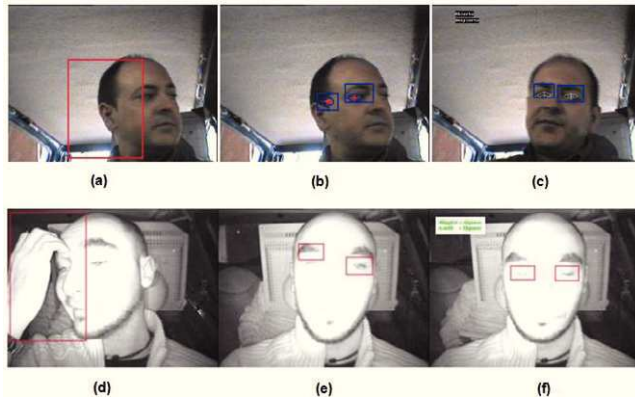


Figura 12: Típicos errores del sistema, (a) falso seguimiento (localización) del rostro, (b) falso seguimiento (localización) del ojo derecho, (c) falsa clasificación del estado del ojo.

6. Conclusiones

En este artículo se han presentado dos sistemas para detectar automáticamente la somnolencia del conductor y paralelamente su distracción, utilizando para este propósito técnicas de Visión por Computador e Inteligencia Artificial. En términos generales, estos sistemas monitorizan los ojos del conductor en tiempo real, para analizarlos y calcular el índice de somnolencia PERCLOS. Paralelamente se analiza la orientación del rostro para construir un índice de distracción basado el umbral de cinco fotogramas consecutivos. Además, en el diseño de los sistemas se han utilizado características intrínsecas que no dependen de factores subjetivos como el color de los ojos o el color de la piel, esto ha dado robustez en los dos escenarios, como se confirma en los resultados presentados en las distintas tablas a lo largo del documento. Dichos experimentos fueron desarrollados sobre la plataforma de investigación IvvI.

En el sistema para conducción diurna se han realizado las siguientes aportaciones:

- Se ha desarrollado un método basado en la forma geométrica para localizar la posición de los ojos. Para ello se ha extraído la información necesaria para modelar dicha forma y ajustar un modelo elíptico basado en el algoritmo EM. Este método trabaja adecuadamente en ambientes con iluminación variable debido a que utiliza

información invariante a los cambios de iluminación y toma en cuenta la geometría del ojo.

- El uso de SVM en conjunto con el filtro de Gabor ha servido para construir un método eficiente para detectar el estado del ojo. El filtro de Gabor, al variar los parámetros de orientación y escala, resalta ciertos rasgos típicos de un ojo abierto y opaca los de un ojo cerrado, y viceversa.

En el sistema para conducción nocturna se han realizado las siguientes aportaciones:

- Para este caso se ha construido un complejo sistema de visión basado en la luz infrarroja del tipo cercano. Este sistema es capaz de separar la señal de video en dos campos, denominados par e impar, que son la base para localizar la posición de los ojos. Lo novedoso de su diseño, radica en que es posible controlar la cantidad de luz radiada mediante dos maneras: (i) variando la corriente de los leds exteriores e interiores, y (ii) modificando la distancia de los leds exteriores respecto al centro de la cámara. Estas incorporaciones han sido beneficiosas para ajustar la calidad de las imágenes. Además, tiene un tamaño reducido, ideal para distintas aplicaciones del tipo análisis del conductor.
- En cuanto al desarrollo del algoritmo, se ha presentado una metodología para detectar los ojos basada en la imagen diferencia, su imagen de bordes y FRTS y los algoritmos de umbralización desarrollados para segmentar entre el fondo (colores oscuros) y los potenciales ojos (colores blancos). Una vez obtenido el conjunto de cuerpos se los filtra bajo dos estrategias, la primera consiste en eliminar los cuerpos que no cumplan ciertas restricciones geométricas, y la segunda consiste en discriminar utilizando el clasificador SVM. Esta doble estrategia ha permitido localizar la posición de las pupilas y por ende la de los ojos de manera robusta.

Los sistemas procesan entre 4 y 5 fotogramas por segundo. Además, se puede resaltar que en este trabajo se ha presentado una nueva y completa propuesta que puede contribuir significativamente al desarrollo de la tecnología aplicada a prevenir los accidentes de tráfico, en el contexto, de los sistemas de seguridad activa.

Finalmente, en los trabajos a futuro se desea incorporar un modelo en 3D del rostro (cabeza), usando visión estéreo para estimar la posición del rostro, la cabeza y la dirección de la mirada.

English Summary

Advance assistance system for driver's drowsiness detection.

Abstract

Every day, statistics on traffic accidents reveal that human errors are the main cause of deaths and injuries on the world's roads. In order to reduce these fatalities, a system for automatic detection of both drowsiness and distraction is presented. Artificial intelligent, computer vision and infrared illumination technologies are used to compute both drowsiness and distraction indexes in real time. Several examples of different driver images taken in a real vehicle at nighttime are shown to validate the proposed algorithms.

Keywords:

Artificial intelligent, computer vision, drowsiness, driver, traffic accidents, infrared illumination.

Agradecimientos

Este trabajo ha sido realizado gracias al apoyo del gobierno español a través de dos proyectos de la CICYT, VISVIA (TRA2007-67786-C02-02) y POCIMA (TRA2007-67374-C02-01).

Referencias

- ASFA, 2008. Driver fatigue is the number one cause of catastrophic truck accidents. Website, <http://www.autoroutes.fr/>.
- Bergasa, L., Nuevo, J., Sotelo, M., Barea, R., Lopez, E., March 2006. Real-time system for monitoring driver vigilance. *IEEE, Transactions on Intelligent Transportation Systems* 7 (1), 63–77.
- Bergasa, L., Nuevo, J., Sotelo, M., Vázquez, M., Junio 14-17 2004. Real-time system for monitoring driver vigilance. *IEEE, Intelligent Vehicles Symposium* 1 (2).
- Bloemkolk, F., de Lijster, J., van Gelderen, M., July 2007. ITS strategy: the Japanese formula for success. Study to promote ITS implementation in the Netherlands. Technical report, International Affairs Office, Ministry of Transportation, Public Works and Water Management.
- Branzan, A., Widsten, B., Wang, T., Lan, J., Mah, J., June 2008. A computer vision-based system for real-time detection of sleep onset in fatigued drivers. *IEEE, Intelligent Vehicles Symposium*, 25–30.
- Brookshear, J., 1983. Theory of computation : Formal Languages; Automata and Complexity. Vol. 1. Addison Wesley Iberoamericana.
- Chang, B., Lim, J., Kim, H., Seo, B., September 2007. A study of classification of the level of sleepiness for the drowsy driving prevention. *IEEE, SICE Annual Conference*, 3084–3089.
- Cristianini, N., Shawe-Taylor, J., 2006. An introduction to Support Vector Machines and other kernel-based learning methods. Cambridge University Press.
- Daugman, J., 1985. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional cortical filters. *J. Optical Soc. Am.* 2 (7), 1160–1169.
- de la Escalera, A., 2001. Visión por Computador, Fundamentos y Métodos. Vol. 1. Prentice Hall, Pearson Educación, Madrid.
- Dong, W., Wu, X., 2005. Driver fatigue detection based on distant eyelid. *IEEE, Int. Workshop VLSI Design & Video Tech.*
- Doucet, A., N. Freitas de, Gordon, N., 2001. Sequential Monte Carlo Methods in Practice. Vol. 1. Springer-Verlag.
- D'Orazio, T., Leo, M., Distant, A., June 2004. Eye detection in face images for a driver vigilance system. *IEEE, Intelligent Vehicle Symposium*, 95–98.
- Durrett, R., 1991. Probability: Theory and Examples. Vol. 1. Library of Congress Cataloging-in-Publication Data.
- Evgeniou, T., Pontil, M., Papageorgiou, C., Poggio, T., 2000. Image representations for object detection using kernel classifiers. In *Asian Conference on Computer Vision*.
- Fletcher, L., Petersson, L., Zelinsky, A., 2003. Driver assistance systems based on vision in and out of vehicles. *IEEE, Proceedings of Intelligent Vehicle Symposium*, 322–327.
- Freund, Y., Schapire, R., 1995. A decision-theoretical generalization of online learning and an application to boosting. In *Second European Conference on Computational Learning Theory*.
- Gejgus, P., Sperka, M., 2003. Face tracking in color video sequences. *Association for Computing Machinery*, 245–249.
- Grace, R., Byrne, V., Bierman, D., Legrand, J., Grcourt, D., Davis, R., Staszewski, J., Carnahan, B., October 1998. A drowsy driver detection system for heavy vehicles. *IEEE, Proceedings of Digital Avionics System Conference* 2, 1–8.
- Guo, J., Guo, X., July 2009. Eye state recognition based on shape analysis and fuzzy logic. *IEEE Intelligent Vehicle Symposium*, 78–82.
- Hagenmeyer, L., August 2007. Development of a multimodal, universal human-machine-interface for hypovigilance-management-systems. Ph.D. thesis, Mechanical Engineering, University of Stuttgart, Institute for Human Factors and Technology Management.
- Hanmi, I., 2005. Drowsy truck drivers. Website, <http://www.gohanmi.com/NREC-COPILOT.htm>.
- Hansen, D., Ji, Q., March 2010. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (3), 478–500.
- Hayami, T., Matsunaga, K., Shidoji, K., Matsuki, Y., September 2002. Detecting drowsiness while driving by measuring eye movement - a pilot study. *IEEE International Conference on Intelligent Transportation Systems*, 156–161.
- Hilario, C., Octubre 2008. Detección de peatones en el espectro visible e infrarrojo para un sistema avanzado de asistencia a la conducción. Ph.D. thesis, Departamento de Ingeniería de Sistemas y Automática, Universidad Carlos III de Madrid.
- Hong, W., Chen, C., Chang, Y., 2004. Driver fatigue detection based on eye tracking and dynamic template matching. *IEEE Proceedings of, International Conference on Networking, Sensing and Control*.
- Isard, M., Blake, A., 1998. Condensation - conditional density propagation for visual tracking. *International Journal of Computer Vision* 29 (1), 5–28.
- Isard, M. A., September 1998. Visual motion analysis by probabilistic propagation of conditional density. Ph.D. thesis, Department of Engineering Science, University of Oxford.
- Ji, Q., Yang, X., 2001. Real-time visual cues extraction for monitoring driver vigilance. *Lectures Notes in Computer Science, Proceedings of the Second International Workshop on Computer Vision Systems* 2095, 107–124.
- Ji, Q., Yang, X., 2002. Real-time eye, gaze and face pose tracking for monitoring driver vigilance. *Elsevier Science Ltd., Real Time Imaging* 1 (8), 357–377.
- Ji, Q., Zhu, Z., Lan, P., Junio 2004. Real time nonintrusive monitoring and prediction of driver fatigue. *IEEE, Transaction on Vehicular Technology* 53 (4).
- Jiangwei, C., Lisheng, J., Lie, G., Keyou, G., Rongben, W., June 2004a. Driver's eye state detecting method design based on eye geometry feature. *IEEE, Intelligent Vehicles Symposium*, 357–362.
- Jiangwei, C., Lisheng, J., Lie, G., Keyou, G., Rongben, W., June 2004b. A monitoring method of driver mouth behaviour based on machine vision. *IEEE, Intelligent Vehicles Symposium*, 351–356.
- Knipping, R., Wierwille, W., 1994. Vehicle-based drowsy driver detection: Current status and future prospects. *IVSH America Fourth Annual Meeting*.
- Koller-Meier, E., Ade, F., 2000. Tracking multiple objects using the condensation algorithm.
- Küçükay, F., Bergholz, J., 2005. Driver assistant systems. *Lectures of Institute of Automatic Engineering*.

- Kuttila, M., December 2006. Methods for machine vision based driver monitoring applications. Ph.D. thesis, Tietotalo Building, Auditorium TB104.
- Lisheng, J., Xuan, S., Yuying, J., Haijing, H., Yuqin, S., June 2009. Study on driver's mouth segmentation and location based on color space. IEEE Intelligent Vehicles Symposium, 500–506.
- Liu, C., May 2004. Gabor-based kernel pca with fractional power polynomial models for face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (5), 572–582.
- Longhurst, G., ??? Understanding driver visual behaviour. Seeing Machine Pty Limited.
- Looney, C., 1997. Pattern Recognition Using Neural Networks: theory and algorithms for engineers and scientists. Oxford University Press Inc.
- Loy, G., January 2003. Computer vision to see people: a basis for enhanced human computer interaction. Ph.D. thesis, Robotics Systems Laboratory, Department of Systems Engineering, Research School of Information Sciences and Engineering, Australian National University.
- Loy, G., Barnes, N., September 2004. Fast shape-based road sign detection for a driver assistance system. IEEE, International Conference on Intelligent Robots and Systems (IROS'04) 1, 70–75.
- Loy, G., Zelinsky, A., August 2003. Fast radial symmetry for detecting points of interest. IEEE, Transactions on Pattern Analysis and Machine Intelligence 25 (8), 959–973.
- Martinez, W., Martinez, A., 2002. Computational Statistics Handbook with Matlab. Chapman & Hall/CRC.
- NHTSA, April 1998. Evaluation of techniques for ocular measurement as an index of fatigue and the basis for alertness management. Final Report DOT HS 808 762, National Highway Traffic Safety Administration, Virginia 22161, USA.
- Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Systems, Man and Cybernetics, 62–66.
- Pitas, I., 2000. Digital Image Processing Algorithms and Applications. A Wiley-Interscience Publication. John Wiley & Sons, Inc.
- Ristic, B., Arulampalam, S., Gordon, N., 2004. Beyond the Kalman Filter: Particle Filters for Tracking Applications. Vol. 1. Artech House.
- Rogers, C., 1998. Hausdorff measures. Vol. 1. Cambridge : Cambridge University Press.
- Rongben, W., Keyou, G., Shuming, S., Jiangwei, C., June 2003. A monitoring method of driver fatigue behavior based on machine vision. IEEE, Proceedings on Intelligent Vehicles Symposium, 110–113.
- Tian, Z., Qin, H., October 2005. Real-time driver's eye state detection. IEEE, International Conference on Vehicular Electronics and Safety, 285–289.
- Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. Computer Vision and Pattern Recognition, Proceedings of the 2001 IEEE Computer Society Conference on 1, 1–511–1–518.
- Viola, P., Jones, M., 2002a. Fast and robust classification using asymmetric adaboost and a detector cascade. Advances in Neural Information Processing System, MIT Press, Cambridge, M. A. (14).
- Viola, P., Jones, M., 2002b. Robust real-time object detection. International Journal of Computer Vision - to appear.
- Vlacic, L., Parent, M., Harashima, F., 2001. Intelligent Vehicle Technologies. A division of Reed Educational and Professional Publishing Ltda. Library of Congress Cataloguing in Publication Data.
- Wang, Q., Yang, J., Ren, M., Zheng, Y., June 2006. Driver fatigue detection: A survey. IEEE, Proceedings of the 6th World Congress on Intelligent Control and Automation, 8587 – 8591.
- Wu, Y., Liu, H., Zha, H., June 2004. A new method of detection humand eyelids based on deformable templates. IEEE International Conference on Systems, Man and Cybernetics, 604–609.
- Zhou, M., Wei, H., 2006. Face verification using gabor wavelets and adaboost. IEEE, 18th. International Conference on Pattern Recognition ICPR06 1, 404–407.
- Zhu, Z., Fujimura, K., Ji, Q., 2002a. Real-time eye detection and tracking under various light conditions. Proceedings of the 2002 Symposium of Eye tracking research & applications, 139–144.
- Zhu, Z., Ji, Q., Fujimura, K., Lee, K., 2002b. Combining Kalman filtering and mean shift for real time eye tracking under active ir illumination. Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02) 4, 318–321.