

# Identificación de edificios y monumentos a partir de fotografías tomadas con dispositivos móviles

Esteban C. Fornal, Christian N. Pfarher, Mauro J. Torrez  
Trabajo práctico final de “Procesamiento Digital de Imágenes”, II-FICH-UNL.

**Resumen**—Se presenta un método para la identificación de edificios y monumentos, a partir de fotografías tomadas con la cámara de un dispositivo móvil. Se extrae un vector de características globales y locales de cada imagen. Se identifica una imagen de entrada minimizando el error cuadrático medio entre sus vectores de características y los de una base de datos de imágenes conocidas. Se presentan dos técnicas para la extracción de características, una basada en la transformada de Hough y otra que utiliza estadísticas de los histogramas. Se evalúa el desempeño de ambas técnicas por separado y en conjunto, para una base de datos de prueba.

**Palabras clave**—Identificación de edificios, *building recognition*, histograma, transformada de Hough, clasificación.

## I. INTRODUCCIÓN

EN la actualidad, la gran difusión de los dispositivos móviles nos permite llevar la información importante siempre con nosotros. Además, es muy común utilizar nuestros celulares, PDAs, GPS para obtener información desde cualquier lugar donde estemos, gracias a la proliferación de redes de datos inalámbricas. Dado que una gran parte de estos dispositivos posee una cámara digital, surge la idea de utilizarla junto con las conexiones de datos para obtener información acerca del lugar donde nos encontramos. Se presenta aquí un método que combina técnicas de Procesamiento Digital de Imágenes orientadas en este sentido, que permite la identificación de edificios, monumentos y otros puntos de interés, a partir de una fotografía capturada con un dispositivo móvil.

## II. MÉTODO PROPUESTO

En el método propuesto se pone énfasis en la etapa de extracción de características de la imagen, considerando un método trivial de clasificación.

### Extracción de características

La extracción de características se realiza mediante dos técnicas diferentes:

- 1) transformada de Hough, y
- 2) estadísticas del histograma.

1) *Extracción de características mediante Transformada de Hough*: La transformada de Hough nos permite visualizar, a partir de una imagen de bordes, los parámetros de aquellas rectas<sup>1</sup> que son principales en la imagen [1].

Para la extracción de características con esta técnica se siguen los siguientes pasos:

- 1) A partir de la imagen original, se obtiene su versión en escala de grises promediando los tres canales RGB, y se la escala a un tamaño normalizado.

<sup>1</sup>El procedimiento es general, sirve para cualquier geometría que se pueda expresar en términos de sus parámetros. En este trabajo, se utiliza el espacio de los parámetros de las rectas, el más sencillo.

|    |    |    |    |   |   |
|----|----|----|----|---|---|
| -1 | -2 | -1 | -1 | 0 | 1 |
| 0  | 0  | 0  | -2 | 0 | 2 |
| 1  | 2  | 1  | -1 | 0 | 1 |

Fig. 1. Máscaras de filtrado (operadores gradiente) de Sobel.

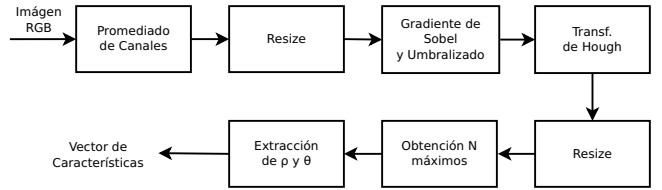


Fig. 2. Proceso de extracción de características por Transformada de Hough

- 2) Se obtiene una imagen de sólo bordes, aproximando la magnitud del gradiente según

$$\nabla f \approx |G_x| + |G_y|, \quad (1)$$

donde  $G_x$ ,  $G_y$  son el resultado de aplicar los operadores gradiente de Sobel [1] (fig. 1) a la imagen. Finalmente se umbraliza esta imagen de bordes utilizando un parámetro  $U$ :

$$f(I) = \begin{cases} 0, & I \leq U \\ 255, & I > U \end{cases} \quad (2)$$

donde  $I = I(x,y)$  es el valor de la intensidad en el punto  $(x,y)$ .

- 3) Con la imagen de bordes umbralizada se calcula la transformada de Hough para rectas.
- 4) Se aplica un escalado (submuestreo) a la transformada obtenida, llevándola a un tamaño pequeño para obtener mayor tolerancia tanto en el parámetro angular  $\theta$  como en el de distancia  $\rho$ .
- 5) Se toman los  $N$  máximos de esta transformada y se guardan en el vector de características sus coordenadas  $(\rho, \theta)$ , mapeadas al rango  $[-1, 1]$ , obteniendo así un vector de  $2N$  valores.

El proceso completo puede verse en la Fig. 2.

2) *Extracción de características por estadísticas del histograma*: A partir de la imagen original, se convierte a escala de grises, se normaliza su tamaño y se toman 2 “perfiles de intensidad”: uno horizontal, calculado promediando cada columna de la imagen, y otro vertical obtenido al promediar cada fila. Estos “perfiles” se relacionan con la distribución de la energía promedio en las filas y columnas de la imagen. Luego se calculan tres histogramas, uno para cada perfil y otro para la imagen entera. De estos 3 vectores, se calculan y se guardan en el vector de características la media aritmética

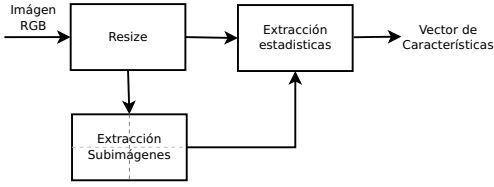


Fig. 3. Proceso de extracción de características utilizando estadísticas del histograma.

$m$ , la mediana  $M$  (posición del percentil 50), y la desviación absoluta  $D_{\text{abs}}$  respecto de la mediana:

$$D_{\text{abs}} = \sum_i |x_i - M| \quad (3)$$

Así, se ha obtenido un vector de 9 valores que caracterizan a la imagen entera.

Con la idea de incorporar características locales, se subdivide la imagen en cuatro cuadrantes y se obtienen, para cada uno, las mismas medidas que se calcularon para la imagen entera. Como resultado, se obtiene un vector de 45 características asociado a cada imagen.

#### Clasificación

El entrenamiento de la base de datos se realiza obteniendo las características para cada imagen, junto a su etiqueta. Se extraen las características de todas las imágenes con la misma etiqueta, y a partir de ellas se genera un “prototipo”, que surge de promediar estas características.

La clasificación de las imágenes consiste en obtener la etiqueta del prototipo  $p_{\text{ganador}}$  cuyas características minimicen el error cuadrático medio con las de la imagen a identificar:

$$p_{\text{ganador}} = \arg \min_i \left\{ \frac{1}{\sum N_j} \sum_{j=1}^K \sum_{n=1}^{N_j} (t_j[n] - p_{ij}[n])^2 \right\} \quad (4)$$

donde  $t_j$  es el  $j$ -ésimo vector de características extraído de la imagen, con longitud  $N_j$ , y  $p_{ij}$  es el  $j$ -ésimo vector de características del  $i$ -ésimo prototipo generado en el entrenamiento, también de longitud  $N_j$ .

### III. EXPERIMENTOS Y RESULTADOS

Como se pretende evaluar el método en situaciones reales, fue necesario generar una base de datos propia de imágenes.

#### Base de datos

La base de datos se armó con imágenes de prueba tomadas con teléfonos celulares corrientes, a una resolución VGA estándar de 640 píxeles de ancho por 480 de alto, en diferentes condiciones de iluminación: día a pleno sol, día nublado, noche, e interior, a monumentos/estatuas y edificios en diferentes lugares de la ciudad de Santa Fe (Argentina). Se obtuvieron 195 imágenes, 13 por cada uno de las 15 etiquetas diferentes. Se muestran en la figura 4 dos ejemplos de las imágenes utilizadas.

#### Experimentación

Se generaron 3 particiones de la base de datos, tomando 5 etiquetas al azar (no repetidas) en cada caso. Para la prueba del método en cada caso se utilizó validación cruzada, entrenando la base de datos con 10 imágenes por etiqueta, y dejando las 3 imágenes restantes para la prueba.

Se clasificó utilizando extracción de características mediante transformada de Hough y mediante histogramas por



Fig. 4. Ejemplo de dos imágenes utilizadas para probar el método.

separado. Para la técnica de Hough se probaron diferentes valores de  $U$  (el umbral aplicado a la imagen de bordes) y  $N$  (el número de máximos considerados en el espacio de la transformada de Hough); mientras que, para el método de histograma se consideró el histograma y los perfiles del canal de intensidad en el espacio de colores HSI.

Luego se evaluó el rendimiento del método utilizando todas las características en forma conjunta, para los parámetros de Hough  $U$  y  $N$  óptimos encontrados.

Se realizó además un caso de prueba con las 15 imágenes de la base de datos, con el objetivo de tener una estimación de cómo responde el método para un mayor número de imágenes.

Se considera la tasa de error (%) del método según:

$$E_{\%} = 100 \cdot \frac{\text{número de errores}}{\text{número de pruebas}}, \quad (5)$$

considerando como error a cada prueba en que la imagen es mal etiquetada.

Cabe destacar que la clasificación se realiza en términos de clases, que representan a un edificio/monumento en cualquier condición de iluminación (puede haber varias etiquetas/prototipos que corresponden a la misma clase), luego no es un error que la imagen de prueba se etiquete en forma incorrecta si la clase es la acertada. Por ejemplo, no es error si la imagen de un monumento tomada de día es etiquetada como la del mismo monumento, pero tomada de noche. Esta consideración se hace debido a que la detección de bordes elimina la información de iluminación de la imagen, y está claro que no influye en la técnica por histograma, ya que éste varía significativamente entre las versiones de día y noche; en particular, la media y mediana tendrán valores bastante menores en la imagen nocturna que en aquella tomada de día. Esto es además consistente con el objetivo del método, que es la correcta identificación de la clase (clasificación), independientemente de las condiciones (etiqueta) en que se toma la imagen.

#### Resultados

Los resultados de las pruebas con la técnica de Hough se muestran en la figura 5. Se ha tomado un rango de valores representativos de  $U$  y  $N$  basados en pruebas previas, donde hemos identificado regiones de mínimo error para  $U \in (60, 120)$  y  $N \in (20, 60)$ . En la misma figura, se puede apreciar que la tasa de error mínima se obtuvo para un umbral  $U = 100$  y unos  $N = 30$  máximos (en promedio).

En la tabla I se presenta el resultado promediado de los 3 casos de prueba de 5 etiquetas y para el caso de 15 etiquetas.

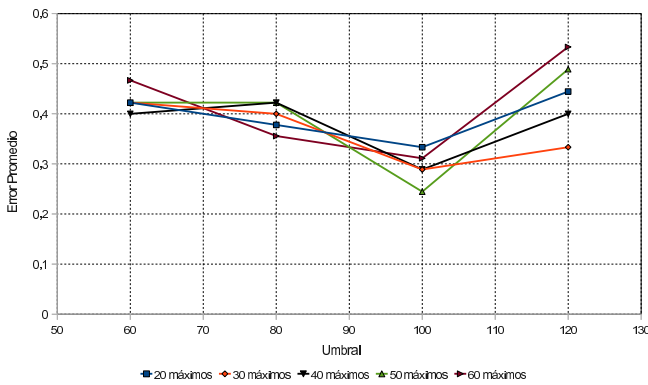


Fig. 5. Error de clasificación promedio en función de distintos  $U$  y  $N$  para la técnica de Hough.

TABLA I  
TASAS DE ERROR PARA LAS TÉCNICAS DE EXTRACCIÓN DE CARACTERÍSTICAS.

| Técnica           | 5 etiquetas | 15 etiquetas |
|-------------------|-------------|--------------|
| Histogramas       | 0%          | 0%           |
| Hough             | 35.5%       | 60.43%       |
| Histogramas+Hough | 2.22%       | 4.17%        |

Se puede observar que para la técnica de histogramas, la tasa de error fue cero en ambas pruebas. En tanto que para la de Hough, se obtiene menor tasa de error en los casos de 5 etiquetas que con el de 15, de 35.5% y 60.43% respectivamente. En la última fila, se muestran los resultados de considerar todas las características, asignándole un peso equivalente a cada una.

#### Discusión

A pesar de la muy buena performance de la técnica de histograma, debemos tener presente que las imágenes de entrenamiento/prueba han sido obtenidas en todos los casos a partir de una secuencia tomada en el lapso de unos minutos, con el mismo dispositivo, lo cual implica que las condiciones de iluminación en cada etiqueta –con la que luego se generan los prototipos– son prácticamente iguales.

Los histogramas miden la distribución estadística de los diferentes niveles de intensidad presentes en la imagen, luego están estrechamente relacionados con las condiciones de iluminación de la escena. Por este motivo, tenemos que los histogramas entre las imágenes de entrenamiento/prueba son similares, lo cual explica el buen funcionamiento de este método.

Se ha probado la técnica de histogramas sobre el canal I del espacio de color HSI; por lo antes expuesto, debemos tener en cuenta consideraciones similares para el resto de los canales de la imagen.

En general, se deberá poner énfasis en definir técnicas de “normalización” orientadas a mejorar los resultados en condiciones más realistas. Se deberá considerar la utilización de métodos de histograma mejorados, por ejemplo el *histograma de combinación espacial de DCT ponderada* [2], *vectores de coherencia de color* [3], o el procesamiento de histogramas borrosos presentado en [4]. Algunos tipos de preprocesamiento sencillos, como la utilización de histogramas en los canales H ó V, o el filtrado homomórfico, nos permitirán lograr una mayor independencia respecto de las condiciones de iluminación.

Respecto de la extracción de características mediante la transformada de Hough, vemos que el rendimiento no es tan bueno como en la técnica de histogramas, sin embargo en varios casos ha sido capaz de identificar correctamente edificios a pesar de las diferentes condiciones de iluminación (día y noche), lo cual es un resultado alentador.

Se deberá considerar también la posibilidad de utilizar extracción de características locales mediante la técnica de Transformada de Hough.

No se puede evitar la mención al costo computacional del proceso, que aunque no es tan elevado como para considerar impráctico el método, sí será una limitante al considerar implementaciones en tiempo real o de alta velocidad de respuesta: en una computadora promedio el cálculo se realiza en 1–2 segundos, luego es esperable que este tiempo se triplique en un dispositivo móvil.

Se deberá considerar la utilización de un mejor detector de bordes, como el propuesto por Canny [5], así como también técnicas de pre-procesamiento de la imagen, como puede ser el filtrado homomórfico, en pos de mejorar los resultados obtenidos en nuestras pruebas.

En lo que respecta a la escalabilidad de la base de datos, los resultados no son particularmente alentadores, lo que nos podría estar indicando que las características extraídas no son suficientemente *únicas* para las imágenes, obligándonos a considerar otra vez la utilización de métodos más refinados de histograma y transformada de Hough.

#### IV. CONCLUSIONES

Se ha presentado una técnica para la identificación de edificios, monumentos, esculturas con extracción de características mediante medidas de histograma y transformada de Hough.

El rendimiento ha sido satisfactorio considerando las restricciones a las que se han sometido las pruebas.

Se debe optimizar la implementación para portarlo a dispositivos móviles con capacidad de procesamiento limitada.

Se hace necesario un preprocesamiento de la imagen, así como la incorporación de métodos más refinados de extracción de características, para mejorar los resultados obtenidos y así poder usar el método con una base de datos de mayor magnitud.

#### V. TRABAJOS FUTUROS

A partir del diseño aquí presentado, han surgido líneas para continuar investigando y lograr un método más robusto:

- Aplicación de filtrado homomórfico y otros tipos de pre-procesamiento en las imágenes.
- Aplicación de técnicas de *warping* y otras transformaciones en busca de lograr invarianza respecto de rotación y escalado de la imagen.
- Desarrollo de técnicas de extracción de características más refinadas.
- Desarrollo de una implementación óptima para dispositivos móviles con poder de procesamiento limitado.

#### REFERENCIAS

- [1] R. C. Gonzalez y R. E. Woods, *Digital Image Processing*. Prentice-Hall, 2001.
- [2] L. S. Yonghua Xie y H. Burkhardt, “Color image retrieval based on weighted dct spatial combination histogram,” *International Journal of Information Technology*, vol. 14, no. 2, pp. 12–23.

- [3] G. Pass y R. Zabih, "Histogram refinement for content-based image retrieval," en *IEEE Workshop on Applications of Computer Vision*, pp. 96–102, 1996.
- [4] K. Konstantinidis, A. Gasteratos, y I. Andreadis, "Image retrieval based on fuzzy color histogram processing," *Optics Communications*, vol. 248, no. 4-6, pp. 375 – 386, 2005. [Online]. Disponible: <http://www.sciencedirect.com/science/article/B6TVF-4F2MDYT-3/2/074827181182cd309f59f57ee820d26f>
- [5] J. F. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 679–698, 1986.