

Análisis de imágenes mediante el método de los momentos usando funciones de base continuas a intervalos (PCBF)

Sergio Domínguez

Centro de Automática y Robótica UPM-CSIC, Sede Castellana. c/ José Gutiérrez Abascal, 2, 28006 Madrid, España.

Resumen

Los invariantes generados a partir de momentos extraídos de una imagen aparecen recurrentemente en la bibliografía como uno de los métodos más potentes para la descripción de imágenes, y más concretamente de formas. En este artículo se propone el uso de funciones de base continuas a intervalos (PCBF) como alternativa a las bases que se vienen utilizando tradicionalmente en la aplicación de este método, todas ellas continuas como puedan ser las muy conocidas de Zernike, Legendre o Tchebichev. El uso de funciones discontinuas se justifica en la naturaleza discontinua de los objetos analizados, en este caso las imágenes: es de sobra conocido que los contornos de los objetos visibles en ellas se caracterizan como discontinuidades en la serie de valores de luminancia cuando nos desplazamos de un lado a otro de dichos contornos. El análisis de tales señales con funciones continuas produce resultados no deseados, como el fenómeno de Gibbs, que pueden ser evitados mediante el uso de funciones como las propuestas, generando mejores aproximaciones a la forma analizada. Adicionalmente, las bases propuestas permiten fácilmente, como se demuestra, la generación de invariantes a rotación, característica altamente deseable para un descriptor de forma, puesto que a priori no se conoce con qué orientación aparecerá esta en la imagen objeto del análisis. La invarianza a traslación y escala se consigue mediante un sencillo proceso de normalización. Se presentan los test que confirman esta hipótesis, comenzando por un análisis del comportamiento de los invariantes ante el ruido en la imagen que permitirá determinar en qué número deben ser extraídos. A continuación, y una vez definida esta longitud de descripción, se realizan sendos experimentos para determinar el comportamiento de los invariantes propuestos en una tarea de recuperación de imágenes, tanto libres de ruido como corrompidas con distintos grados de ruido gaussiano. Los resultados avalan la hipótesis de idoneidad para la tarea, demostrando que se pueden alcanzar resultados similares a los de la base de referencia, Zernike, utilizando descripciones hasta un 40 % más cortas. *Copyright © 2015 CEA. Publicado por Elsevier España, S.L. Todos los derechos reservados.*

Palabras Clave:

Método de los momentos, descriptores invariantes, bases ortonormales, recuperación basada en contenido, análisis de imágenes

1. Introducción

El análisis de imágenes mediante invariantes extraídos de momentos de distinta índole se ha constituido en un tema de investigación activo desde su propuesta inicial por Hu (Hu, 1962). Desde entonces las líneas de trabajo se han orientado tanto hacia una mayor profundización en las bases teóricas de su definición, como hacia nuevas y mejores formas de construir tales descriptores (Reiss, 1991) (Flusser et al., 2009). Uno de los temas que más atención ha recibido dentro de esta investigación ha sido la definición de momentos utilizando bases ortonormales, dada la conexión que existe entre tales bases y las descomposiciones de Fourier propuestas en Análisis Funcional, y más

concretamente en el álgebra de Hilbert (Rudin, 1991); conforme a estas disciplinas, el uso de bases ortonormales, además de las ventajas del uso de invariantes en la descripción de una imagen, ofrece adicionalmente la posibilidad de realizar descripciones y reconstrucciones óptimas, en términos de compacidad y unicidad de la representación, lo que conduce igualmente a un óptimo computacional (Rudin, 1991) (Teague, 1980). De esta forma, se ha dedicado mucho esfuerzo al estudio de las bases ortonormales conocidas, como las de Zernike, Legendre o Tchebichev, y a su comportamiento como generadores de descriptores, tanto desde el punto de vista comparativo (Teh and Chin, 1988), como desde la exploración de sus capacidades específicas (Khotanzad and Hong, 1990). Dado el buen comportamiento observado con esta estrategia de descripción en formas 2D, se han dirigido esfuerzos a su extensión como herramientas de descripción de objetos 3D (Xu and Li, 2008). Recientemente

Correo electrónico: sergio.dominguez@car.upm-csic.es (Sergio Domínguez)

URL: www.car.upm-csic.es (Sergio Domínguez)

se han realizado esfuerzos para unificar la capacidad de los invariantes generados a partir de momentos con las posibilidades de análisis que ofrecen las wavelets (Chen and Xie, 2011).

De otra parte, también es conocido que la descomposición de Fourier de una función utilizando bases continuas produce defectos no deseados en el análisis de funciones con discontinuidades, siendo la manifestación más palpable de este hecho el conocido fenómeno de Gibbs (Hewitt and Hewitt, 1979). Más concretamente, este comportamiento se puede observar en el análisis de imágenes, donde es habitual encontrar grandes discontinuidades que, de hecho, dan lugar al concepto de contorno tan explotado en Visión por Computador. Se tiene constancia de la aparición de este tipo de fenómenos no deseados en la reconstrucción de imágenes tras un proceso de filtrado en el dominio de Fourier (Bovik, 2009). Estos inconvenientes han generado un creciente interés en el uso de funciones de base discontinuas para el análisis de este tipo de señales, como puede ser la aplicación de la transformada de Hadamard-Walsh en análisis de imágenes (Sasikala and Neevelani, 2010), su fusión con el análisis de Fourier (Hamood and Boussakta, 2011), o el uso de wavelets como las de Haar, presentes en multitud de propuestas en la bibliografía, como el algoritmo de Viola-Jones, el análisis de fallos en maquinaria industrial (Lee and Tarn, 1999), o el análisis de imágenes en dominios específicos (Krommweh, 2009).

De este modo, el trabajo que se presenta tiene por objeto explorar la potencialidad de funciones discontinuas, o más concretamente las conocidas como funciones de base continuas a intervalos (*Piecewise Continuous Basis Functions*, PCBF) (Rao, 1983), como base para la definición de familias de momentos a partir de las que luego extraer invariantes para la descripción de imágenes; incluso, en este caso, se podría introducir una restricción adicional, puesto que las funciones de base que se proponen no sólo son continuas a intervalos, sino también constantes en esos intervalos.

La metodología que se utiliza se basa en el análisis de dos tipos de PCBF, una base ortonormal y otra que no lo es, de forma que no sólo se puedan extraer conclusiones al respecto del uso de bases discontinuas, sino adicionalmente la ventaja que puede suponer el que tales bases sean además ortonormales. De esta forma, el trabajo que ahora se presenta, extiende y explica los resultados preliminares reflejados en (Domínguez, 2013) en varios sentidos: en primer lugar, se presenta una metodología generalizada para la conversión de bases ortonormales en coordenadas cartesianas a bases en coordenadas polares, de gran interés en análisis de imágenes por la facilidad que ofrecen para la generación de descriptores invariantes a rotación; en segundo lugar se completa el experimento presentado en el trabajo citado, en el sentido de que ahora se diferencian claramente los efectos de dos características distintas de las funciones presentadas, al detallar por un lado los beneficios de usar bases continuas a intervalos, y por otro al añadir a esta característica la ortonormalidad de las funciones de la base. Se presentan diversos experimentos que persiguen determinar el correcto dimensionamiento de cada una de las familias de invariantes analizando su comportamiento frente al ruido y cuando se utilizan como descriptores de una imagen en una aplicación de recuperación

de imágenes por contenido, tanto en un caso ideal, donde las imágenes objeto de análisis se presentan sin ruido, como cuando, en un escenario más realista, tales imágenes se presentan con distintos grados de contaminación con ruido gaussiano.

Finalmente, cabe reseñar que como elemento de control se han repetido todos estos experimentos utilizando momentos de Zernike, de los que se dispone de información en la bibliografía que corrobora que se trata de la familia de invariantes con mejor comportamiento en la descripción de imágenes (Teh and Chin, 1988) (Chen and Xie, 2011).

2. Definición de las bases

El punto de partida para la propuesta de las bases de funciones con las que se va a trabajar está en la propia definición de los momentos; dada una imagen descrita en coordenadas cartesianas $I(x, y)$, su momento M_{nm} se calcula como (Flusser et al., 2009):

$$M_{nm} = \int_{x_0}^{x_f} \int_{y_0}^{y_f} I(x, y) p_{nm}^*(x, y) dx dy \quad (1)$$

donde $p_{nm}(x, y)$ es una función que pertenece a la base de los momentos, y que tiene orden n en la coordenada x y orden m para y , $*$ es el conjugado complejo, y $[x_0, x_f]$ e $[y_0, y_f]$ los intervalos en ambas variables que definen la región en la que tanto la imagen como $p_{nm}(x, y)$ están definidas.

De acuerdo con el álgebra de Hilbert, la ecuación 1 describe igualmente el producto escalar entre ambas funciones implicadas, $\langle p_{nm}(x, y), I(x, y) \rangle$. Gracias a esta correspondencia, es posible introducir en el análisis mediante momentos el concepto de ortonormalidad entre cada par de funciones integrantes de la base que se elija. Este concepto es muy importante, puesto que es conocido que la proyección de un punto de un espacio vectorial sobre una base ortonormal es óptima en el sentido de que cada componente proyectada contiene información única y no repetida en ninguna otra proyección. La consecuencia de esta propiedad es que el análisis por el método de los momentos utilizando bases ortonormales genera proyecciones óptimas (Rudin, 1991) (Teague, 1980) puesto que a igual longitud de representación se da cabida a una mayor cantidad de información referente al objeto representado, en nuestro caso la imagen.

El siguiente paso a tener en cuenta es un cambio en la definición de la propia imagen que se va a analizar. Tradicionalmente, y debido a la morfología de los sensores CCD utilizados en la captura de imágenes, éstas se representan como una matriz, direccionándose por tanto cada uno de sus píxeles mediante un par de números que representan las coordenadas cartesianas (i.e. fila y columna) en las que dicho píxel se ubica en la imagen. No obstante, a la hora de realizar un análisis de la imagen mediante el método de los momentos, se ha probado muy útil modificar esta representación de forma que la imagen pasa a manejarse en coordenadas polares, y dentro del círculo unidad. Por tanto, ahora la imagen pasa a denominarse $I(\rho, \theta)$ y se define como:

$$I(\rho, \theta) : \Omega \mapsto [0, 1], \quad \Omega = \{(\rho, \theta) \in \mathbb{R}^2 : 0 \leq \rho \leq 1, 0 \leq \theta < 2\pi\}$$

Como es lógico, esta definición de la imagen fuerza a modificar igualmente la definición del producto escalar y, como consecuencia, de los propios momentos, dada en la ecuación 1 para coordenadas cartesianas. De esta forma, ahora el cálculo de cada uno de los momentos se realiza mediante la expresión:

$$M_{nm} = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 I(\rho, \theta) \varphi_{nm}^*(\rho, \theta) \rho \, d\rho \, d\theta \quad (2)$$

donde $\varphi_{nm}(\rho, \theta)$ representa la función de orden n en ρ y m en θ de la base de momentos, y $*$ representa el conjugado complejo.

Cuando esta representación de la imagen en coordenadas polares se conjuga con una definición apropiada de las funciones de base para el análisis, se obtienen resultados muy deseables. Una posible forma de construir estas funciones de base es desacoplando la dependencia de ambas variables como un producto de funciones, una que exprese la variación sobre ρ y otra sobre θ ; de esta forma, se podría escribir:

$$\varphi_{nm}(\rho, \theta) = \Pi_n(\rho) \Phi_m(\theta) \quad (3)$$

Según esta descomposición, cuando la componente angular se define como la correspondiente función de la base de Fourier, es decir:

$$\Phi_m(\theta) = e^{im\theta} \Rightarrow \varphi_{nm}(\rho, \theta) = \Pi_n(\rho) e^{im\theta} \quad (4)$$

el resultado es que el módulo del momento así extraído es invariante a rotación, como puede comprobarse fácilmente. Supóngase una imagen, $I(\rho, \theta)$ y una copia de la misma girada un ángulo α , $I'(\rho, \theta) = I(\rho, \theta + \alpha)$. Es fácil ver que la relación entre los momentos de ambas imágenes es tal que $M'_{nm} = M_{nm} e^{-in\alpha}$, por lo que $\|M'_{nm}\| = \|M_{nm}\|$ (Flusser, 2000).

Fijada la invarianza a rotación, queda por garantizar la invarianza a traslación y escalado del descriptor utilizado. En este trabajo, tales propiedades se consiguen mediante un proceso previo de normalización de las imágenes (Khotanzad and Hong, 1990).

Adicionalmente, las funciones de base definidas según la expresión 4 gozan de otra propiedad interesante, y es que la condición de ortonormalidad entre ellas se cumple cuando simultáneamente las componentes radial y angular la cumplen. Sean $\varphi_{nm}(\rho, \theta)$ y $\varphi_{kl}(\rho, \theta)$ dos funciones de la base de momentos. Su producto escalar es, según la expresión 2:

$$\langle \varphi_{nm}(\rho, \theta), \varphi_{kl}(\rho, \theta) \rangle = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 \varphi_{nm}(\rho, \theta) \varphi_{kl}^*(\rho, \theta) \rho \, d\rho \, d\theta \quad (5)$$

pero dada la forma propuesta en la ecuación 4 se puede escribir como:

$$\langle \varphi_{nm}(\rho, \theta), \varphi_{kl}(\rho, \theta) \rangle = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 \Pi_n(\rho) e^{im\theta} \Pi_k(\rho) e^{-il\theta} \rho \, d\rho \, d\theta \quad (6)$$

En esta expresión, dado que ninguna de las funciones implicadas en el integrando depende de ambas variables, se puede secuenciar el proceso de integración:

$$\langle \varphi_{nm}(\rho, \theta), \varphi_{kl}(\rho, \theta) \rangle = \frac{1}{\pi} \int_0^1 \Pi_n(\rho) \Pi_k(\rho) \left[\int_0^{2\pi} e^{im\theta} e^{-il\theta} d\theta \right] \rho \, d\rho \quad (7)$$

donde el término entre paréntesis sólo depende de θ , y dado que lo que se integra es el producto de dos funciones de la base de Fourier, se sabe que, por su ortonormalidad, el resultado es:

$$\int_0^{2\pi} e^{im\theta} e^{-il\theta} d\theta = 2\pi \delta_{ml} \quad (8)$$

donde

$$\delta_{ml} = \begin{cases} 1 & m = l \\ 0 & m \neq l \end{cases} \quad (9)$$

lo que deja:

$$\langle \varphi_{nm}(\rho, \theta), \varphi_{kl}(\rho, \theta) \rangle = 2\delta_{ml} \int_0^1 \Pi_n(\rho) \Pi_k(\rho) \rho \, d\rho \quad (10)$$

Lo que se traduce en que $\varphi_{nm}(\rho, \theta)$ y $\varphi_{kl}(\rho, \theta)$ son ortonormales si a su vez lo son $\Pi_n(\rho)$ y $\Pi_k(\rho)$, es decir, si se cumple que:

$$\langle \varphi_{nm}(\rho, \theta), \varphi_{kl}(\rho, \theta) \rangle = 2\delta_{ml} \langle \Pi_n(\rho), \Pi_k(\rho) \rangle = 2\delta_{ml} \delta_{nk} \quad (11)$$

Finalmente, y para permitir una más cómoda definición de la componente radial de las funciones de base, se presenta una última propiedad. Sean $p(x)$ y $q(x)$ dos funciones de una base ortonormal, por lo que según el producto escalar definido en la ecuación 1:

$$\langle p(x), q(x) \rangle = \int_{x_0}^{x_f} p(x) q^*(x) \, dx = \delta_{pq} \quad (12)$$

donde

$$\delta_{pq} = \begin{cases} 1 & p = q \\ 0 & p \neq q \end{cases}$$

Dada esta definición, se propone un cambio de variable, de tal forma que:

$$x = f(\rho) = \rho^2 \Rightarrow dx = f'(\rho) \, d\rho = 2\rho \, d\rho \quad (13)$$

por lo que sustituyendo en la ecuación 12 se tiene que:

$$\langle p(\rho^2), q(\rho^2) \rangle = 2 \int_{\rho_0^2}^{\rho_f^2} p(\rho^2) q^*(\rho^2) \rho \, d\rho = \delta_{pq} \quad (14)$$

siendo $x_0 = \rho_0^2$ y $x_f = \rho_f^2$. No obstante, aunque la expresión 14 es muy similar al producto escalar definido sobre la coordenada radial de un sistema polar, como el que aparece en la ecuación 10, aún depende de ρ^2 , y no de ρ . Para completar el desarrollo, es necesario por tanto realizar la composición de funciones planteada. Para ello, sea $x_1 = f(\rho_1) = \rho_1^2$ y $p(x_1) = p_1$; entonces:

$$p(x_1) = p(f(\rho_1)) = p(\rho_1^2) = p_1 \Rightarrow (p \circ f)(\rho_1) = p_1 \quad (15)$$

con $\rho_1 = \sqrt{x_1}$. Dicho de otra forma, la composición de ambas funciones supone que los pasos por cero en la coordenada ρ se producen en la raíz cuadrada de los valores de los pasos por cero para la coordenada x . Uniendo entonces las ecuaciones 14 y 15 queda:

$$\langle (p \circ f)(\rho), (q \circ f)(\rho) \rangle = 2 \int_{\rho_0}^{\rho_f} (p \circ f)(\rho) (q \circ f)^*(\rho) \rho \, d\rho = \delta_{pq} \quad (16)$$

donde $\rho_0 = \sqrt{x_0}$ y $\rho_f = \sqrt{x_f}$.

Esta propiedad permitirá más adelante, en conjunción con la ecuación 10, deducir fácilmente bases ortonormales para un sistema de coordenadas polares a partir de bases ortonormales en coordenadas cartesianas.

Una vez que se ha justificado la forma que se ha elegido para las funciones de las bases que se emplearán en el desarrollo del trabajo, queda determinar la definición de la componente radial de cada una de ellas. En este trabajo se trata de establecer una comparación entre la descripción utilizando funciones de base continuas, lo que representa la metodología de trabajo utilizada hasta ahora, frente al uso de funciones continuas a intervalos (*Piecewise Constant Basis Functions*, PCBF) (Rao, 1983) para el mismo fin. Como referencia de sistema de funciones de base continuas, en este trabajo se han reproducido todos los experimentos descritos utilizando momentos de Zernike, citados en la bibliografía como familia de momentos con una mejor capacidad de representación (Teh and Chin, 1988) (Chen and Xie, 2011). Su definición comienza con los polinomios de Zernike:

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \rho^{n-2s} \quad (17)$$

$$R(\rho) = \{R_{nm}(\rho) \mid n = 0, 1, \dots, \infty, |m| \leq n, n - |m| \text{ par}\}$$

y a partir de ellos, se define el momento de Zernike de orden nm de una imagen $I(\rho, \theta)$ como:

$$A_{nm} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 [V_{nm}(\rho, \theta)]^* I(\rho, \theta) \rho \, d\rho \, d\theta \quad (18)$$

donde $V(\rho, \theta)$ se define como:

$$V_{nm}(\rho, \theta) = R_{nm}(\rho) e^{im\theta} \quad (19)$$

Como puede verse en la ecuación 19, la función $V_{nm}(\rho, \theta)$ responde a la estructura descrita en la ecuación 4, por lo que es inmediato deducir que $\|A_{nm}\|$ es un invariante a rotación, que se utiliza como descriptor en este caso.

Quedan por definir las PCBF que se utilizarán sobre la componente radial en este estudio. En este caso se ha optado por manejar dos alternativas, representativas de las dos posibles vías de elección en este punto.

2.1. PCBF no ortonormales: tren de pulsos

En primer lugar, y como alternativa no ortonormal, se ha optado por incluir como funciones de base un tren de pulsos. Sea $\Delta_n : 0 = t_0 < t_1 < \dots < t_{n-1} < t_n < t_{n+1} = 1$, $n \in \mathbb{N}$ una partición uniforme en el intervalo $[0, 1]$, donde:

$$t_i = \frac{i}{n+1}, \quad i = 1, \dots, n+1 \quad (20)$$

Sobre esta partición, se define la función de base de orden n como:

$$\gamma_n(x) = \begin{cases} 1 & t_k < x < t_{k+1} \\ -1 & t_{k-1} < x < t_k \end{cases} \quad (21)$$

donde k es un entero par tal que $0 \leq k \leq n+1$, $k \in \mathbb{N}$. De este modo se consigue que la función de orden n tenga el mismo

número de cambios de signo. Posteriormente se ha aplicado el cambio de variable descrito en la ecuación 13 para adecuar su definición a la componente radial de un sistema de coordenadas polares. El resultado puede verse en la figura 1, donde se han representado las funciones hasta orden $n = 5$.

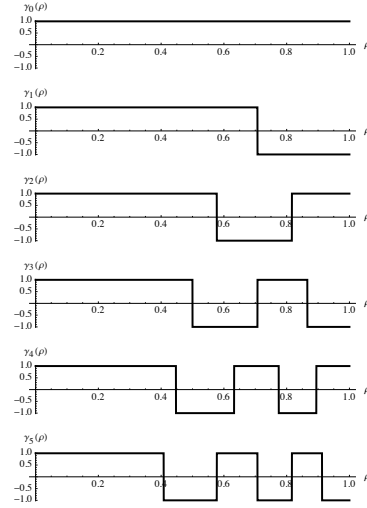


Figura 1: Funciones de la base tren de pulsos hasta orden $n = 5$.

Por tanto, la definición de la función de orden nm de esta base de momentos en coordenadas polares, aplicando la expresión general propuesta en la ecuación 4 y el cambio de variable 13, queda como:

$$\Gamma_{nm}(\rho, \theta) = \gamma_n(\rho^2) e^{im\theta} \quad (22)$$

a partir de la que se genera el descriptor invariante a rotación $\|\Gamma_{nm}(\rho, \theta)\|$.

2.2. PCBF ortonormales: funciones de Walsh

En segundo lugar, y como alternativa ortonormal, en este estudio se han incluido como funciones de base las conocidas funciones de Walsh (Walsh, 1923). Para su definición, considérense $k > 0$, $k \in \mathbb{N}$, es decir un entero no negativo, y $x \in [0, 1]$ un número real. La representación binaria de k viene dada por:

$$k = \kappa_0 2^0 + \kappa_1 2^1 + \dots + \kappa_{a-1} 2^{a-1} \quad (23)$$

mientras que la de x es:

$$x = \frac{x_1}{2^1} + \frac{x_2}{2^2} + \dots \quad (24)$$

donde $\kappa_0, \kappa_1, \dots, \kappa_{a-1}, x_1, x_2, \dots \in \{0, 1\}$. Entonces, la función de Walsh de orden k es:

$$w_k : [0, 1] \mapsto \{-1, 1\}, \quad w_k(x) = -1^{\kappa_0 x_1 + \kappa_1 x_2 + \dots + \kappa_{a-1} x_a} \quad (25)$$

comenzando la serie con $w_0(x) = 1$. Al igual que en el caso anterior, las funciones de Walsh se han ordenado de forma que el n -ésimo elemento de la base tenga n cambios de signo. A continuación, se aplica el cambio de variable propuesto en la

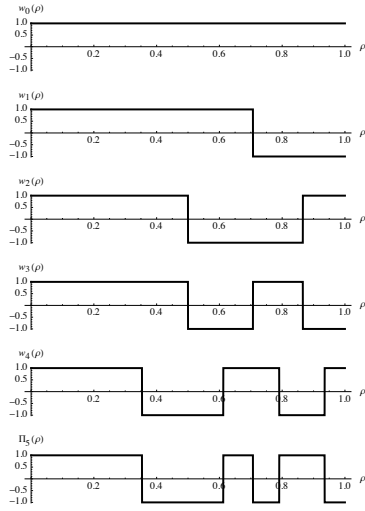


Figura 2: Funciones de la base de Walsh hasta orden $n = 5$.

ecuación 13 para su adaptación al caso polar. El resultado es el que se observa en la figura 2 donde igualmente se han representado las funciones de la base hasta orden 5.

De este modo, la definición de la función de esta base de momentos de orden nm en coordenadas polares, aplicando la expresión general propuesta en la ecuación 4 y el cambio de variable 13, queda como:

$$W_{nm}(\rho^2, \theta) = w_n(\rho^2)e^{im\theta} \quad (26)$$

y, al igual que en el caso anterior, se utiliza como descriptor $\|W_{nm}(\rho, \theta)\|$.

Mediante la elección de las dos bases descritas, es posible realizar un estudio en el que no sólo se determinará la idoneidad de las PCBF como base para la descripción de una imagen mediante invariantes basados en momentos, sino que también será posible, de la comparación entre los resultados obtenidos con la base tren de pulsos y con la base de Walsh, determinar si la ortonormalidad en tales funciones supone una ventaja adicional, como se podría deducir de las propiedades de la proyección sobre tales bases descritas en el álgebra de Hilbert.

3. Comportamiento ante ruido

Antes de poder utilizar los invariantes explicados en la sección anterior, es necesario determinar, para cada base, cuál va a ser la longitud de la descripción que se genere, es decir, cuántos momentos se van a extraer de cada una de las imágenes. Esta cuestión tiene una primera respuesta en términos de la precisión en la descripción, que indica que a mayor número de descriptores, mejor representada está una imagen. Pero aumentar la descripción aumenta igualmente el tiempo necesario para su cálculo y uso, así como los recursos necesarios para su almacenamiento y transmisión. Para ayudar a resolver este dilema, existe un criterio adicional que consiste en evaluar cómo se ven afectados los valores de los momentos cuando, como ocurre en

la mayoría de aplicaciones, las imágenes de partida están corrompidas con ruido. Experimentos anteriores demuestran que la sensibilidad al ruido aumenta a medida que se incrementa el orden de los momentos (Teh and Chin, 1988), por lo que es necesario evaluar hasta dónde se puede extender la descripción manteniendo su robustez.

Para llevar a cabo este test se han tomado cien muestras al azar de nuestra base de datos de test, que consta de más de 30.000 imágenes binarias, normalizadas a un tamaño de 128x128 píxeles¹, lo que, como se ha mencionado, garantiza la invarianza a traslación y escalado de la descripción (Khotanzad and Hong, 1990). Se ha decidido utilizar esta base de datos frente a otras de dominio público (Fisher, 2011), debido a que ninguna de las consultadas cumplía con los requisitos de simplicidad (los problemas de segmentación quedan fuera de este trabajo) y de diversidad para presentar suficientes imágenes similares a una dada, distractores y una escala comparable a la de un problema real. La más adecuada de entre las consultadas es SIID de Brown University, pero su reducido tamaño nos condujo a rechazarla finalmente en favor de la utilizada. Algunas imágenes contenidas en nuestra base de datos se muestran en la figura 3. Cada una de ellas se ha corrompido con ruido gaussiano de distinta intensidad para crear cuatro copias, correspondientes a valores de $SNR_{in} = \{4, 2, 1, 0, 5\}$. En este caso, se define:

$$SNR_{in} = \frac{\sum_{\forall(\rho, \theta)} [I(\rho, \theta) - \bar{I}(\rho, \theta)]^2}{\sum_{\forall(\rho, \theta)} [n(\rho, \theta) - \bar{n}(\rho, \theta)]^2} \quad (27)$$

donde $n(\rho, \theta)$ es el ruido aditivo, y $\bar{I}, \bar{n}$ representa los respectivos valores medios de la imagen y de la señal de ruido (Chen and Xie, 2011).



Figura 3: Ejemplos de imágenes en la base de datos de test.

Cada una de estas 400 imágenes con ruido se han transformado utilizando las tres familias de momentos propuestas para generar los respectivos invariantes: Zernike, $\|A_{nm}\|$, tren de pulsos, $\|I_{nm}\|$ y Walsh, $\|W_{nm}\|$.

Para evaluar la robustez de cada uno de ellos ante el ruido, se ha calculado el SNR_{nm} de cada orden de invariantes, siendo:

$$SNR_{nm} = \frac{\sum_{i=1}^{100} [\phi_{nm}^i - \bar{\phi}_{nm}]^2}{\sum_{i=1}^{100} [\kappa_{nm}^i - \bar{\kappa}_{nm}]^2} \quad (28)$$

¹La base de datos de test está disponible enviando un e-mail al autor

donde ϕ_{nm}^i representa el invariante de orden nm ($\|A_{nm}\|$, $\|\Gamma_{nm}\|$ o $\|W_{nm}\|$) de la i -ésima imagen, y κ_{nm}^i representa el invariante de orden nm ($\|A_{nm}\|$, $\|\Gamma_{nm}\|$ o $\|W_{nm}\|$) del ruido añadido a dicha imagen (Teh and Chin, 1988).

En los resultados presentados a continuación se han utilizado los siguientes órdenes:

- $\|A_{nm}\|$: $m = 0$ para n par, y $m = 1$ para n impar.
- $\|\Gamma_{nm}\|$: $m = 0$ para todos los valores de n .
- $\|W_{nm}\|$: $m = 0$ para todos los valores de n .

Los resultados se presentan en la figura 4, en la que puede corroborarse el patrón de degradación de SNR_{nm} para las tres familias de momentos, en los cuatro casos propuestos correspondientes a los cuatro niveles de corrupción con ruido gaussiano ($\text{SNR}_{in} = \{4, 2, 1, 0.5\}$). En línea continua se observan los resultados para $\|A_{nm}\|$, en punteada para $\|\Gamma_{nm}\|$ y en discontinua para $\|W_{nm}\|$.

De los resultados observados en la figura 4, y tomando como referencia $\|A_{nm}\|$, se pueden extraer varias conclusiones:

- El comportamiento ante ruido de $\|\Gamma_{nm}\|$ es muy parecido al de $\|A_{nm}\|$, lo que indica que se puede extender la descripción en ambos casos al mismo número de invariantes manteniendo la misma calidad de la representación en términos de robustez ante el ruido. Por lo tanto en las pruebas posteriores se utilizará la misma longitud de descripción en ambos casos. La ventaja de utilizar PCBF se demostrará, por tanto, si se consiguen resultados significativamente mejores en el desarrollo de dichas pruebas.
- Por contra, el comportamiento de $\|W_{nm}\|$ es ligeramente peor, puesto que se observa que alrededor del orden 7 hay un empeoramiento significativo para todos los SNR_{in} . Por lo tanto, en las pruebas posteriores es preciso limitar la longitud de la representación a órdenes inferiores de invariantes que garanticen la robustez ante el ruido. La ventaja de esta PCBF se demostrará si con menor longitud de representación se alcanzan los mismos o mejores resultados que usando $\|A_{nm}\|$.

4. Aplicación en recuperación de imágenes

En esta sección se detallan las pruebas realizadas para la evaluación de las PCBF como descriptores de imágenes en una tarea de recuperación de imágenes basada en contenido, usando para ello dos bases de datos diferentes: por un lado la base de datos pública de la Universidad de Brown, con 216 imágenes de objetos en 18 categorías (Sharvit et al., 1998), y por otro nuestra base de datos de test, con más de 30.000 imágenes heterogéneas. Cabe reseñar que las imágenes contenidas son binarias, puesto que tradicionalmente, los momentos y sus invariantes se han utilizado como descriptores de forma, por lo que en estos tests no se tiene en cuenta la información de luminancia.

Para el dimensionamiento de la descripción asociada a cada tipo de invariantes, se han tomado como base los momentos de Zernike, $\|A_{nm}\|$. En este experimento se ha utilizado hasta

orden 17 en este descriptor, que representa un compromiso entre longitud y calidad avalado por los resultados presentes en la literatura (Teh and Chin, 1988) (Mukundan, 2004), y que adicionalmente mantienen un buen comportamiento ante ruido; la extensión de esta descripción queda fijada entonces en 90 invariantes. A partir de esta referencia, y teniendo en cuenta las consideraciones expuestas en la sección anterior, i.e. misma longitud para $\|\Gamma_{nm}\|$, y menor para $\|W_{nm}\|$, se ha diseñado el experimento según las condiciones reflejadas en las tablas 1 y 2

Tabla 1: Detalles de los descriptores en el experimento de recuperación de imágenes

	$\ A_{nm}\ $	$\ \Gamma_{nm}\ $	$\ W_{nm}\ $
Orden radial	0...17	0...8	0...6
Orden angular	0...17	0...9	0...7
N. descriptores	90	90	56

Tabla 2: Detalles de la recuperación

	Brown	Propia
Tamaño DB	216	30,000
N. Consultas	18	100
Imágenes recuperadas	25	25
Total pares evaluados	450	2500

La comparación de los resultados alcanzados utilizando cada una de las familias de invariantes se realiza mediante la medida de precisión, que se define como:

$$\text{precisión} = \frac{\text{imágenes correctamente recuperadas}}{\text{tamaño del conjunto recuperado}} \quad (29)$$

Cabe reseñar que su medida complementaria, la medida de *recall*, no está disponible por una cuestión de escala: realizar el etiquetado como similar o no de todos los posibles pares generados en 100 consultas sobre una base de datos de más de 30.000 imágenes supondría evaluar más de 3,000,000 de casos, lo que por sí sólo constituye una tarea enorme.

En el primer experimento con la base de datos de Brown University, se realizan 18 consultas, una correspondiente a cada categoría presente en la base de datos, y se recuperan 25 imágenes por cada una de ellas. Se calcula la precisión para cada tamaño del conjunto de recuperación, siendo los resultados alcanzados los mostrados en la figura 5

A continuación, se repite el experimento con la base de datos propia, realizando cien consultas, y siendo los resultados alcanzados los que se muestran en la figura 6.

En estas gráficas se puede observar que:

- La comparación entre $\|A_{nm}\|$ y $\|\Gamma_{nm}\|$ deja un resultado en el que las precisiones medias alcanzadas en ambos casos para cada tamaño del conjunto recuperado son muy similares, si bien algo desfavorables a $\|\Gamma_{nm}\|$. Si se tiene en cuenta que el conjunto de invariantes $\|A_{nm}\|$ proviene

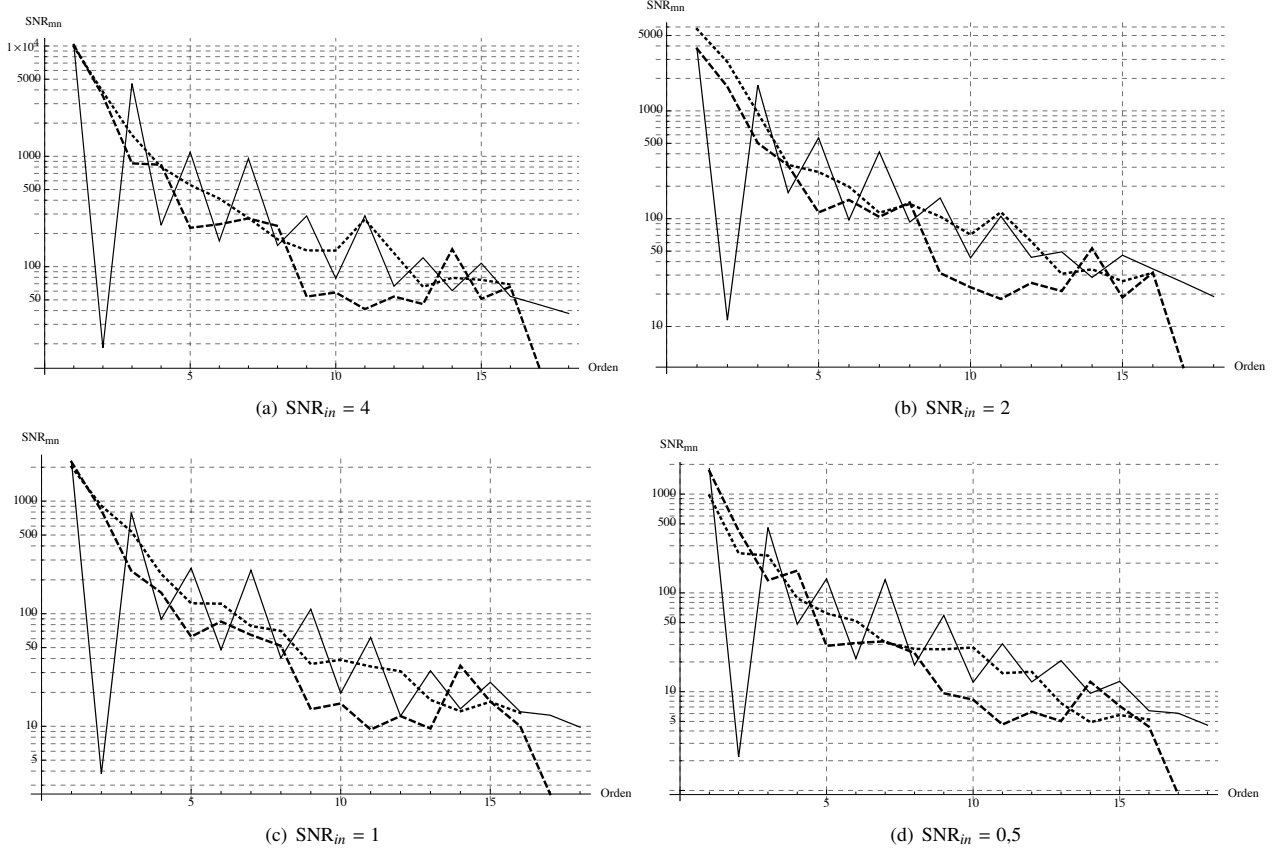


Figura 4: Comportamiento ante ruido de las tres familias de momentos: $\|A_{nm}\|$ línea continua, $\|\Gamma_{nm}\|$ punteada, $\|W_{nm}\|$ discontinua.

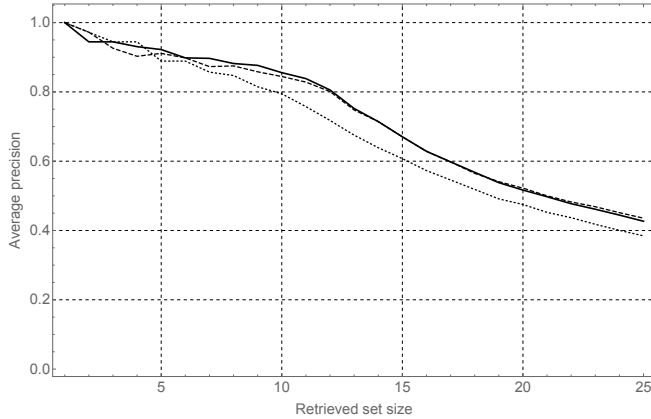


Figura 5: Comparación entre las precisiones alcanzadas con cada familia de momentos en el experimento de recuperación de imágenes con la base de datos de la Universidad de Brown. $\|A_{nm}\|$, trazo continuo, $\|\Gamma_{nm}\|$ puntos, $\|W_{nm}\|$ discontinuo.

de la proyección de la imagen sobre una familia de funciones ortonormales, $V_{nm}(\rho, \theta)$, mientras que el conjunto $\|\Gamma_{nm}\|$ se genera mediante un conjunto de funciones no ortonormales, $\gamma_{nm}(\rho, \theta)$, unido a que ambos conjuntos de invariantes tienen el mismo tamaño, se puede dedu-

cir que la capacidad de descripción de estas últimas es mayor, puesto que existiendo redundancia entre los descriptores generados (en el sentido de que la información contenida en cada uno no es exclusiva (Rudin, 1991)) se alcanzan los mismos resultados. Dicho de otra forma, eliminando la redundancia se generaría una descripción más corta con los mismos resultados. Lo que nos lleva a la siguiente observación.

- La comparación entre $\|A_{nm}\|$ y $\|W_{nm}\|$ revela también un resultado con precisiones medias muy similares, aunque en este caso ligeramente a favor de $\|W_{nm}\|$. Teniendo en cuenta que la descripción mediante estas PCBF es, aproximadamente, un 40 % más corta que la realizada mediante la base de Zernike, se demuestra la mejor capacidad de representación de estas PCBF ortonormales, puesto que a menor longitud de la descripción se obtienen los mismos resultados.

Como conclusión se puede establecer, sobre los casos propuestos, que no se justifica el uso de $\|\Gamma_{nm}\|$ frente a los invariantes de Zernike, puesto que se mantiene la misma longitud de representación, pero la no ortonormalidad de estas PCBF supondrían un inconveniente en algunas tareas, como por ejemplo la reconstrucción de la imagen original a partir de los momentos extraídos (Rudin, 1991) (Teague, 1980). Sin embargo, en el

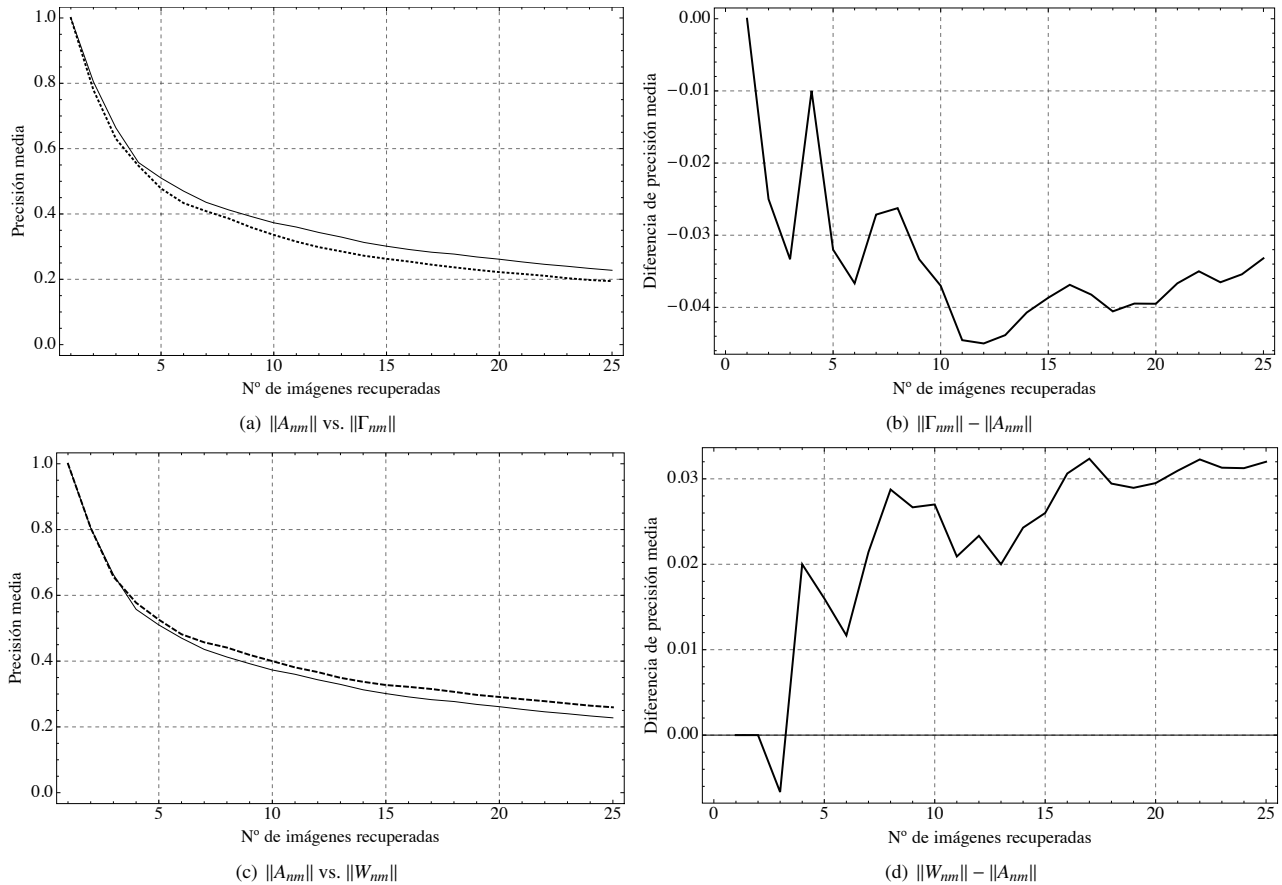


Figura 6: Comparación entre las precisiones alcanzadas con cada familia de momentos en el experimento de recuperación de imágenes. Derecha (a) y c): series de precisiones medias. Izquierda (b) y d): diferencia entre las series. $\|A_{nm}\|$, trazo continuo, $\|\Gamma_{nm}\|$ puntos, $\|W_{nm}\|$ discontinuo.

caso de $\|W_{nm}\|$, las ventajas de utilizar esta descripción frente a $\|A_{nm}\|$ es muy clara, y se puede cifrar en:

- Una representación más corta supone menos tiempo de cómputo para la generación de los descriptores, tanto de la imagen de consulta como de las almacenadas en la base de datos. En nuestro caso, esta reducción se cifraría en un 40 % menos (asumiendo el mismo coste computacional por cada invariante generado)
- La reducción también se aplica en la misma proporción a los recursos computacionales necesarios para el almacenamiento y transmisión de los datos.
- También se produce una reducción significativa, que en su caso habría que cuantificar, en la gestión de la base de datos, al reducirse la dimensionalidad de los datos sobre los que generar los índices multidimensionales.
- La reducción también se refleja en el coste computacional de cada proceso de consulta, puesto que la comparación de cada par de imágenes se realiza sobre descripciones mucho más cortas, lo que supone menos operaciones y menos ciclos de CPU necesarios para alcanzar una conclusión.

- También se reducirían las necesidades computacionales en el caso de tener que reconstruir la imagen original a partir de la descripción, puesto que manteniéndose la simplicidad del proceso debido a la ortornormalidad de esta base, el número de componentes es radicalmente menor.

No obstante, tal y como ya se ha apuntado anteriormente, no es frecuente que en una aplicación real las imágenes estén libres totalmente de ruido, por lo que para confirmar las ventajas de utilizar PCBF frente a bases continuas, como la de Zernike, en la generación de invariantes, es necesario repetir el experimento detallado anteriormente en unas condiciones más realistas. Con este fin, se tomaron las mismas 100 imágenes utilizadas en el experimento anterior y se corrompieron con ruido gaussiano para generar dos copias de cada una de ellas, una caracterizada por un $\text{SNR}_{in} = 1$ y otra con un $\text{SNR}_{in} = 0,5$. Con cada uno de estos grupo de de cien imágenes, se realizaron otras tantas consultas a la base de datos, y se calcularon finalmente las precisiones medias para cada tamaño del conjunto recuperado. Los resultados alcanzados se muestran en la figura 7. Algunos de los valores más significativos de los experimentos de recuperación de imágenes con y sin ruido se han recogido en la tabla 3

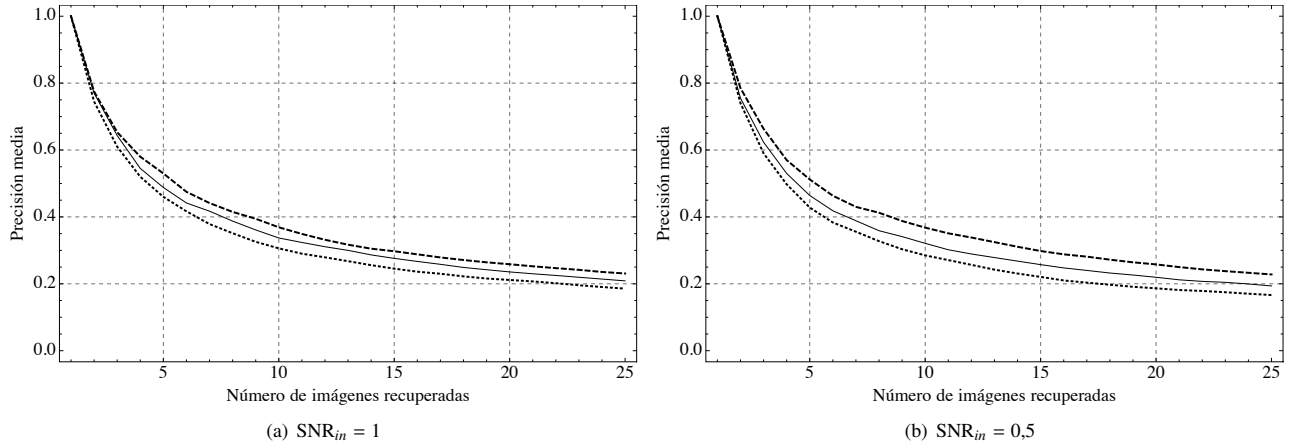


Figura 7: Comparación entre las precisiones alcanzadas con cada familia de momentos en el experimento de recuperación de imágenes con presencia de ruido. $\|A_{nm}\|$, trazo continuo, $\|\Gamma_{nm}\|$ puntos, $\|W_{nm}\|$ discontinuo.

Tabla 3: Recopilación de resultados de los experimentos de recuperación de imágenes. P_i se refiere a la precisión media alcanzada para un conjunto de i elementos recuperados.

SNR_{in}	Tipo	# invariantes	P_5	P_{10}	P_{15}	P_{20}	P_{25}
Sin Ruido	$\ A_{nm}\ $	90	0.510	0.373	0.301	0.262	0.227
	$\ \Gamma_{nm}\ $	90	0.478	0.336	0.263	0.222	0.194
	$\ W_{nm}\ $	56	0.526	0.376	0.317	0.279	0.252
1.0	$\ A_{nm}\ $	90	0.488	0.337	0.276	0.235	0.209
	$\ \Gamma_{nm}\ $	90	0.46	0.306	0.245	0.211	0.186
	$\ W_{nm}\ $	56	0.520	0.369	0.297	0.259	0.231
0.5	$\ A_{nm}\ $	90	0.464	0.321	0.257	0.219	0.194
	$\ \Gamma_{nm}\ $	90	0.428	0.285	0.22	0.186	0.166
	$\ W_{nm}\ $	56	0.512	0.368	0.298	0.258	0.228

Tanto en la figura 7 como en la tabla 3 se puede comprobar que los tres invariantes estudiados en este trabajo tienen un comportamiento bastante similar en términos generales. No obstante, en ambos casos, $SNR_{in} = \{1, 0.5\}$, se puede comprobar también que se mantienen las tendencias que ya se apuntaban para el experimento de recuperación de imágenes en ausencia de ruido. Por un lado, un comportamiento ligeramente peor para $\|\Gamma_{nm}\|$ en relación con la referencia continua $\|A_{nm}\|$, y por otro, se observa a la vez un comportamiento ligeramente mejor para $\|W_{nm}\|$. Incluso se puede observar una degradación menor de estos últimos invariantes cuando el nivel de ruido aumenta desde $SNR_{in} = 1$ hasta $SNR_{in} = 0.5$. Aunque esta tendencia debería ser confirmada mediante experimentos específicos, el motivo podría ser la reducción en los órdenes de momentos utilizados en la generación de los invariantes, forzada por la mayor presencia de ruido en órdenes altos, como se apuntaba anteriormente en este artículo.

5. Conclusiones

En este artículo se presenta un estudio del uso de funciones de base discontinuas a intervalos (PCBF) como alternativa a las funciones de base continuas en la aplicación del método

de los momentos al análisis de imágenes. Para ello, se propone la construcción de invariantes a partir de estos momentos, y se compara su comportamiento en tareas de recuperación de imágenes por contenido, probando dos alternativas discontinuas (tren de pulsos y funciones de Walsh), frente a la familia de funciones comúnmente aceptada como la más potente en la descripción de imágenes (Zernike). En primer lugar, se adapta la formulación de las dos bases discontinuas para su aplicación a la coordenada radial en un sistema de coordenadas polares. La característica de las bases propuestas es que la primera, el tren de pulsos, está compuesta de funciones discontinuas pero no ortonormales, mientras que la segunda, las funciones de Walsh, si cumplen con esta propiedad.

En las pruebas, los resultados apuntan a las ventajas de utilizar este tipo de funciones frente a la alternativa tradicional de utilizar bases continuas, como Zernike. Por un lado, la base no ortonormal demuestra similares resultados utilizando una descripción del mismo tamaño, lo que no justificaría su empleo. Sin embargo, la base de Walsh, muestra también los mismos resultados pero haciendo uso de una descripción mucho menor, aproximadamente un 40 % más corta; esto implica que su uso supone una reducción del mismo orden en los recursos computacionales necesarios para su generación y mantenimiento, así como una simplificación algorítmica en su manejo, siendo todos estos

ahorros francamente deseables en cualquier aplicación real.

Adicionalmente, los descriptores propuestos no sólo han sido probados en escenarios de aplicación ideales, en los que las imágenes utilizadas están libres de ruido, sino que se les ha sometido a la resolución de tareas más realistas, donde los datos se han corrompido con distintos grados de ruido gaussiano, manteniéndose en todo caso un comportamiento parejo, si no superior, al alcanzado con los momentos de Zernike.

Como conclusión, los resultados aquí presentados demuestran que el uso de PCBF como alternativa a las funciones de base continuas en la aplicación del método de los momentos, y en concreto la adaptación de la base de Walsh al caso de coordenadas polares, es muy ventajoso en términos computacionales frente al uso de funciones de base continuas, apuntando incluso mejor comportamiento en entornos ruidosos, aunque este es un extremo que habrá que confirmar con experimentos destinados a tal fin.

English Summary

Image analysis by the method of moments using Piecewise Continuous Basis Functions (PCBF)

Abstract

Invariants generated departing from moments, previously extracted from an image, appear frequently in the bibliography as one of the most powerful means of describing images, and more precisely shapes. In this paper, the use of Piecewise Continuous Basis Functions (PCBF) is proposed as an alternative to those basis which have been used traditionally in the method of moments, all of them continuous as the well known Zernike, Legendre or Tchebichev basis. The use of discontinuous basis can be justified by the own discontinuous nature of the object of such analysis, namely images: it is thoroughly known that the contours of visible objects are modeled as discontinuities in the series of luminance values as we go from one side of the border to the other. Analyzing such discontinuous objects by means of continuous functions can lead to undesired results, as the Gibbs phenomenon, that can be avoided by simply shifting to discontinuous basis for the analysis, getting better approximations to the described object. Additionally, the proposed basis can easily generate, as shown in this paper, rotation invariants, which is a very desirable feature for a shape descriptor, given that the orientation that the shape will have in an image is not known in advance. Translation and scale invariance is obtained by means of a simple normalization process. Test confirming this hypothesis are presented as well, starting with an analysis of the behavior of the proposed invariants in noisy environments, which allow to fix the number of invariants that have to be extracted. Next, once this description length has been determined, new experiments are carried out to assess the performance of the proposed invariants in a content based retrieval task, both in a noise

free and in noisy environments, having images corrupted with different gaussian noise intensities. Results confirm our hypothesis that these descriptors are very well suited for this task, showing that they can achieve results similar to those obtained using the continuous reference basis, which is Zernike's, but with a description which is roughly a 40 % shorter.

Keywords:

Method of moments, invariant descriptors, orthonormal basis, content based image retrieval, image analysis

Referencias

- Bovik, A., 2009. The Essential Guide to Image Processing. Academic Press.
- Chen, G., Xie, W., 2011. Wavelet-based moment invariants for pattern recognition. *Optical Engineering* 50 (7), 077205–1 – 077205–9.
- Domínguez, S., 2013. Image analysis by moment invariants using a set of setp-like basis functions. *Pattern Recognition Letters* 34, 2065–2070.
- Fisher, R., 2011. URL: homepages.inf.ed.ac.uk/rbf/CVonline/Imagedbase.htm
- Flusser, J., 2000. On the independence of rotation moment invariants. *Pattern Recognition* 33, 1405–1410.
- Flusser, J., Suk, T., Zitova, B., 2009. Moments and Moment Invariants in Pattern Recognition. John Wiley and Sons Ltd.
- Hamood, M. T., Boussakta, S., November 2011. Fast walsh-hadamard-fourier transform algorithm. *IEEE Transactions on Signal Processing* 59 (11), 5627–5631.
- Hewitt, E., Hewitt, R. E., 1979. The gibbs-wilbraham phenomenon: An episode in fourier analysis. *Archive for History of Exact Sciences* 21, 129–160.
- Hu, M.-K., 1962. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory* 8, 179–187.
- Khotanzad, A., Hong, Y. H., 1990. Invariant image recognition by zernike moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (5), 489–497.
- Krommweh, J., 2009. An orthonormal basis of directional haar wavelets on triangles. *Results in Mathematics* 53, 323–331.
- Lee, B. Y., Tarn, Y. S., 1999. Application of the discrete wavelet transform to the monitoring of tool failure in end milling using the spindle motor current. *International Journal of Advanced Manufacturing Technology* 15, 238–243.
- Mukundan, R., 2004. Some computational aspects of discrete orthonormal moments. *IEEE Transactions on Image Processing* 13 (8), 1055–1059.
- Rao, G. P., 1983. Piecewise constant orthogonal functions and their application to systems and control. Springer Verlag.
- Reiss, T. H., 1991. The revised fundamental theorem of moment invariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13 (8), 830–834.
- Rudin, W., 1991. Functional Analysis. McGraw Hill.
- Sasikala, D., Neevelani, R., May 2010. Registration of brain images using fast walsh hadamard transform. *International Journal of Computer Science and Information Security* 8 (2), 96–105.
- Sharvit, D., Chan, J., Tek, H., Kimia, B. B., December 1998. Symmetry-based indexing of image databases. *Journal of Visual Communication and Image Representation* 9 (4), 366–380.
- Teague, M. R., 1980. Image analysis via the general theory of moments. *Journal of the Optical Society of America* 70 (8), 920 – 930.
- Teh, C.-H., Chin, R. T., 1988. On image analysis by the methods of moments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 10 (4), 496–512.
- Walsh, J., 1923. A closed set of normal orthogonal functions. *American Journal of Mathematics* 45, 5–24.
- Xu, D., Li, H., 2008. Geometric moment invariants. *Pattern Recognition* 41, 240–249.