

Medición de flujo vehicular basado en visión Artificial

Andrés Felipe Granados Zárate

Abstract— Se presenta el desarrollo de un sistema automático de medición del flujo vehicular basado en visión artificial. Dicho sistema se encarga de detectar y clasificar vehículos y peatones para determinar el flujo vehicular y peatonal en las intersecciones de calles y avenidas mediante secuencias de imágenes a color tomadas por una cámara de video estática. Estas secuencias de imágenes a colores son filtradas para evitar que los cambios de iluminación producto de movimiento de los objetos causen falsos resultados, empleando para ello un filtrado homomórfico. Para detectar los objetos en movimiento se compara la suma de la diferencia absoluta (SAD) entre dos secuencias de imágenes, mediante la umbralización y segmentación, técnicas que generaran una mascara binaria que contiene los objetos en movimiento y aísla los objetos ajenos a la escena de estudio. Posteriormente, se extraen las características individuales trazando las fronteras de los objetos y utilizando descriptores de Fourier, para luego realizar la distinción entre peatones y vehículos a través de una red neuronal *feed-forward*. En este trabajo se describe cada una de las etapas antes mencionadas y las pruebas de desempeño que permitirán para una futura implementación, el desarrollo de semáforos más eficientes y el monitoreo del tráfico.

Index Terms— Visión artificial, Redes Neuronales, Detección de Cambios, Filtros Homomórficos, Descriptores de Fourier, Umbralización.

I. INTRODUCCIÓN

Como consecuencia del gran desarrollo de la visión artificial y a los acontecimientos que han rodeado al mundo, ha surgido la necesidad de contar con sistemas de monitoreo y vigilancia en situaciones específicas, que superen las limitaciones del sistema de percepción humano en el sentido de atención, vigilancia y monitoreo. Aunque el ser humano se puede concentrar, por periodos cortos de tiempo, en eventos que considere importantes, la cantidad de información que se procesa e interpreta es usualmente limitada. Sin embargo, la concepción de visión artificial en monitoreo y vigilancia no se limita a reemplazar el sistema humano, sino también sirve como una herramienta de procesamiento de información que extienda la percepción y el razonamiento del mismo [1].

El manuscrito se recibió el 15 de julio de 2006. Este trabajo es patrocinado en parte por la Universidad del Quindío desde la academia.

Andrés Felipe Granados Zárate. Estudiante de Ingeniería Electrónica universidad del Quindío anfelec@yahoo.com, gdsproc@uniquindio.edu.co.

Una de las grandes problemáticas en la que la visión artificial puede ser una solución, es el monitoreo y vigilancia del tráfico. Una parte de este problema es reconocer y medir el flujo vehicular y peatonal, para maximizar el uso de los semáforos de tal forma que el tiempo de espera dependa de las necesidades del tráfico.

El primer paso para realizar esta tarea es detectar todos los objetos que realizan cambios en la escena. Como flujo se traduce en movimiento, entonces lo que se busca son objetos en movimiento, sin embargo, esta forma de medición identificará todos los candidatos posibles que deben ser posteriormente clasificados como carros, peatones u objetos que no son de nuestro interés.

Para realizar la detección de cambios en una escena lo más sencillo es analizar las diferencias entre dos imágenes sucesivas y maximizar el resultado mediante la umbralización. Este procedimiento es complicado debido a que se pueden agregar o eliminar cambios en la escena. Una vez que los candidatos son identificados, se procede su clasificación, para ello es necesario tener una descripción de cómo es un carro o un peatón [2]. Es importante escoger características apropiadas para realizar la comparación con los objetos candidatos, ya que de esto depende en gran medida una clasificación exitosa. Realizar la comparación es una tarea tediosa. Una solución poderosa y sencilla es la implementación de una red neuronal entrenada con las características escogidas [7]. Si los datos de entrenamiento y las características son apropiados, la clasificación será una tarea independiente de la acomodación y orientación de la cámara.

En este documento se presenta la detección y clasificación de vehículos y peatones en tres procesos principales: Primero se realiza la detección de cambios en la escena, usando un algoritmo de detección de movimiento invariante a las condiciones de iluminación para obtener los objetos candidatos; a continuación a estos candidatos se les extrae sus características mediante el concepto de descriptores de Fourier. Por último se realiza la clasificación de los objetos en peatones o vehículos, a través de una red neuronal *feed-forward* y se lleva un registro de aciertos para realizar el rastreo hasta que desaparezcan de la escena, permitiéndonos así el análisis de las escenas de tráfico.

En las siguientes secciones se describe el algoritmo de detección de movimiento invariante a las condiciones de iluminación, seguido del método de umbralización utilizado

para maximizar el resultado de la detección (sección II) y el concepto de descriptores de Fourier y la clasificación de los objetos (sección III). En la sección IV se describe el sistema completo y por último, en la sección V, se realiza la presentación de los resultados experimentales.

II. DETECCIÓN DE OBJETOS EN MOVIMIENTO INVARIANTE A LAS CONDICIONES DE ILUMINACIÓN

A. Adaptación de detección de objetos en movimiento

Para evitar los falsos positivos debido a las variaciones de iluminación, la detección de movimiento se adapta para trabajar imágenes independientes de la iluminación, y se hace necesario implementar un enfoque estadístico para decidir las regiones de cambio/no-cambio y evitar los falsos positivos o negativos producidos por el ruido [3].

El propósito de la detección de movimiento es generar una máscara binaria con intensidades de cambio/no cambio en la escena, las regiones estáticas pertenecen a intensidades cercanas al negro (0) y las regiones de cambio o movimiento al blanco (1). Comenzamos realizando la diferencia entre dos imágenes, mediante el cálculo de la suma de la diferencia absoluta (SAD) para detectar los cambios en la escena [4]. Para las variaciones debidas a los cambios de iluminación se filtran las imágenes antes de realizar el cálculo para obtener la reflectancia. Para maximizar los resultados de la SAD se utiliza un enfoque estadístico para umbralizar la imagen, modelando la distribución espacial de ruido y eliminar los cambios debido al mismo. En la Fig. 1 se presenta el principio del algoritmo.

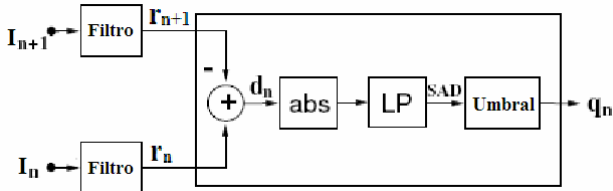


Fig. 1. Adaptación detección de movimiento. I_n y I_{n+1} son dos imágenes sucesivas, r_n y r_{n+1} son la reflectancia de las imágenes respectivamente, d_n es la diferencia, LP filtro pasa-bajos, umbral es umbralización estadística y q_n es la máscara binarizada.

B. Filtrado Homomórfico

En un modelo de imagen simple, la radiancia de la escena o intensidad l se caracteriza por dos componentes: 1) la cantidad de luz que incide en la escena i (irradiación o iluminación); y 2) la cantidad de luz que reflejan los objetos en la escena r (reflectancia), su producto define l . Para evitar la influencia de las variaciones de iluminación, cada imagen es prefiltrada usando filtrado homomórfico y se trabaja solamente con la reflectancia [5]. La componente de iluminación se caracteriza generalmente por variaciones espaciales lentas, mientras que la componente de reflectancia tiende a variar bruscamente, en especial en las uniones entre objetos distintos, por lo que se utiliza un filtro pasa-bajos para eliminar la iluminación [8],

como se muestra en la Fig. 2.

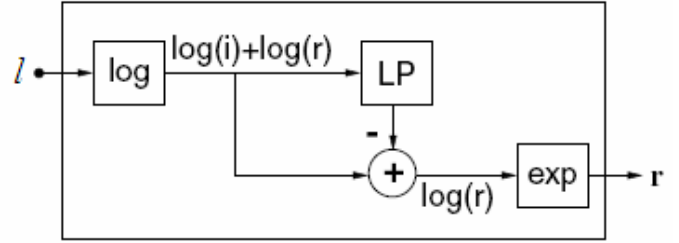


Fig. 2. Diagrama Filtro Homomórfico para señales multiplicadas.

C. Umbralización

La diferencia de imágenes generalmente se binariza por umbralización con algún valor predeterminado para obtener una clasificación de cambio/no-cambio. Sin embargo el valor del umbral es crítico. Un valor demasiado bajo inundará el mapa de diferencias con cambios falsos, mientras un valor alto suprimirá los cambios significativos. El valor apropiado del umbral debe ser dependiente de la escena y por tanto a sus cambios. Esto indica que en general el valor de umbral debe ser calculado dinámicamente basado en el contenido de la imagen. Aunque existen gran cantidad de métodos de umbralización, son muy pocos los disponibles en la literatura específica y apropiada a la detección de cambios. El método disponible, con mejores resultados y utilizado en esta aplicación, es modelo de Ruido de Poisson. Si asumimos que el ruido es blanco entonces su distribución espacial sobre la imagen será aleatoria. Para el análisis espacial de los datos hay muchas medidas de la aleatoriedad, a menudo basadas en la suposición de que las observaciones sigan una distribución de Poisson. Como la distribución de Poisson tiene su media igual a su varianza entonces la razón entre la varianza muestral y la media muestral, llamada *varianza relativa* $V_r = s^2/\bar{x}$, es una prueba natural para esta distribución. Este valor se calcula particionando la imagen en n ventanas, y por cada ventana se cuenta el número de observaciones (en nuestro caso el número de píxeles por encima del umbral en el mapa de diferencia), finalmente \bar{x} y s^2 se calculan como la media y la varianza de dichos conteos. Para nuestro propósito no es necesario detectar el ruido espacial aleatorio, pero en cambio si debemos evitarlo en la imagen umbralizada. Por lo tanto se escoge el umbral que llevan al máximo la variación relativa, de modo que se maximizan las regiones de cambio y se aminora la distribución de Poisson (el ruido) [3].

En la Fig. 3 se muestra el resultado de aplicar el proceso de umbralización a la imagen de diferencias obtenida con los procedimientos de las secciones A y B.



Fig. 3. Diferencia de imágenes (izquierda) y binarización (derecha)

III. EXTRACCIÓN DE CARACTERÍSTICAS Y CLASIFICACIÓN

A. Figures and Tables

A partir de la imagen binarizada, los objetos candidatos son marcados utilizando un análisis de conectividad. El paso siguiente, es extraer las características describiendo cada objeto individualmente. Como se utilizan descriptores de Fourier, es necesario estimar primero la frontera de los objetos utilizando detectores de bordes. Esto permite reducir el número de puntos en los bordes, lo que reduce los costos computacionales. Los descriptores de Fourier tienen como característica ser invariantes a transformaciones geométricas y tolerantes ante ruido. Una figura cerrada puede ser representada mediante series de Fourier con una parametrización adecuada. Para mostrarlo, considere un contorno $z(k)$ constituido por N puntos en el plano XY :

$$z(k) = [x_k, y_k], \quad k = 1, 2, \dots, N$$

donde cada punto se puede considerar como un número complejo:

$$z(k) = x_k + jy_k$$

El eje x es relacionado con la parte real y el eje y con la parte imaginaria de la sucesión de números complejos. La gran ventaja de esta consideración es que se pasa de un problema en dos dimensiones a un problema en una dimensión.

La transformada discreta de Fourier de $a(n)=z(k)$ es:

$$a(n) = \sum_{k=0}^{N-1} z(k) e^{-j2\pi kn/N} \quad 0 \leq n \leq N-1 \quad (1)$$

De esta forma los descriptores de Fourier pueden representar un contorno cerrado arbitrario [6]. Las propiedades más importantes de los descriptores de Fourier y que potencian su utilización son:

- 1) *Compactación*: Para la representación de la mayoría de las figuras son necesarios pocos descriptores sin perder fidelidad. Esto permite desechar los descriptores poco relevantes sin ocasionar pérdidas de información considerables.
- 2) *Insensibilidad a desplazamientos*: Cuando se mueve una figura solamente se modifica el primero de sus descriptores. Dicho descriptor es el que proporciona el "centro" de la figura y se puede asociar a la componente DC de la señal. Como sólo se busca comparar formas, este coeficiente debe ser eliminado.

- 3) *Insensibilidad a la rotación*: Como la rotación está decodificada en la fase de los coeficientes, se puede evitar los cambios de rotación usando solo la magnitud de cada $a(n)$.
- 4) *Insensibilidad de Escalado*: Los descriptores de dos contornos continuos g y f que tienen la misma forma pero diferente tamaño están relacionados únicamente por un factor de escala A , $g(l)=A*f(l)$. De esta forma, el factor de escala se puede eliminar normalizando adecuadamente los descriptores, lo cual se hace dividiendo por $a(1)$.
- 5) *Insensibilidad al punto de comienzo del contorno*: se puede conseguir restando la fase $e^{j\phi_1}$ del peso $a(1)$ con $a(n)$.

Al aplicar las anteriores propiedades a los descriptores calculados con (1), se obtiene un nuevo conjunto de coeficientes que son invariantes a la traslación, rotación, escalado y comienzo del contorno:

$$fd = \left[\frac{|a(n)|}{|a(1)|} e^{-j\phi_n} \right], \quad n = 2, 3, \dots, N-1$$

Otro método para representar contornos de mejor desempeño, hace uso de la distancia de los puntos del contorno $(x(k), y(k))$ al centro del objeto (x_c, y_c) :

$$r(k) = \left((x(k) - x_c)^2 + (y(k) - y_c)^2 \right)^{1/2}$$

La transformada de Fourier de $r(k)$ es análoga a la transformada de $z(k)$, lo que permite obtener los coeficientes $b(n)$:

$$b(n) = \frac{1}{N} \sum_{k=0}^{N-1} r(k) e^{-j2\pi kn/N} \quad 0 \leq n \leq N-1$$

Debido a que $r(k)$ tiene valores reales, solo se necesita la mitad de los descriptores de Fourier $b(n)$ para representar la forma correspondiente. Por efecto de la resta del centro, $r(k)$ sigue siendo insensible al desplazamiento. Las otras propiedades de insensibilidad para $b(n)$ se pueden obtener de igual manera. Esto lleva a un conjunto de descriptores fd_c (denota el centro):

$$fd_c = \left[\frac{|b(n)|}{|b(0)|} \right], \quad n = 1, 2, \dots, N/2$$

Finalmente, para reducir el número de coeficientes y debido a que no necesitamos todos los detalles del contorno, se descartan los componentes de alta frecuencia usando solo los 10 primeros coeficientes de fd y fd_c para clasificar los objetos. Esta cantidad de coeficientes es un resultado empírico. La experiencia ha demostrado que usando solo los 10 primeros coeficientes para reconstruir el contorno es suficiente para representar la forma global de los objetos (ver Fig. 4). En este trabajo se realizó la comparación de ambos métodos fd y fd_c con el fin de encontrar el de mejor desempeño. Estos resultados se encuentran en la sección V.

B. Clasificación

La clasificación es implementada por una red neuronal *feed-forward* de cuatro capas, una capa de entrada con una neurona por característica, dos capas ocultas con siete neuronas cada una y una capa de salida con una neurona por clase. Se usó una función de activación sigmoideal y un entrenamiento *back-propagation*. Para que la red obtenga buenos resultados es necesario una gran cantidad de ciclos de entrenamiento (aprox. 10000). En la capa de salida se utilizan dos umbrales para tomar la decisión, un umbral alto y uno bajo. Con esto se pretende disminuir las decisiones equivocadas.

C. Clasificación

Los objetos clasificados como humanos o vehículos son guardados y comparados con las nuevas entradas a la red hasta que desaparezcan de la escena.

IV. DIAGRAMA GENERAL DEL SISTEMA

El sistema propuesto para este problema consiste en tres procesos principales descritos en la Fig. 5. En el primer proceso, todos los objetos candidatos en movimiento son detectados usando el algoritmo de detección de objetos en movimiento invariante a las condiciones de iluminación descrito en la sección II. El proceso siguiente es extraer las características de los objetos candidatos mediante descriptores de Fourier y por ultimo una red neuronal clasifica cada objeto como humano, vehículo o cualquier forma.

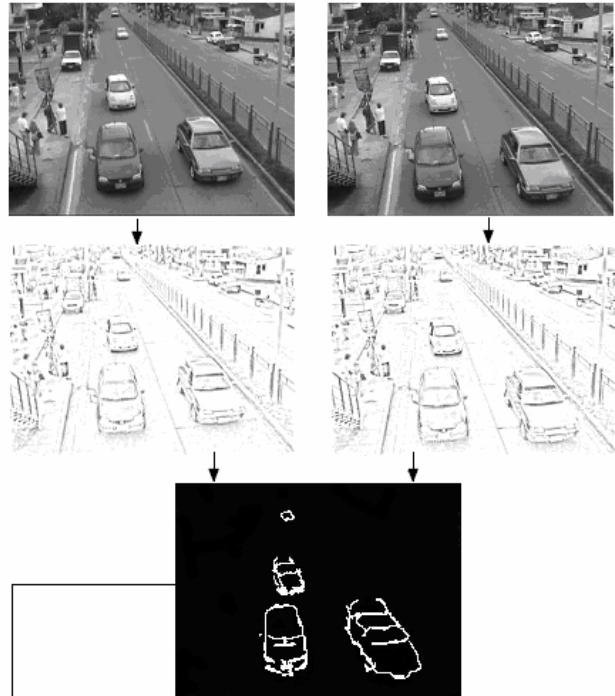
V. RESULTADOS

El algoritmo se probó en un PC 2.3 GHz Pentium 4 HT sistema operativo Windows XP, implementado en Matlab 7. Para imágenes de 320x240 píxeles el tiempo aproximado de procesamiento es de 10 imágenes/s usando del filtro homomórfico y 14 imágenes/s sin el uso del filtro. La diferencia computacional debido al uso de los fd y fd_c es insignificante. Sin embargo el rendimiento usando el filtrado homomórfico se puede mejorar optimizando su algoritmo.

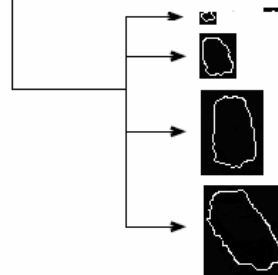
Se utilizaron cerca de 10 horas de grabación de humanos y vehículos en movimiento, se extrajeron cerca de 600 imágenes donde se detectaron y clasificaron cerca de 500 vehículos y 200 humanos. La tabla I muestra claramente los resultados de clasificación para los fd y fd_c . La primera fila corresponde al número de objetos humanos y vehículos, la segunda fila corresponde a número aciertos y desaciertos en al clasificación. En la última fila se encuentra el porcentaje de aciertos.

El principal problema de clasificación son los objetos obstruidos, porque no se detectan completamente sino por separado. Otro problema para la clasificación es que el cuerpo humano no es un cuerpo rígido como lo es un vehículo, por lo que su forma varía constantemente y se deben modelar como cuerpos deformables, lo cual dificulta el proceso de clasificación. Además, los humanos se mueven en ciertas ocasiones muy lentamente causando que la binarización no

1. detección de movimiento invariante a la iluminación



2. Extracción de características: descriptores de Fourier



3. Clasificación: Red Neuronal

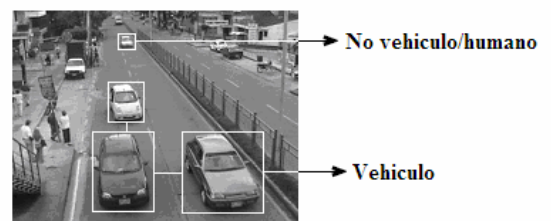


Fig. 5. Diagrama general de sistema

sea suficientemente exacta. Como se esperaba en los resultados, la clasificación es independiente del tamaño del objeto y su orientación.

TABLE I

Método	fd		fd _c	
	huma	Vehi	huma	vehi
no. Objetos	600	400	600	400
no. Incorrectos	36	16	72	12
no. Correctos	564	384	528	388
% Correctos	94	96	88	97

La Fig. 6 muestra algunos ejemplos de clasificación incorrecta. Los vehículos que obstruyen uno al otro se identifican erróneamente como un sólo vehículo (Fig. 6a). Los humanos parcialmente obstruidos no son clasificados en algunas ocasiones (Fig. 6b). Los vehículos que están entrando en algunas escenas pueden ser clasificados como humanos (Fig. 6c). Objetos con formas similares a la humana, como banderas, ondeando (Fig. 6d), pueden ser clasificadas erróneamente. Sin embargo, en términos generales, el sistema es robusto, si los objetos son enteramente visibles.



Fig. 6. Ejemplos de clasificación errónea.

VI. CONCLUSIONES

En resumen, se construyó un sistema para detectar y clasificar el tráfico de peatones (humanos) y vehículos. En el primer paso, un algoritmo de detección de objetos en movimiento invariante a las condiciones de iluminación nos proporciona los candidatos a clasificar. Estas formas binarias se utilizan para el cálculo de descriptores de Fourier. Finalmente, una red neuronal *feed-forward* se utiliza para clasificar los objetos en cuestión. Se probaron dos métodos para el cálculo de descriptor de Fourier (FD): El tradicional, donde números complejos se construyen para todos los píxeles objetivos de borde y se transforman para obtener el FDs. Y un método alternativo, basado la transformada de Fourier del conjunto de las distancias del centro a los puntos de borde. En este sistema ambos métodos tienen buen desempeño en la clasificación de objetos en más del 90 % en todos los casos. Sin embargo, el método tradicional trabaja igualmente bien para ambos humano y el reconocimiento de vehículo, que no es el caso para el segundo método. Hay dos elementos que hace la clasificación sea robusta: en primer lugar, los descriptores de Fourier solo están basados en la forma del objeto, que los hace una buena elección para varias aplicaciones de reconocimiento de patrones. En segundo lugar, un doble de umbral en la decisión de la red neuronal lo hace más seguro. Además, la parte de detección de cambio proporciona candidatos objetivos bien formados. No obstante, esto se puede mejorar en términos de la extracción de sombra, porque sombras especialmente largas pueden confundir el sistema.

REFERENCES

- [1] K. N. Plataniotis and C. S. Regazzoni. Visual-Centric Surveillance Networks and Services. *Signal Processing: Image: Surveillance Networks*, 22, 2005.
- [2] D. Koller, K. Daniilidis, and H. H. Nagel. Model-based object tracking in monocular image sequences of road traffic scenes. *International Journal of Computer Vision*, 10(3):257-281, 1993.
- [3] J.P. L. Rosin, "Thresholding for change detection," *Comput. Vis. Image Understanding*, vol. 86, no. 2, pp. 79-95, 2002.
- [4] G. L. Foresti, C. Micheloni, L. Snidaro, P. Remagnino, and T. Ellis, "Active Video-Based Surveillance" *Signal Processing: Image: Surveillance Networks*, 22, 2005.
- [5] H. Y. Ryu; K. Lee; B. D. Kwon; "Change detection for urban analysis with high-resolution imagery: homomorphic filtering and morphological operation approach". 2004 IEEE International vol. 4, 20-24.
- [6] E. G. García, V. F. Batlle, A. A. Oliver y L. S. "Descriptores de Fourier para identificación y posicionamiento de objetos en entornos 3D" XXV Jornadas de Automática Ciudad Real, España, 8-10 sep. 2004.
- [7] F. Tadeo, S. González, M. T. Álvarez y J. L. Arribas; "Clasificación de microorganismo Mediante Procesado de Imagen" XI Jornadas de Investigación México DF, México 25-29 Abr. 2004.
- [8] J.J. G. Proakis, and D. G. Manolakis, *Digital Signal Processing: principles, algorithms and applications*, 3rd ed., Prentice-Hall, 1998.



Andrés Felipe Granados Zárate nacido el 27 de Octubre de 1980 en la ciudad de Bogota, realizo y finalizó sus estudios de secundaria en el Colegio San Francisco Solano en la ciudad de Armenia, en el año 1998, realiza sus estudios universitarios actualmente en la Universidad del Quindío,. Auxiliar docente en el área de control durante el segundo periodo académico del año 2005 Asistió al "Segundo Congreso de Electrónica y Telecomunicaciones" realizado en la Universidad del Quindío en octubre de 2002, aspirante al título de Ingeniero electrónico.