

Sistema de reconocimiento automático de la información de pórticos de carreteras usando visión computacional

Manuel López, Luis M. Bergasa, Miguel Gavilán, Álvaro González, Miguel Á. Sotelo

Abstract—Los sistemas de visión artificial aplicados al mantenimiento de carreteras, ya sea de la señalización o del estado de las mismas, cobran un protagonismo creciente en muchos países debido a la cada vez mayor inversión en obras públicas de este tipo. Estos sistemas permiten recopilar información diversa de forma automática y rápida con el objetivo de llevar a cabo acciones para la mejora de la seguridad en las carreteras. En este sentido, la adecuación de la información contenida en las señales situadas sobre la carretera o pórticos, es de vital importancia para un uso eficiente y seguro de la vía por parte del usuario. El presente estudio abordará las diferentes partes para llevar a cabo un sistema capaz de realizar labores de inventariado y comprobación de la fidelidad de la información contenida en éstos y cuyo fin será formar parte de un sistema de tiempo real empotrado en un vehículo de mantenimiento equipado con una cámara de vídeo.

Index Terms— Pórtico, tracking, segmentación de la información, clustering, OCR, k-medias, modelo oculto de Markov.

I. INTRODUCCIÓN

En todo proyecto de infraestructuras destinadas al tránsito de personas o vehículos de cualquier tipo es imprescindible llevar a cabo un gran esfuerzo en materia de señalización. Ésta, más aún si se trata de infraestructuras destinadas al uso público, debe responder a criterios como son la claridad en la presentación de la información, la sencillez y la uniformidad o uso de estándares. Todo ello, en virtud de un correcto uso de la misma y de una mayor seguridad para el usuario.

En esta materia, se ha hecho un especial esfuerzo en los últimos años por parte de las administraciones públicas y privadas debido principalmente al aumento considerable de proyectos de este tipo en países como España. Se ha puesto un especial énfasis no solo en la creación y emplazamiento de señalización más intuitiva y clarificadora, sino además en la planificación y puesta en marcha de tareas de mantenimiento acordes con la importancia de dar al usuario un servicio de mayor calidad y capaz de garantizar su propia seguridad.

En esta línea se engloban los estudios descritos en este

artículo, centrándose éstos esencialmente en el reconocimiento y análisis de la información contenida en los pórticos de las carreteras de forma automática. Ésta es, en líneas generales, más susceptible a cambios que aquella que proporcionan las señales icónicas. Debido al hecho de que éstas contienen elementos léxicos tales como nombres de poblaciones, vías, lugares importantes, etc, deberá ser objeto de una mayor atención por parte de las administraciones en cuanto a la corrección, facilidad de visualización para cualquier persona y actualización de la misma.

El presente trabajo sirve como complemento a otro llevado a cabo por el Departamento de Electrónica de la Universidad de Alcalá [1] que se ocupa de la comprobación del correcto estado de las señales de tráfico en la vía, sean estas del tipo que sean, para que cumplan con los estándares en cuanto a lo que se refiere a los niveles de retroreflectancia de las mismas. Para ello, se dispone de un vehículo equipado con una cámara de vídeo en escala de grises capaz de captar señal en la banda infrarroja y un emisor en dicha banda. La toma de la señal de vídeo se realizará de forma que se alternen frames en los que se iluminen las señales mediante el emisor infrarrojo con otros en los que no. Este método, importante para las medidas de retroreflectancia, también será de utilidad para el caso que nos atañe, tal y como se comentará más adelante.

II. OBJETIVOS

El objetivo principal del presente estudio es, por tanto, el desarrollo de un sistema basado en técnicas de visión computacional para la realización automática del inventariado de la información contenida en los pórticos, así como la verificación del estado y corrección de la misma.

Las principales tareas a desarrollar serán, en un primer lugar, la identificación de los pórticos en el frame capturado por la cámara, distinguiéndolos de carteles que pudiera haber a ambos lados de la carretera así como de otros elementos de formas similares. Una vez que éstos hayan sido identificados, se realizará un tracking de los mismos con el objetivo de no perder su referencia aún en casos en los que se encuentren total o parcialmente ocluidos. Al mismo tiempo, y siempre que sea posible, se llevará a cabo la extracción de la información contenida en el pórtico representando ésta de una forma jerarquizada en función de los elementos informativos y subpaneles que puedan hallarse. Esta información estará compuesta por elementos léxicos, formantes de palabras; identificadores alfanuméricos para los nombres de carreteras, salidas, etc; símbolos y por último, por elementos de separación como

Manuel López, Luis M. Bergasa, Miguel Gavilán, Álvaro González, Miguel Á. Sotelo pertenecen al Departamento de Electrónica de la Universidad de Alcalá.

E-mail: manuel.lopez@madrimasd.net, bergasa@depeca.uah.es, miguel.gavilan@depeca.uah.es, alvaro.g.arroyo@gmail.com, sotelo@depeca.uah.es

cuadros informativos, cajas de texto, etc.

Todo esto se llevará a cabo teniendo en cuenta que los algoritmos desarrollados deberán implementarse en un sistema de tiempo real empotrado en un vehículo habilitado para ello, por lo que es de vital importancia que el procesamiento de cada frame capturado no sea muy pesado y pueda realizarse a tiempo.

III. IDENTIFICACIÓN DE LAS DISTINTAS PARTES DEL PÓRTICO

Es necesario, antes de pasar a un estudio más exhaustivo, definir las partes en las que se dividirá un pórtico. En la Fig. 1 se presenta un ejemplo de lo que llamaremos *pórtico completo*. Éste, a su vez, podrá estar dividido en *pórticos individuales* o pórticos sin más, cada uno de los cuales contendrá uno o más *cuadros informativos*. La forma de localizar éstos últimos será identificando los recuadros de esquinas redondeadas por los que están delimitados. Dentro de éstos es donde se encuentra la información propiamente dicha. En la Fig. 2 puede verse la jerarquización realizada a partir del pórtico completo bajo estudio en este ejemplo. En ocasiones, dentro de un cuadro informativo puede existir otro cuadro de texto con información relevante. En este caso, se añadirá un elemento más a la jerarquía por debajo del cuadro informativo correspondiente, que volverá a tener textos, identificadores o símbolos.



Fig. 1. Ejemplo de un pórtico completo

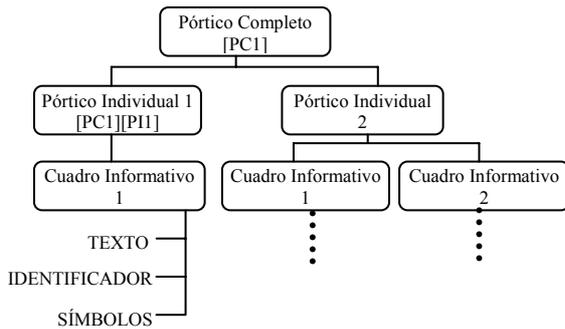


Fig. 2. Jerarquización de la información en el pórtico

IV. DETECCIÓN DE PÓRTICOS

Una aproximación de la detección de los pórticos vendrá heredada del otro trabajo desarrollado por el Departamento de Electrónica, tal y como se comentó en la *Introducción*. La información recibida serán los rectángulos contenedores de los candidatos a ser pórticos. Estos candidatos, en algunos casos se tratarán de carteles publicitarios situados en los márgenes de la vía, matrículas de coches, carteles de señalización kilométrica, etc. Por ello, la primera labor será la de descartar todos éstos. Esta discriminación se realizará

en base a su localización en la imagen utilizando la máscara definida por la Fig. 3 Sólo se aceptarán aquellos candidatos contenidos completa o parcialmente en la zona blanca de ésta. Con ello se elimina la práctica totalidad de los erróneos detectándose más adelante aquellos que hayan podido pasar desapercibidos.

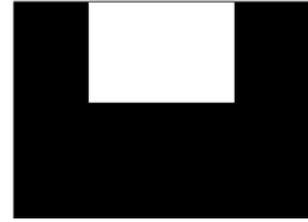


Fig. 3. Máscara aplicada a la imagen para la discriminación entre los candidatos a pórticos.

V. TRACKING

Una vez que se ha detectado un posible candidato a ser un pórtico a partir de la información de un sólo frame, el sistema esperará a que éste aparezca en frames sucesivos para seleccionarlo como tal y comenzar el tracking o seguimiento del mismo. Este proceso es importante debido a que, durante el transcurso de la secuencia desde la cual el pórtico es reconocido hasta que desaparece de la imagen, se pueden producir oclusiones de tipo parcial o total, así como circunstancias en las cuales la iluminación o cualquier otro factor hagan que no se detecte el mismo de forma correcta, perdiendo su referencia. Por ello, se implementará un modelo predictivo de la posición del pórtico en cada frame a partir de la información de su movimiento en los predecesores. La técnica utilizada para ello se basará en dos filtros de Kalman [2] que predecirán el movimiento de la esquina superior izquierda y de la inferior derecha del pórtico por separado. Para ello se definirá el vector de estados para ambos puntos:

$$\begin{aligned} x_1 &= (c_1, r_1, v_{c1}, v_{r1}) \\ x_2 &= (c_2, r_2, v_{c2}, v_{r2}) \end{aligned} \quad (1)$$

Donde (c_i, r_i) será el vector de posición y (v_{ci}, v_{ri}) el vector de velocidades del punto i .

Las ecuaciones de predicción del siguiente estado serán:

$$\begin{aligned} x_i(n+1/n) &= F \cdot x_i(n) + W_t \\ z_i(n+1/n) &= H \cdot x_i(n+1/n) + V_t \end{aligned} \quad (2)$$

Con z_i el vector de observación, F la matriz de actualización de estado en función del estado anterior, H la matriz que relaciona el estado con la medida y por último W_t y V_t las componentes de ruido blanco gaussiano del proceso y de la medida respectivamente.

$$\begin{aligned} W_t &\rightarrow N(0, Q) \quad Q = cte \\ V_t &\rightarrow N(0, R) \quad R = cte \end{aligned} \quad (3)$$

Se utilizarán las siguientes matrices de covarianza de

ruido:

$$Q = \begin{bmatrix} 16 & 0 & 1 & 0 \\ 0 & 16 & 0 & 1 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix} \quad R = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} \quad (4)$$

Las matrices principales, F y H serán:

$$F = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (5)$$

Con ello el modelo quedará completamente definido.

VI. SEGMENTACIÓN

El problema de la segmentación tendrá como objetivo distinguir los diferentes elementos dentro del pórtico. Por ello, se trata de la tarea más difícil y específica del sistema completo. Los elementos en los que nos centraremos serán sobre todo aquellos que marcarán la división de la información dentro del propio pórtico, que suelen estar delimitados por líneas, textos y símbolos. Será importante conocer las propiedades de los pórticos para así adaptar el algoritmo a éstas y conseguir mejores resultados.

En primer lugar se realizará un estudio del color de los mismos, teniendo en cuenta que la imagen de la que se dispondrá para su análisis estará en escala de grises. Atendiendo a este criterio se distinguirán dos tipos de pórticos:

a) *Pórticos de fondo blanco (Tipo 1)*, en los cuales las letras, líneas y símbolos suelen ser de color negro u otro que destaque sobre el blanco. En una imagen en escala de grises estos pórticos se verán de una forma similar a la mostrada en la Fig. 4(a).

b) *Pórticos de fondo coloreado (Tipo 2)*, donde tanto letras como líneas y símbolos suelen ser blancos. Independientemente del color utilizado para el fondo, dentro de los definidos en las reglamentaciones correspondientes, la imagen en escala de grises de un pórtico de este tipo será del estilo de la presentada en la Fig. 4(b).

Como se observa en ambas figuras, el contraste entre el fondo y la información es bastante pronunciado. Una primera aproximación al problema de separar el fondo del primer plano (el de la información) se basa en el estudio del histograma para cada tipo de pórtico. Ver Fig. 5 (a) y (b). A

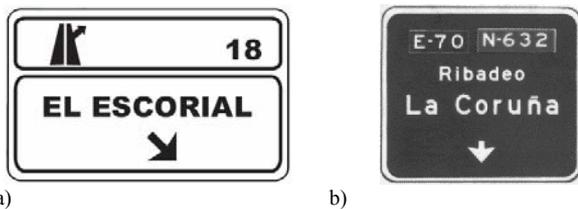


Fig. 4. a) Ejemplo de pórtico Tipo 1 b) Ejemplo de pórtico Tipo 2

la vista está como en ambos histogramas se aprecian claramente dos máximos. Uno de ellos, el mayor, será el debido al fondo y el otro a la información. Con esta información proporcionada por el histograma se puede distinguir el tipo de pórtico identificando el mayor de los máximos y comprobando si es el de la derecha (el más cercano al blanco), en cuyo caso se trataría de un tipo 1, o el de la izquierda (el más cercano al negro), tratándose en este caso de un tipo 2. El siguiente paso es el de hallar el umbral que separe perfectamente ambas regiones en el histograma, tomando como válido alguno de los situados entre los dos máximos. Con ello, sea $f(u,v)$ la función definida por la imagen en la que u define las coordenadas sobre el eje vertical y v las coordenadas sobre el eje horizontal, se definirá una nueva función o imagen umbralizada $g(u,v)$ a partir del umbral γ , de la siguiente forma:

$$f(u_0, v_0) \geq \gamma \rightarrow g(u_0, v_0) = \begin{cases} 0 & \text{si Tipo 1} \\ 255 & \text{si Tipo 2} \end{cases} \quad (6)$$

$$f(u_0, v_0) < \gamma \rightarrow g(u_0, v_0) = \begin{cases} 255 & \text{si Tipo 1} \\ 0 & \text{si Tipo 2} \end{cases}$$

La nueva imagen será en binaria, donde las zonas en negro (0) son el fondo mientras que las zonas en blanco (255) se tratarán de elementos pertenecientes a la información del pórtico.

Esta técnica de segmentación, sin embargo, presenta muchos inconvenientes. El principal es el debido al hecho de que la iluminación de los pórticos, en la gran mayoría de los casos, es altamente no uniforme. En tramos de carretera iluminados se logrará una mayor uniformidad en la iluminación sobre el pórtico ya que la fuente de luz más influyente será la procedente de las luminarias situadas al borde de la vía. Por otra parte, en tramos sin iluminación, será la luz procedente de los faros de los coches la que más influya y esta es bastante puntual. La utilización de la reflexión infrarroja, de la que se habló anteriormente, y la utilización de aquellos frames en los que se capta ésta, hará que se reduzca en gran medida la no uniformidad de la iluminación del pórtico. Véase en la Fig. 6(a) y (b) la diferencia tan notable entre el frame en el que se capta la iluminación infrarroja sobre el pórtico y aquel en el que no. No obstante, se ha observado que este efecto depende en gran medida de factores como la altura del pórtico, la distancia a este o la iluminación de la carretera, por lo que la

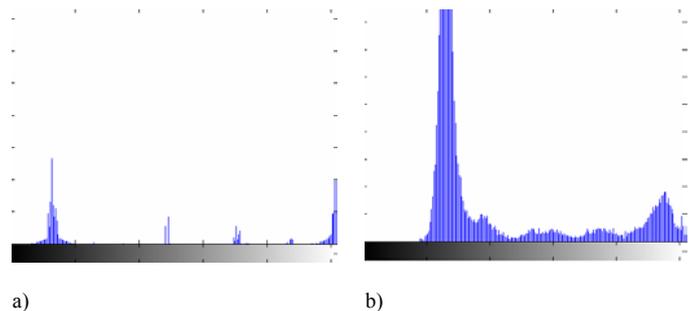


Fig. 5. a) Ejemplo de histograma de un pórtico Tipo 1 b) Ejemplo de histograma de un pórtico Tipo 2

casuística se hace muy grande.

El análisis experimental que da lugar a las conclusiones que acaban de presentarse justifica la necesidad de un método más robusto independiente de la distribución de la iluminación sobre el pórtico. Por ello, las técnicas de segmentación implementadas se basarán en la detección de bordes en la imagen y a partir de ellos la determinación de contornos. Sobre éstos se aplicarán reglas de decisión en función de propiedades conocidas tanto de los pórticos como de la información contenida en éstos para clasificarlos.

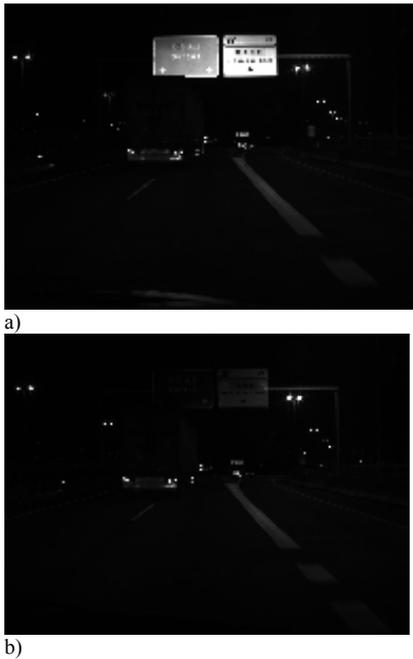


Fig. 6. a) Frame captado en el momento en el que se ilumina el pórtico con el emisor de IR b) Frame consecutivo, captado en el momento en el que el emisor de IR no emite.

La detección de bordes se basa en la obtención del gradiente direccional para cada píxel de la imagen, por lo que lo importante ahora será que la información destaque lo suficiente sobre el fondo para poder delimitar los bordes de textos, recuadros, símbolos, etc. El tipo de detector de bordes usado será el de Canny [3] debido a sus propiedades, como son, en primer lugar, la reducción de la probabilidad de falsa alarma (probabilidad de detectar borde cuando no lo hay) mediante la aplicación de un filtro de tipo Gaussiano que suaviza la imagen y, en segunda instancia, la decisión en base a dos umbrales para conseguir una menor probabilidad de pérdida (probabilidad de no detectar bordes, generalmente reducidos, cuando éstos existen). Los principales parámetros de este detector, por tanto, serán la desviación típica para el filtro Gaussiano (σ), y los dos umbrales (Th_1, Th_2). Experimentalmente se determinan los valores usados para éstos con los cuales los resultados son aceptables para las etapas posteriores:

$$\sigma = 1 ; Th_2 = 0.2 ; Th_1 = 0.08 \quad (7)$$

A la salida del detector obtendremos una imagen en la cual se aprecian los bordes detectados en color blanco y el

resto en negro. Sobre ésta se aplicará el algoritmo que nos proporcionará los contornos. Con ellos, el primer paso será realizar una primera decisión sobre el tipo de información de que se trata. El procedimiento será el siguiente:

a) Comprobar si se trata de un elemento de tipo rectángulo, candidato a ser aquellos que delimitan los elementos informativos, cajas de texto o incluso el borde del pórtico completo. En caso contrario,

b) Comprobar si se trata de textos o números atendiendo a un margen de proporcionalidad experimental entre la altura y la anchura de un carácter con respecto a la altura del pórtico. En caso contrario,

c) Se considerará como un objeto sin determinar, entre los cuales estarán los símbolos como flechas, iconos, etc.

Con respecto al caso a) los supuestos que deben darse para que un contorno sea seleccionado como un cuadro contenedor de información son los siguientes:

a.1. La altura y la anchura de la caja que encierra al contorno candidato deben ser mayores que el 10% de la altura y anchura del pórtico. Esto evitará seleccionar contornos demasiado pequeños.

a.2. Definiéndose como paso el avance de un píxel al inmediatamente posterior de entre los que componen un contorno, debe cumplirse que la proporción de pasos verticales y horizontales sumados debe ser mayor del 85% del total de pasos que definen el contorno a la vez que se da cualquiera de estas tres condiciones: 1.) el perímetro teórico definido por la caja contenedora del contorno es similar al real; 2.) el contorno es de tipo U o U invertida; 3.) el contorno es de tipo C o C invertida. Estas comprobaciones se llevarán a cabo con la ayuda de la representación del contorno en forma de código cadena [3]. El umbral del 85% es suficiente para que el efecto de la perspectiva sea tenido en cuenta, sabiendo que ésta nunca va a ser muy pronunciada.

Una vez que se han determinado los contornos candidatos a ser cuadros contenedores se decidirá en función de información sobre relaciones padre-hijo entre ellos, cruces de dos candidatos, dimensiones y características conocidas de los pórticos. En este punto, si el pórtico es suficientemente grande y no se ha detectado ningún recuadro con las características de aquellos que delimitan a los elementos informativos, se decidirá descartar a éste como pórtico.

En cuanto al caso b), los candidatos a ser letras o dígitos deberán cumplir las siguientes características:

b.1. El cociente entre la altura del pórtico (H) y la de la caja contenedora del contorno candidato (h) debe estar dentro de unos márgenes obtenidos de forma experimental (γ_1 y γ_2).

$$\begin{aligned} \gamma_1 < (H/h) < \gamma_2 \\ \gamma_1 &= 7.5 \\ \gamma_2 &= 25 \end{aligned} \quad (8)$$

b.2. El cociente entre la altura del pórtico (H) y la

anchura de la caja contenedora del contorno candidato (w) debe estar, de la misma forma, dentro de otros márgenes obtenidos de forma experimental (δ_1 y δ_2).

$$\begin{aligned} \delta_1 &< (H/w) < \delta_2 \\ \delta_1 &= 10 \\ \delta_2 &= 50 \end{aligned} \quad (9)$$

En la Fig. 7(a) se presenta la salida de este algoritmo, marcándose los contornos de las letras identificadas. Como puede verse, no será del todo efectivo debido a que habrá contornos que englobarán varias letras y por ello no serán seleccionados. Para solucionar esto se identificarán las posiciones en las que hay líneas de texto y posteriormente se verificará si los contornos no asignados y posicionados en línea con alguna de ellas cumplen la relación de altura definida en *b.l.*. En ese caso se marcarán como texto y se añadirán a la línea correspondiente. En la Fig. 7(b) se observa como después de aplicar esto último se mejora considerablemente la detección, distinguiéndose además cada una de las líneas de texto mediante un recuadro que las engloba. Para llevar a cabo este algoritmo se definirá una distancia mínima entre líneas igual a la mitad de la media de las alturas de todas las letras, valor obtenido de forma experimental y con el que se han conseguido buenos resultados. Si bien la detección de estos contornos no será útil para el reconocedor individual de caracteres descrito en el siguiente apartado sí será muy importante para más adelante saber delimitar palabras y que, en todo caso, el sistema sepa la extensión de ésta. El siguiente paso es la separación de palabras, cifras compuestas por varios números, etc. Para ello se tendrá en cuenta la separación media entre dígitos de forma similar a como se hace con el interlineado.

Una vez hecho esto, los dígitos que se han conseguido identificar individualmente estarán listos para ser introducidos a un OCR (Optical Character Recognition) para su clasificación. Por otra parte, aquellos contornos que fueron añadidos más tarde, y que en general englobaban varios dígitos, podrían ser objeto de un estudio más exhaustivo para conseguir individualizar cada una de las letras. Una técnica propuesta para ello es la erosión seguida de una umbralización y una nueva búsqueda de contornos.

Por último, se deben eliminar aquellas asignaciones que se encuentren aisladas en una línea y cuyo contorno no sea conexo para eliminar así asignaciones fallidas.

VII. RECONOCEDOR DE CARACTERES

Una vez han sido aislados los dígitos o caracteres, éstos deben ser reconocidos por el reconocedor de caracteres (OCR) con el objetivo de extraer la información que nos aportan de forma individual [4]. La implementación del OCR se lleva a cabo en dos etapas bien distintas: entrenamiento y test. En la etapa de entrenamiento, previa a la puesta en marcha del sistema, se ajustan los parámetros del modelo en base a unas entradas seleccionadas para el correcto ajuste del mismo. Debido a la disparidad de las muestras de entrada, éstas deben ser normalizadas y

descritas con un vector de características para cada una de ellas. Sea $c(u,v)$ la imagen en la que se encuentra aislado el carácter, se extrae el vector de características $[c_1, c_2, c_3, \dots, c_i]$ normalizadas según las dimensiones del mismo. Esta circunstancia hace que cada entrada sea representada por un punto en un espacio de i dimensiones. Si las características son lo suficientemente representativas del carácter, todos aquellos vectores que se correspondan con caracteres iguales se agruparán en el espacio formando *clusters*. La etapa de entrenamiento trata de formar una población de vectores en la que se contemple una casuística lo más elevada posible para que, más adelante, en la etapa de test un nuevo vector representante de un carácter de entrada pueda ser clasificado de forma correcta. El clasificador está basado en el algoritmo de las *k-medias*, cuyo objetivo es el de particionar el espacio de características en k particiones, intentando para ello localizar los centroides o puntos medios de cada cluster. Esta localización se hace en base a un criterio de minimización de la varianza total de cada cluster:

$$V = \sum_{i=1}^k \sum_{x_j \in S_i} (x_j - \mu_i)^2 \quad (10)$$

Donde μ_i se refiere al centroide del cluster S_i y x_j a cada uno de los elementos del mismo.

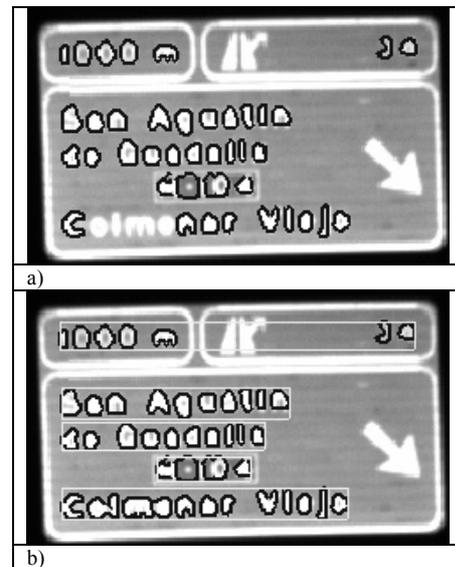


Fig. 7. a) Salida de la primera etapa del detector de caracteres. Sólo los hallados de forma individual son obtenidos b) Salida de la segunda etapa del detector. Se han añadido algunos que en el anterior no se consiguieron detectar y se han recuadrado las líneas de texto.

En este caso de estudio $k=37$ (27 letras del abecedario en español + 10 dígitos). Al comienzo del algoritmo los k centroides serán repartidos de forma semi-aleatoria y para cada vector de entrenamiento ha modificando la posición de los mismos atendiendo al criterio expuesto anteriormente.

En este tipo de algoritmos es muy importante saber elegir qué características son las que diferencien mejor unos caracteres de otros. Algunas de ellas, adaptadas al caso de estudio, son:

a) *Parámetros elípticos*, como pueden ser la

excentricidad o el ángulo del eje mayor.

b) *Masa total del carácter*, o lo que es lo mismo, el número de píxeles de éste normalizados respecto de la caja que lo contiene.

c) *Centro de masa* del carácter en coordenadas relativas a las dimensiones de la caja.

d) *Parámetros estadísticos* como el coeficiente de apuntamiento o *skewness* y la *curtosis*.

VIII. FORMACIÓN DE PALABRAS MEDIANTE MODELOS OCULTOS DE MARKOV

El número de palabras que contiene la información de los pórticos está limitada a los nombres de las poblaciones y a algunas otras contenidas en frases tales como: “vía de servicio”, “cambio de sentido”, etc. El hecho de que el conjunto de palabras sea finito hará que si el OCR no ha conseguido hallar todas las letras de una palabra, ésta se pueda intentar descubrir a partir de las que sí se han hallado correctamente. Para ello se aplicarán los modelos ocultos de Markov o HMM (*Hidden Markov Model*) [5], los cuales se fundamentan en los procesos de Markov. Estos últimos se basan en un diagrama de estados entre los cuales existen unas probabilidades de transición determinadas y para cada uno de los cuales existe una salida fija. No obstante, en los modelos ocultos de Markov, la función que asigna a cada estado una salida es una función de distribución de probabilidad, siendo $p_k(j)$ la distribución resultante de observar la salida j sobre el estado k . Esto se puede ver en la Fig. 8 de forma esquematizada. En el caso de reconocimiento de palabras a partir de letras reconocidas en la etapa anterior, cada palabra que contiene nuestro “diccionario” finito vendrá representada por un estado. Las salidas observables serán, en este caso las letras del abecedario, con lo que, para cada estado (palabra) tendremos una función densidad de probabilidad que lo relacione con cada una de las letras del vector de salidas observables. De esta forma, a partir de un vector de observaciones concreto, es decir, una secuencia de caracteres reconocidos, se decidirá que palabra o estado es el que lo genera con mayor probabilidad. Con ello queda resuelto el problema del reconocimiento de la información léxica en los pórticos de carreteras, aportando además una característica muy interesante como es la capacidad de detectar errores y efectuar correcciones.

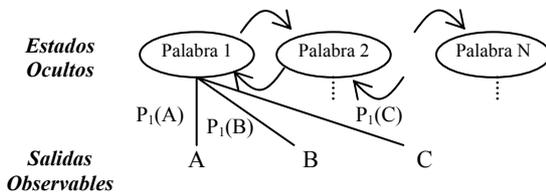


Fig. 8. Esquema de el modelo basado en cadenas ocultas de Markov

IX. MONITORIZACIÓN CONTINUA DE LA INFORMACIÓN

Una característica relevante del sistema es que el proceso de obtención de información de cada pórtico se realiza de forma progresiva sobre los frames en los que aparece éste, lo que quiere decir que no se trata de obtener toda desde el primer momento en el que se observa un pórtico

determinado. Esto es, en primer lugar, porque generalmente la primera observación se realiza cuando el pórtico está aún demasiado lejos para realizar un análisis con resultados lo suficientemente fiables, en segundo lugar, para solventar circunstancias en las que éste quede parcial o totalmente ocluido, frames en los que la iluminación no es óptima, etc; y por supuesto, para conseguir un algoritmo lo suficientemente robusto y cuyos resultados puedan darse con un nivel de fiabilidad muy alto. Mediante un breve análisis previo de las proporciones de los elementos del pórtico, puede estimarse si los resultados que se desprenden son o no significativos y de esta forma ahorrar tiempo de proceso.

Para identificar el momento a partir del cual el análisis en profundidad de los elementos del pórtico dan como resultado información fiable, se identifican elementos candidatos a ser textos y se compara su altura con la altura de la imagen captada por el video. En el momento en el que ésta sobrepasa un cierto umbral hallado de forma experimental se pasa a la siguiente etapa en la cual se aplica el Reconocedor de caracteres previamente descrito.

La altura umbral del texto es aquella para la cual los caracteres se reconocen con éxito en el 95% de los casos, utilizando para ello un OCR de características idénticas al que se aplicará en la etapa correspondiente del procesado.

Con todo ello, el sistema consigue realimentarse de la propia información captada en momentos anteriores dotándolo de una gran robustez.

X. RESULTADOS

El proyecto se encuentra en la actualidad en una fase intermedia previa a la aplicación de los modelos ocultos de Markov, habiéndose obtenido resultados experimentales de las salidas ofrecidas por el reconocedor de caracteres descrito. En base a las pruebas realizadas en la etapa de segmentación llegamos a concluir que aproximadamente el 90% de las letras son reconocidas como tal y el algoritmo es capaz de aislarlas para pasar posteriormente a la etapa de reconocimiento de caracteres. Esta cifra se da a partir del cómputo global de las que son aisladas desde que el pórtico empieza a ser reconocido a gran distancia hasta el último frame en el cual se obtiene información del mismo. Este 90% de letras son enviadas al OCR consiguiendo en éste unos resultados satisfactorios teniendo en cuenta que todavía restaría una etapa más para obtener la información final. Como se explicó, el objetivo de la etapa de OCR no es el de reconocer todas las letras de cada palabra sino ser capaz de clasificar correctamente una cantidad suficiente de letras como para que la salida del modelo de Markov sea correcta. Como ejemplo de los resultados obtenidos hasta el momento, se aplica como entrada el pórtico de la Fig. 7. Para éste, el OCR da la siguiente salida: “1000 m / 34 / San Agustln de Guadallx / (desconocido) / C(desconocido)nar Viejo”. Como se comprueba, queda pendiente un mejor ajuste y entrenamiento del reconocedor en determinados casos, pero incluso en el caso de introducir errores de clasificación en esta etapa éstos pueden ser corregidos en la siguiente siempre y cuando la información adicional sobre la palabra sea suficiente. Esto hace que la robustez del sistema final sea bastante elevada. La tabla representada en

la Fig. 9 muestra los resultados obtenidos con el OCR para las letras con las que se ha estado trabajando de forma experimental.

A	95	P	86	f	78	u	77
B	71	Q	68	g	92	v	73
C	70	R	74	h	84	w	81
D	70	S	83	i	69	x	79
E	85	T	89	j	91	y	84
F	90	U	72	k	75	z	81
G	72	V	74	l	68	1	72
H	87	W	87	m	90	2	84
I	73	X	84	n	93	3	90
J	92	Y	86	ñ	58	4	88
K	68	Z	82	o	60	5	83
L	85	a	88	p	89	6	75
M	89	b	72	q	84	7	82
N	76	c	74	r	72	8	73
Ñ	57	d	75	s	76	9	72
O	58	e	73	t	69	0	60

Fig. 9. Aciertos de detección en % del OCR.

Como puede comprobarse el sistema es susceptible de una cierta mejora. Con la ayuda de los modelos ocultos de Markov y, tal y como se ha comentado, una base con las estructuras léxicas apropiadas en las que se incluyan los nombres de las poblaciones, se pueden obtener resultados aceptables en lo que se refiere al reconocimiento de palabras al completo. Además, todo ello puede completarse con una mejora en el OCR que proporcione unos mejores resultados. Una posible línea podría ser implementar una red neuronal como reconocedor adicional o mejorar la segmentación de los caracteres para conseguir unas muestras de entrada con una mejor calidad. De esta forma, podría desarrollarse un sistema realmente fiable que solventase el problema con la robustez que demanda la aplicación.

REFERENCIAS

- [1] García-Garrido, M.A Sotelo, M.A Martin-Gorostiza, E. "Fast traffic sign detection and recognition under changing lighting conditions." *Dept. of Electron. , Alcalá Univ., Madrid*. This paper appears in: Intelligent Transportation Systems Conference, 2006. ITSC '06 IEEE. On pages: 811-816.
- [2] Greg Welch, Gary Bishop, "An Introduction to the Kalman Filter". *Department of Computer Science, University of North Carolina at Chapel Hill*. Technical Report 95-041.
- [3] David A. Forsyth and Jean Ponce. "Computer vision: A Modern Approach." *Prentice Hall. Pearson Education International*
- [4] Qing Chen, "Evaluation of OCR Algorithms for Images with Different Spatial Resolutions and Noises" A thesis submitted to the School of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Applied Science. *Ottawa-Carleton Institute for Electrical Engineering School of Information Technology and Engineering Faculty of Engineering University of Ottawa*.
- [5] Jeff Bilmes, "What HMMs Can Do" *Dept. of EE, University of Washington*

AGRADECIMIENTOS

Este trabajo ha sido financiado mediante fondos del Ministerio de Educación y Ciencia a través de los proyectos MOVICON (TRA2005-08529-C02-01) y CABINTEC (PSE-370100-2007-2), del proyecto RoboCity2030 (CAM-S-0505/DPI/000176) y del proyecto SLAM-MMEX (CCG07-UAH/DPI-1736) de la Comunidad de Madrid.

