

# Monitorización de ocupantes en el interior de vehículos usando visión computacional

J. Javier Yebe, Luis M. Bergasa, Pablo F. Alcantarilla

**Abstract**—En este artículo se presentan los resultados preliminares de un sistema de monitorización de ocupantes en el interior de vehículos usando visión computacional. El objetivo final consiste en contar el número de ocupantes, clasificarlos entre adultos, niños y bebés, y grabar el último minuto de video en caso de accidente. Esta información servirá de base al moderno sistema paneuropeo de llamada de emergencia (eCall) que conectará automáticamente con una central del servicio de emergencias 112 en caso de accidente. Se estudia la viabilidad de colocar una o dos cámaras VGA con lentes de “ojo de pez” y su mejor ubicación en el interior del vehículo para capturar las imágenes que han de ser analizadas. Debido a la distorsión radial que introducen dichas ópticas, se aborda también la calibración y rectificación de las imágenes como paso previo a la aplicación de las técnicas de extracción de la información útil. Se estudia la utilización del algoritmo Viola-Jones para detección de caras como núcleo del sistema de conteo de ocupantes y se presentan algunos resultados que validan la propuesta.

**Index Terms**— Computer vision, eCall, Viola-Jones, Haar-classifier, face detection, occupant monitoring.

## I. INTRODUCCIÓN

Las técnicas de visión por computador y los importantes avances en el procesamiento de imágenes están mostrando al mundo un gran abanico de posibilidades en multitud de campos de aplicación y no solamente en el ámbito científico e investigador, sino también con desarrollos y productos que gozan del interés empresarial y social. En este documento se aborda una propuesta concebida dentro del marco del sistema paneuropeo de llamada de emergencia eCall: la monitorización de ocupantes en el interior del vehículo. El sistema eCall parte de la preocupación por la seguridad vial y los tiempos de rescate en caso de accidente, ya que en base a diversas fuentes, entre el 10 y el 25% de las víctimas podrían salvarse o disminuir sus secuelas si recibieran la atención

adecuada en el momento del accidente.

La Unión Europea ya se ha marcado el objetivo de implantar el sistema eCall en todos sus países miembros para 2009 y será a partir de 2010 cuando todos los vehículos nuevos comiencen a estar equipados con dicho sistema [1].

Aprovechando los retos y oportunidades que ofrece el sistema eCall en materia de seguridad vial e innovación para vehículos, este artículo introduce un estudio a cerca de la colocación de cámaras VGA en el interior del vehículo para monitorizar sus ocupantes y grabar el último minuto de video en caso de accidente, aportando una información útil que pueda ser enviada a la central de emergencias.

El sistema propuesto analizará las imágenes capturadas mediante un algoritmo basado en el detector de caras Viola-Jones [2] para determinar el número de ocupantes en el interior del vehículo. Se estudiará adicionalmente, la clasificación según adultos, niños o bebés, así como las distintas condiciones de iluminación que podrían darse en el entorno de grabación. Este artículo aborda la primera parte del trabajo presentando los resultados preliminares del sistema de conteo de ocupantes.

## II. ARQUITECTURA DEL SISTEMA

Para dar solución a nuestra propuesta, se plantea la instalación de una o dos cámaras del tipo Unibrain Fire-I [3] con resolución VGA, CCD de 1/4” sobre las que se montarán lentes tipo “ojo de pez” modelo DSL215 de Sunex [4]. Dicho montaje permite conseguir un campo de visión de unos 145° aproximadamente que captura video en tiempo real del interior del vehículo.

En un planteamiento inicial se pensó en la disposición de dos cámaras como las indicadas en el párrafo anterior centradas en el techo interior del vehículo, orientadas de tal forma que enfocasen por un lado al piloto y copiloto y por otro, a los pasajeros de los asientos traseros.

Tras las primeras pruebas en el vehículo se descubrió una posibilidad más sencilla y perfectamente válida como primera opción: la utilización de una única cámara colocada en el techo justo encima del espejo retrovisor interior y enfocada de tal modo que todos los pasajeros aparezcan en las imágenes capturadas. Dicha elección proporciona además las ventajas de facilidad de integración en el vehículo (ubicación, tamaño pequeño, menor consumo eléctrico) y reducción de los costes. En la figura 1 se puede ver la disposición elegida:

José Javier Yebe del Departamento de Electrónica de la Universidad de Alcalá 28805.

E-mail: jjavier.yebes@gmail.com

Luis M. Bergasa del Departamento de Electrónica de la Universidad de Alcalá 28805.

E-mail: bergasa@depeca.uah.es

Pablo F. Alcantarilla Del Departamento de Electrónica de la Universidad de Alcalá 28805

E-mail: pablo.alcantarilla@depeca.uah.es

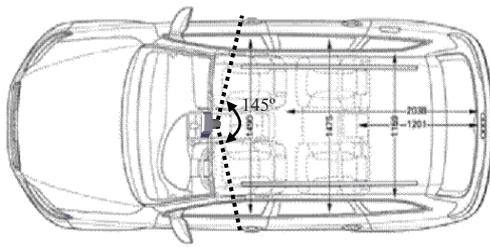


Fig. 1. Esquema del vehículo y colocación de la cámara en su interior así como el ángulo de visión cubierto.

Los riesgos principales de esta disposición conciernen al ángulo de visión máximo del montaje cámara más lente tipo “ojo de pez”. Las caras de los pasajeros podrían verse cortadas ligeramente en los bordes de las imágenes en función de la postura que éstos adopten. Este pequeño inconveniente podría solucionarse mediante el uso de cámaras con un CCD mayor, que además facilitaría el proceso de eliminación de distorsión como se verá más adelante.

En la siguiente figura 2 tomada de Sunex [4] observamos con mayor claridad lo explicado en el párrafo anterior:

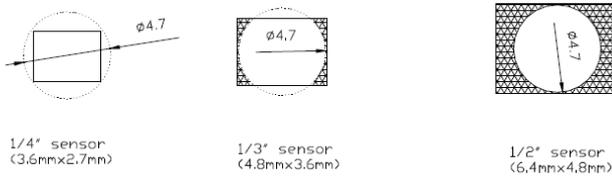


Fig. 2. Círculo de la imagen proporcionada por la lente tipo “ojo de pez” y rectángulo sensible de la cámara (CCD). En el CCD de ¼”, la imagen es recortada y se reduce el campo de visión por los cuatro lados de la imagen.

En la figura 3 puede verse un ejemplo de imagen capturada en el interior del vehículo con varios pasajeros.



Fig. 3. Imagen capturada en el interior del vehículo con la cámara montada según se ha explicado en los párrafos anteriores.

### III. PREPROCESADO DE IMAGEN

#### A. Calibración.

El proceso de calibración permite identificar un modelo geométrico para la cámara y un modelo de distorsión para la

lente, que se traducen en una serie de parámetros intrínsecos y extrínsecos que caracterizan el montaje explicado en la sección anterior. En nuestro caso estos parámetros son de interés para conseguir eliminar la distorsión radial (la más significativa que introducen las lentes tipo “ojo de pez”) y posteriormente trabajar con las imágenes ya rectificadas.

En los párrafos siguientes se describen los fundamentos teóricos del proceso de calibración [6] [11]:

Los parámetros intrínsecos relativos a la distancia focal y el punto principal (intersección del plano imagen con el eje óptico de la cámara) se calculan mediante la matriz de transformación mostrada en la siguiente ecuación (1):

$$q = MQ, \text{ donde } q = \begin{bmatrix} x \\ y \\ w \end{bmatrix} \quad M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad Q = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (1)$$

La ecuación (1) describe la proyección de un punto en el espacio real 3D (X,Y,Z) a un punto en el plano de imagen 2D (x,y). En el plano de proyección (CCD) se han de considerar unas coordenadas homogéneas, tal que los puntos en dicho plano se representarán por un vector 3D (q<sub>1</sub>,q<sub>2</sub>,q<sub>3</sub>), pudiéndose obtener las coordenadas originales del punto dividiendo entre q<sub>3</sub>; de ahí que q=(x,y,w).

Para completar la caracterización interna de la cámara, se calculan 5 coeficientes (k<sub>1</sub>, k<sub>2</sub>, k<sub>3</sub>, p<sub>1</sub>, p<sub>2</sub>). Los k<sub>i</sub> modelan la distorsión radial de la lente y se obtienen a partir de las series de Taylor considerando hasta el coeficiente de tercer orden. Las ecuaciones (2) y (3) muestran la relación existente entre las coordenadas de un punto en la imagen corregida sin distorsión y el punto en la imagen original con distorsión radial.

$$x_{\text{corregida}} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \quad (2)$$

$$y_{\text{corregida}} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \quad (3)$$

El parámetro *r* representa el radio del punto analizado con respecto al centro óptico de la imagen. Los 2 parámetros p<sub>i</sub> restantes modelan la distorsión tangencial debida a defectos de fabricación, ya que lente no será exactamente paralela al plano imagen. Las ecuaciones (4) y (5) añaden dicho efecto al cálculo de las coordenadas de un punto en la imagen corregida con respecto a la imagen original.

$$x_{\text{corregida}} = x + [2p_1 y + p_2 (r^2 + 2x^2)] \quad (4)$$

$$y_{\text{corregida}} = y + [p_1 (r^2 + 2y^2) + 2p_2 x] \quad (5)$$

La implementación práctica del proceso de calibración hace uso de las funciones de la librería OpenCV-1.1.0 de Intel [5] [6]. Para calibrar el conjunto de cámara más lente, se necesita un objeto plano de dimensiones conocidas que consiste en una cuadrícula con varias filas y columnas de cuadros negros y blancos alternados a modo de tablero de

ajedrez. Para que los resultados de la calibración sean correctos, se capturan varias imágenes de dicho objeto en diferentes posiciones con respecto a la cámara. En base a las pruebas realizadas, es suficiente tomar 20 imágenes para obtener finalmente los parámetros intrínsecos con la exactitud apropiada.

### B. Eliminación de la distorsión.

Los parámetros calculados mediante la calibración, se utilizarán en este punto para rectificar la imagen y corregir principalmente la distorsión radial tan característica de las lentes tipo “ojo de pez” y así facilitar el posterior tratamiento de las imágenes. Sin embargo, tras las pruebas realizadas, se ha demostrado que la aplicación directa del algoritmo Viola-Jones sobre las imágenes distorsionadas, ha ofrecido resultados satisfactorios con algunas pequeñas salvedades y modificaciones. Estos aspectos se detallarán en la siguiente sección III “Detección de ocupantes”.

De nuevo se utilizan las funciones de Opencv para crear 2 matrices o mapas de conversión, uno por cada coordenada del plano imagen, en base a los parámetros intrínsecos calculados. Estos mapas definen la correspondencia píxel a píxel entre la imagen original distorsionada y la imagen final sin distorsión.

Los resultados son muy satisfactorios para los píxeles centrales de la imagen, sin embargo en los bordes, la imagen final queda demasiado distorsionada debido al tamaño del CCD ( $1/4''$ ) perjudicando el posterior procesamiento de la imagen. El efecto causado puede verse en las figuras 4 y 5:



Fig. 4. Imagen 640x480 tras la rectificación o eliminación de la distorsión. Solo se ve la zona central de la imagen debido a la transformación de píxeles que implica la rectificación.



Fig. 5. Imagen 640x480 rectificada y escalada previamente para verla al completo. Los bordes quedan muy distorsionados tras la rectificación.

La solución del problema pasa por la utilización de una cámara con CCD mayor ( $1/3''$  o  $1/2''$ ) lo cual añadiría mejoras al obtener una imagen cuadrada tras el proceso de eliminación de distorsión y un mayor ángulo de visión posibilitando la captura de imágenes que conservarían un mayor margen en los bordes con respecto a los pasajeros del vehículo. Esta opción se ha propuesto como uno de los trabajos futuros.

## IV. DETECCIÓN DE OCUPANTES

El método elegido para la detección de ocupantes se basa en la detección de caras existentes en las imágenes obtenidas de la cámara instalada en el interior del vehículo. Éstas aparecerán de perfil o frontalmente, con una determinada inclinación variable a lo largo de la secuencia de video, y en diferentes condiciones de iluminación. Serán necesarias varias transformaciones en torno al algoritmo base de detección de caras Viola-Jones [2].

### A. Haar-Classifíer y Viola-Jones

El algoritmo elegido para la detección y conteo de pasajeros está basado en el clasificador “Haar-Classifíer” [7] [8] implementado en la librería Opencv-1.1.0 de Intel [5] que a su vez procede de una mejora del detector de caras Viola-Jones [2] [7] [8] previamente desarrollado por Paul Viola y Michel Jones.

Este clasificador permite la detección de un amplio abanico de objetos en una imagen a partir de una serie de características del objeto reunidas en una base de datos, que se obtienen mediante una secuencia de imágenes de entrenamiento. El clasificador trabaja de tal modo que va sumando y restando regiones cuadradas de la imagen y delimitando mediante unos umbrales aquellas regiones candidatas que puedan contener el objeto buscado, a partir de las comparaciones de determinadas propiedades (denominadas *Haar features*) entre los píxeles de la imagen y los datos almacenados en la base.

El algoritmo está optimizado para no penalizar el tiempo de procesamiento, aspecto muy importante para el proyecto que nos ocupa dado que analizamos videos en tiempo real a la máxima velocidad posible (el objetivo sería procesar a 25 fps). Además, en base a Viola-Jones, se utiliza una forma de clasificación [9] con nodos de rechazo en cascada. Es decir, se trata de un árbol de decisión en el que cada nodo evalúa si el candidato ha de ser rechazado o no. Según éstas consideraciones, para detectar positivamente una región candidata que contiene el objeto, habría que recorrer todo el árbol hasta el final y se deberían cumplir todas las condiciones impuestas en los nodos intermedios. Por otro lado, cuando una región candidata fuese rechazada en algún nodo, se pasaría a procesar la siguiente región.

Adicionalmente, los nodos se agrupan para que los primeros en la cadena de detección sean aquellos con las características más significativas, de tal modo que, se

reduzcan los fallos de detección del algoritmo y los rechazos se produzcan lo antes posible reduciendo el coste de procesamiento.

Todo lo anterior resulta en un elevado rendimiento de detección y una baja tasa de rechazo como características esenciales del clasificador. Esto es, puede haber falsos positivos y es raro que haya falsos negativos.

### B. Detección de caras

Para poder detectar el número de ocupantes en el interior del vehículo, nos centraremos únicamente en la utilización del Haar-Classifier y Viola-Jones para detección de caras aplicados directamente sobre las imágenes originales con distorsión grabadas por el conjunto cámara VGA más lente tipo “ojo de pez”.

Junto con las librerías Opencv se proporcionan unos ficheros xml [10] de referencia para detecciones de caras frontales y de perfil, que contienen datos almacenados previamente mediante el largo proceso de entrenamiento con un gran número de imágenes, los cuales permitirán al algoritmo decidir si en las imágenes analizadas aparecen formas sospechosas de ser una o varias caras. Dichos ficheros serán los utilizados para detectar las caras de los pasajeros en las imágenes capturadas.

A pesar de las características esenciales que se indicaban en el apartado anterior, el proceso de detección es delicado y muy sensible a la orientación/rotación de las caras en la imagen, sobre todo cuando se intentan detectar caras de perfil. También hay que destacar la dependencia con los datos almacenados en los ficheros xml, ya que hemos recurrido a los datos ya existentes por defecto proporcionados junto con la librería Opencv. Las secuencias de entrenamiento podrían personalizarse y optimizarse para el caso que abordamos, el interior de vehículos, pero se trata de un proceso lento y tedioso. Debido a estos motivos hay que buscar soluciones para guiar el algoritmo clasificador [7] y facilitar la detección de las caras. Para ello, se tendrá en cuenta las siguientes consideraciones:

1. Según la colocación de la cámara explicada en la sección II, las caras del piloto y copiloto estarán de perfil en las imágenes capturadas la mayor parte del tiempo, salvo movimientos repentinos.

2. Las caras de los pasajeros traseros se capturarán frontalmente, salvo movimientos o posturas de perfil (por ejemplo al mirar por la ventanilla hacia el exterior).

3. El algoritmo “Haar-Classiffier” más “Viola-Jones” solo actúa sobre imágenes en escala de grises. Además, existen una serie de parámetros configurables del clasificador que pueden ajustarse para buscar los mejores resultados en lo que respecta a detecciones positivas y consumo de recursos.

4. Se supone que el análisis de los vídeos en tiempo real y el tiempo de respuesta del procesamiento realizado sobre ellos permitirá integrar fragmentos grandes del mismo, en torno a 30 segundos, de tal modo que se filtren los resultados del algoritmo Haar-Classiffier más Viola-Jones para entregar finalmente el número de ocupantes del vehículo.

### C. Adecuación de las imágenes como paso previo a la aplicación del detector de caras

Teniendo en cuenta las consideraciones del apartado anterior, hay que llevar a cabo una serie de acciones previas para adecuar las imágenes facilitando de este modo la labor de detector de caras y haciendo más robusta y fiable su ejecución.

En primer lugar se han diferenciado tres regiones rectangulares de interés con diferente tamaño a partir de la imagen de la cámara cuya resolución es de 640x480 píxeles. La primera región aísla la zona del piloto del resto de la imagen, la segunda región hace lo propio con el copiloto y por último la tercera permite el análisis del número de ocupantes en los asientos traseros del vehículo. En la figura 6 aparece un esquema de la distribución planteada teniendo en cuenta que la cámara se coloca muy cerca del espejo retrovisor interno y está orientada hacia el interior del vehículo. Se observa como existen zonas de la imagen que quedan fuera de las regiones de interés y otras solapadas. Esto está justificado por el análisis empírico de las secuencias, que ha permitido seleccionar las regiones de interés y descartar aquellas zonas que no contienen información relevante.

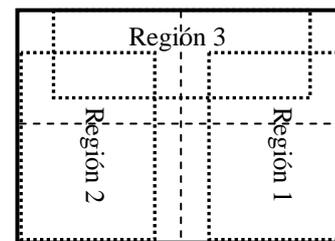


Fig. 6. Esquema de selección de las regiones de interés en las imágenes capturadas con la cámara VGA del interior del vehículo.

Con carácter general, las caras del piloto y copiloto aparecerán de perfil en las imágenes capturadas por la cámara, a excepción de los movimientos naturales de la persona a lo largo del tiempo. Obviando en este momento la inclinación del rostro, las posibles rotaciones de la cabeza harán que aparezca una mayor o menor área visible de la cara en la imagen bajo análisis. Existirán pocas ocasiones en las cuales la región de interés contenga una cara frontal al completo, por tanto, las regiones 1 y 2 delimitadas anteriormente son procesadas aplicando el detector de caras de perfil [10].

Por otro lado, en la región 1 aparece el perfil derecho de

la cara del piloto y en la región 2 aparece el perfil izquierdo de la cara del copiloto. Tras una serie de pruebas iniciales y para mejorar el rendimiento de detección de la cara del piloto, es necesario realizar una reflexión sobre el eje vertical de la región 1 que convierte la cara del piloto en un perfil izquierdo.

Además de la rotación de la cara descrita antes, que provoca la visibilidad de una mayor o menor superficie de la misma, se ha planteado la problemática referente a otros movimientos de la cabeza que implican una inclinación del rostro en la imagen. El algoritmo “Viola-Jones” más “Haar-classifier” es sensible a la orientación de las caras contenidas en las regiones de interés bajo estudio, principalmente tomando como referencia la línea horizontal imaginaria que pasaría por los ojos. Para ayudar y facilitar la labor de dicho algoritmo es necesario rotar las regiones 1 y 2 un cierto ángulo, aumentando así el rendimiento en el número de detecciones positivas.

En relación a la región de interés número 3, en general contendrá las caras frontales de los posibles ocupantes de los asientos traseros. Aunque de nuevo se han de considerar los posibles movimientos éstos. Tras los primeros análisis, no se requieren pasos previos y puede aplicarse directamente el detector de caras frontales ya que permite detectar varias caras a la vez en caso de que existan en la imagen. Aún así, podrían incluirse mejoras combinando la detección con caras de perfil que contemplaría aquellos casos en los cuales los pasajeros de atrás permanezcan durante mucho tiempo con la mirada al exterior del vehículo de tal modo que se encontrarían de perfil con respecto a la cámara.

#### D. Aplicación del detector de caras a las diferentes regiones de interés preseleccionadas

A continuación se describen los parámetros de ajuste del detector y los valores utilizados para cada una de las regiones:

→ *Rectángulo mínimo de búsqueda.* Indica el tamaño del rectángulo más pequeño que usará el detector para comenzar la búsqueda de posibles caras. Afecta a la velocidad de procesamiento y a la granularidad del algoritmo y depende de la resolución de las imágenes. Para la cámara VGA usada, se ha configurado un tamaño de 70x70 píxeles para las regiones 1 y 2 (primer plano de la cámara), y de 30x30 para la región 3 (plano alejado de los pasajeros de atrás). Obteniendo unos resultados de compromiso entre velocidad y granularidad.

→ *Escalado.* El clasificador en cascada realiza varias pasadas por las subregiones rectangulares candidatas en la imagen bajo análisis. Es decir, a partir del mínimo rectángulo definido, efectúa varias iteraciones con saltos de tamaño en dichos rectángulos según el factor de escala

indicado. Para acelerar el algoritmo manteniendo un buen rendimiento de detección en videos de tiempo real, el valor aconsejado es de 1.2; esto es, aumento del 20% del tamaño en cada nueva iteración.

→ *Número mínimo de vecinos.* Este parámetro relaciona el número mínimo de detecciones positivas solapadas en píxeles cercanos, que han de producirse para considerar un área rectangular candidata como verdadera zona que contiene la cara frontal o de perfil buscada. Básicamente permite rechazar falsos positivos, aunque el valor no debe ser muy alto porque se reduciría significativamente el rendimiento de la detección. El valor escogido en base a las pruebas realizadas es de 2.

→ *Filtro de Canny más algoritmo “pruning”.* Se ha configurado el detector para que internamente haga uso de un filtro de Canny para rechazar ciertas regiones de la imagen bajo análisis que pueden contener muchos bordes y no son candidatas a contener una cara. En resumen, el algoritmo “pruning” agiliza la carga de procesamiento rechazando componentes parásitos no deseados en la imagen.

Para resumir los pasos anteriores y detallar los parámetros de configuración del detector, enumeramos a continuación las características de cada una de las regiones:

- Región 1. Piloto. Requiere reflexión respecto al eje vertical, y rotación negativa un cierto ángulo. Tamaño mínimo de búsqueda 70x70 píxeles, y aplicación del detector de caras usando la base de datos para caras de perfil.

- Región 2. Copiloto. Requiere rotación negativa un cierto ángulo. Tamaño mínimo de búsqueda 70x70 píxeles, y aplicación del detector de caras usando la base de datos para caras de perfil.

- Región 3. Pasajeros traseros. No requiere reflexión ni rotación. Tamaño mínimo de búsqueda 30x30 píxeles, y aplicación del detector de caras usando la base de datos para caras frontales.

Por último, los resultados que nos entrega el “Haar Classifier” son integrados durante una determinada ventana temporal (entre 20 y 30 segundos) ya que los videos en tiempo real que captura la cámara se consideran bastante estables en el tiempo en lo referente al número de personas por detectar en el interior del vehículo. De este modo podemos permitirnos usar un tiempo de establecimiento para el filtrado relativamente alto aumentando la fiabilidad del resultado final, esto es, el número calculado de ocupantes del vehículo.

## V. RESULTADOS EXPERIMENTALES

Se han grabado cuatro videos de distinta duración y con diferente número de ocupantes en el interior del vehículo tanto parado como en movimiento. Usando dicho material, se ha procedido a realizar un análisis offline mediante el método explicado en las secciones anteriores.

En relación al ángulo de rotación necesario para el procesamiento de las regiones 1 y 2 seleccionadas a partir de las imágenes capturadas, se ha probado con un valor fijo no adaptativo, elegido a partir de diferentes pruebas:  $-30^\circ$  para la región 1 que contiene al piloto y  $-45^\circ$  para la región 2 que contiene al copiloto.

Se ha establecido una ventana temporal de 20 segundos para la integración de los resultados que entrega el detector de caras después de haber sido aplicado a cada región de interés, obteniendo como resultado el número de ocupantes del vehículo.

Las figuras 7 y 8 que aparecen a continuación corresponden a una imagen de uno de los videos analizados. Estas figuras muestran claramente un ejemplo de los resultados obtenidos tras la aplicación del algoritmo de detección de caras explicado en las secciones anteriores.



Fig. 7. Separación en regiones de una imagen. Reflexión y rotaciones en las regiones de piloto y copiloto tal como se explica en el apartado C de la sección IV. Detección de caras en las regiones.



Fig. 8. Imagen original capturada por la cámara con el resultado de la detección de ocupantes visualizado como un número en el centro.

Los resultados preliminares arrojan unos porcentajes muy interesantes que apoyan la continuidad y ampliación del trabajo expuesto en este artículo. En la tabla I siguiente se muestra el porcentaje de aciertos para cada uno de los videos utilizados considerando el resultado final en número de ocupantes del vehículo, obtenido tras la ejecución del algoritmo explicado en este documento.

TABLA I  
RESULTADOS PRELIMINARES DEL CONTEO DE OCUPANTES

	% aciertos en el conteo de ocupantes
Video 1 (58'')	100%
Video 2 (3' 16'')	84%
Video 3 (1' 44'')	75%
Video 4 (3' 47'')	82%

Tras una serie de pruebas ampliadas considerando ángulos de rotación variables para las regiones 1 y 2, que se adaptan a la secuencia de imágenes capturada, el porcentaje de aciertos aumenta, aunque se penaliza de forma significativa el tiempo de procesamiento.

Si se presta atención al procesamiento de las imágenes en los periodos transitorios cuando los pasajeros suben y bajan del vehículo, pueden aplicarse una serie de mejoras para favorecer los resultados finales.

Otras características que también pueden afectar en el análisis de las imágenes por parte del detector son el tamaño del CCD y el recorte de bordes en la imagen capturada; y en menor medida, las condiciones de iluminación

## VI. TRABAJOS FUTUROS

Para abordar la problemática de la orientación de las caras en distintos ángulos a lo largo del video, se ha planteado la posible aplicación de un filtro de Canny [13] para detectar contornos en conjunción con un análisis posterior de la fase de los píxeles filtrados. Se pretende establecer una relación entre dicha fase y el ángulo de orientación de la cara en la imagen. Tras las primeras pruebas, no ha habido resultados concluyentes y se sigue trabajando en ello.

También se están revisando los procedimientos sugeridos por los métodos FMVFD (Fast Multi-View Face Detection) [14] y WFS (Width First Search) [15] para detección de caras con independencia de su rotación en la imagen.

Otra mejora adicional sería la aplicación del detector de perfiles en conjunción con el frontal que ya se utiliza para los pasajeros de atrás. De este modo aumentaría el rendimiento en el número de detecciones positivas al contemplar los casos en los que los pasajeros traseros tengan la cara girada y no se aprecien bien los ojos en posición frontal a la cámara.

Si abordamos de nuevo el formato de las imágenes

capturadas por la cámara, se ha propuesto un trabajo pendiente de validar que consistiría en la utilización de cámaras con CCD mayor que aumenten el campo de visión (hasta unos 180° aproximadamente). Posteriormente aplicar la eliminación de la distorsión radial para rectificar las imágenes capturadas que serán las muestras entrantes para el “Haar-Classifer” [7]. El resultado esperado es una menor dependencia del ángulo de orientación de las caras, y un mayor rendimiento de detección.

Dado que las condiciones de iluminación en el interior del coche pueden cambiar debido principalmente a cambios externos de la iluminación ambiental, se ha planteado el funcionamiento nocturno del sistema mediante la adición de leds infrarrojos y lentes diseñadas para capturar imágenes a tal efecto.

Otra realización futura es la clasificación de los pasajeros en adultos, niños o bebés. Se estudiará la aplicación de redes neuronales o SVM (Support Vector Machine) [16] para dicho objetivo.

Por último, también se han propuesto nuevas funcionalidades como son: el almacenamiento de los videos en formato 3gp para su transmisión vía GPRS y la variación del tiempo de integración de los resultados de salida del “Haar Classifier” según un flag de puerta abierta, para que el tiempo de establecimiento del filtro se reduzca en dicha situación en la que puede haber un trasiego de personas subiendo/bajando del vehículo.

#### AGRADECIMIENTOS

Este trabajo ha sido financiado mediante fondos del Ministerio de Ciencia e Innovación a través del proyecto DRIVER-ALERT (TRA2008-03600) y del proyecto Robocity 2030 (CAM-5-0506/DPI/000176) concedido por la Comunidad de Madrid.

#### REFERENCIAS

- [1] “Intelligent car: eCall” European Commission, Information Society: [http://ec.europa.eu/information\\_society/activities/intelligentcar/technologies/tech\\_07/index\\_en.htm](http://ec.europa.eu/information_society/activities/intelligentcar/technologies/tech_07/index_en.htm)
- [2] P. Viola and M. Jones. “Rapid object detection using a boosted cascade of simple features”. Published December 2001. In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Kauai, HI.
- [3] Especificaciones de las cámaras modelo Unibrain Fire-I utilizadas: [http://www.unibrain.com/Products/VisionImg/Fire\\_i\\_DC.htm](http://www.unibrain.com/Products/VisionImg/Fire_i_DC.htm)
- [4] Sunex Digital Imaging Optics. Lente tipo “ojo de pez” modelo DSL215: [http://www.optics-online.com/dsl\\_quarter.asp](http://www.optics-online.com/dsl_quarter.asp)
- [5] OpenCV Library: <http://opencvlibrary.sourceforge.net>
- [6] Gary Bradski and Adrian Kaehler, “Learning OpenCV”, O’REILLY, First Edition, September 2008, Chapter 11 “Camera Models and Calibration”.
- [7] Gary Bradski and Adrian Kaehler, “Learning OpenCV”, O’REILLY, First Edition, September 2008, pp. 506-513.
- [8] Stefan Wender, Otto Loehlein, “A cascade Detector approach applied to vehicle Occupant Monitoring with an Omnidirectional camera”. Published 2004. *IEE Intelligent Vehicles Symposium*. University of Parma.
- [9] Gary Bradski and Adrian Kaehler, “Learning OpenCV”, O’REILLY, First Edition, September 2008, pp. 495-501.
- [10] Ficheros xml incluidos en la librería OpenCV [5] que contiene los datos necesarios para aplicar el detector de caras de perfil (haarcascade\_profileface.xml) o el detector de caras frontales (haarcascade\_frontalface\_alt2.xml).
- [11] Hynek Bakstein, Michal Havlena, Petr Pohl, Tomás Pajdla, “Omnidirectional sensors and their calibration for the Dirac project” (version 1.0) – Published Research report 2006. Czech Technical University. <ftp://cmp.felk.cvut.cz/pub/cmp/articles/bakstein/CMP-TR-2006-13.pdf>
- [12] Janne Heikkilä and Olli Silvén, “A Four-step Camera Calibration Procedure with Implicit Image Correction”, Infotech Oulu and Department of Electrical Engineering, University of Oulu. [http://www.vision.caltech.edu/bouguetj/calib\\_doc/papers/heikkila97.pdf](http://www.vision.caltech.edu/bouguetj/calib_doc/papers/heikkila97.pdf)
- [13] Gary Bradski and Adrian Kaehler, “Learning OpenCV”, O’REILLY, First Edition, September 2008, Chapter 6 “Image Transforms”.
- [14] Michael Jones, Paul Viola, “Fast Multi-view face detection”. Published July 2003. Mitsubishi Electric Research Laboratories.
- [15] Chang HUANG, Haizhou AI, Yuan LI and Shihong LAO, “Vector boosting for rotation invariant multi-view face detection”. 2005 Computer Science and Technology Department, Tsinghua University, Beijing, China.
- [16] Gary Bradski and Adrian Kaehler, “Learning OpenCV”, O’REILLY, First Edition, September 2008, pp. 517.