



**Universitat
Autònoma
de Barcelona**

Revisión de la teoría de los *Textons*
Enfoque computacional en color

Tesis presentada por **Susana Álvarez Fernández**
en la Universidad Autónoma de Barcelona para op-
tar al título de **Doctor en Informática**.

Bellaterra, 25 de Mayo del 2010.

Director: **Dra. Maria Vanrell i Martorell**
Universitat Autònoma de Barcelona
Dept. Ciències de la Computació y Centre de Visió per Computador

Co-director: **Dr. Xavier Otazu Porter**
Universitat Autònoma de Barcelona
Dept. Ciències de la Computació y Centre de Visió per Computador

A mi familia

Agradecimientos

En primer lugar agradecer a Maria Vanrell la oportunidad que me brindó de formar parte de su equipo y la dedicación de parte de su tiempo, ese valioso recurso en estos días. También como no, a Xavier Otazu por colaborar en todo cuanto ha estado en su mano. Ambos son en parte responsables de este trabajo, que espero que haya cumplido sus expectativas.

Por otro lado tengo que agradecer a mis compañeros de oficio que en más de una ocasión me han prestado su ayuda, siempre desinteresada, en diversos problemas que han surgido durante el transcurso de esta tesis. Entre ellos se encuentra Santi, compañero de despacho y sufridor directo de mis enfados cuando las cosas no transcurrían como me hubiera gustado. No puedo olvidarme de algunos de los matemáticos del departamento, Toni, Carlos B. y Gerard, siempre solícitos, que en algún momento han solucionado mis dudas. Y sobretodo a Carme y Dolors por su ayuda con la notación matemática, entre otras muchas cosas.

También a Maria, Josep B., Albert F., Francesc, Carlos G.B, Carlos M., Carlos A., todos ellos vecinos de despacho que me han aguantado, sobretodo estos últimos y vertiginosos meses, durante los cuales me han animado a continuar.

A Agustí i Sergi, que siempre me han ayudado a resolver los múltiples problemas que aparecen cuando uno no domina los aspectos técnicos del Latex.

También agradecer a mis compañeros del grupo de investigación, Robert, Anna, Ramón y Javier, que aunque se encontraban a mayor distancia física, siempre que ha surgido la oportunidad me han echado una mano.

Y finalmente agradecer a Marcel su apoyo incondicional en todo momento. Sus consejos en más de una ocasión me han ayudado a seguir adelante.

En definitiva gracias a todos los que de una manera u otra han contribuido a la consecución de este trabajo.

Resumen

El color y la textura son dos estímulos visuales importantes para la interpretación de las imágenes. La definición de descriptores computacionales que combinan estas dos características es aún un problema abierto. La dificultad se deriva esencialmente de la propia naturaleza de ambas, mientras que la textura es una propiedad de una región, el color es una propiedad de un punto.

Hasta ahora se han utilizado tres los tipos de aproximaciones para la combinación, (a) se describe la textura directamente en cada uno de los canales color, (b) se describen textura y color por separado y se combinan al final, y (c) la combinación se realiza con técnicas de aprendizaje automático. Considerando que este problema se resuelve en el sistema visual humano en niveles muy tempranos, en esta tesis se propone estudiar el problema a partir de la implementación directa de una teoría perceptual, la teoría de los *textons*, y explorar así su extensión a color.

Puesto que la teoría de los *textons* se basa en la descripción de la textura a partir de las densidades de los atributos locales, esto se adapta perfectamente al marco de trabajo de los descriptores holísticos (bag-of-words). Se han estudiado diversos descriptores basados en diferentes espacios de *textons*, y diferentes representaciones de las imágenes. Asimismo se ha estudiado la viabilidad de estos descriptores en una representación conceptual de nivel intermedio.

Los descriptores propuestos han demostrado ser muy eficientes en aplicaciones de recuperación y clasificación de imágenes, presentando ventajas en la generación de vocabularios. Los vocabularios se obtienen cuantificando directamente espacios de baja dimensión y la perceptualidad de estos espacios permite asociar semántica de bajo nivel a las *palabras visuales*. El estudio de los resultados permite concluir que si bien la aproximación holística es muy eficiente, la introducción de co-ocurrencia espacial de las propiedades de forma y color de los blobs de la imagen es un elemento clave para su combinación, hecho que no contradice las evidencias en percepción.

Resum

El color i la textura son dos estímuls visuals importants en la interpretació de les imatges. La definició de descriptors computacionals que combinen aquestes dues característiques és encara un problema obert. La dificultat es deriva essencialment de la pròpia natura de les dues, mentre que la textura és una propietat d'una regió, el color és una propietat d'un punt.

Fins ara s'han utilitzat tres tipus d'aproximacions per a la combinació, (a) descrivint la textura directament en cadascun dels canals de color, (b) describint la textura i el color per separat i es combinen al final, i (c) fent la combinació amb tècniques d'aprenentatge automàtic. Considerant que aquest problema es resol en el sistema visual humà en nivells molt primaris, en aquesta tesi es proposa estudiar el problema a partir de la implementació directa d'una teoria perceptual, la teoria dels *textons*, i explorar d'aquesta manera la seva extensió a color.

Donat que la teoria dels *textons* es basa en la descripció de la textura a partir de les densitats dels atributs locals, això s'adapta perfectament al marc de treball dels descriptors holístics (bag-of-words). S'han estudiat diversos descriptors basats en diferents espais de *textons*, i diferents representacions de les imatges. Així mateix s'ha estudiat la viabilitat d'aquest descriptors en una representació conceptual de nivell mig.

Els descriptors proposats han demostrat ser molt eficients en aplicacions de recuperació i classificació d'imatges, presentant avantatges en la generació de vocabularis. Els vocabularis s'obtenen al realitzar la quantificació directament en espais de baixa dimensió y la seva perceptualitat d'aquests espais permet associar semàntica de baix nivell a les paraules visuals. L'estudi del resultat permet concloure que si bé l'aproximació holística és molt eficient, la introducció de co-ocurrència espacial de les propietats de forma i color dels blobs de la imatge és un element clau en la seva combinació, fet que no contradiu les evidències en la percepció.

Abstract

Colour and texture are important visual cues for image understanding. The definition of computational descriptors that combine both features is still an open problem. The difficulty is essentially due to the inherent nature of both cues, while texture is a property of a region, colour is a property of a point.

Since now three approaches have been used for combining cues, (a) texture is directly described in each one of the colour channels, (b) texture and colour are described separately and combined in a latter step, and (c) the combination is done using machine learning techniques. Considering that this issue is solved at early stages of the human visual system, in this work we propose to study the problem using a direct implementation of a perceptual theory, the *texton* theory, and to explore its extension to colour.

Since texton theory is based on the description of texture by the densities of local attributes, this matches perfectly with an holistic framework where descriptors are based on bag-of-words. Some descriptors based on different textons spaces and different image representations have been studied. Furthermore, the feasibility of these descriptors has also been studied for intermediate levels of image representation.

The proposed descriptors have proved high efficiency in retrieval and image classification. They also present some advantages in vocabulary generation. The quantification is done directly on low-dimensional spaces, whose perceptual properties allow low-level semantic associations to the visual words. The results make us to conclude that although the performance of holistic approaches is high, the introduction of spatial co-occurrence of blob properties, shape and colour, is a key element for their combination. This conclusion agrees with perceptual evidences.

Índice general

1. Introducción	1
1.1. Motivación	1
1.2. Objetivos	2
2. Antecedentes	5
2.1. Introducción	5
2.2. Color	5
2.2.1. Física del color	6
2.2.2. Percepción humana del color	7
2.2.3. Representación computacional del color	10
2.2.4. Nombres de colores	15
2.2.5. Descriptores computacionales del color	16
2.3. Textura	17
2.3.1. Antecedentes en la Ciencia Cognitiva	19
2.3.2. Descriptores computacionales de la textura	22
2.4. Textura y Color	25
2.4.1. Descriptores computacionales de la textura en color	26
2.5. Problemas abiertos	29
3. Representación holística de texturas en color	31
3.1. Introducción	31
3.2. Teoría de los <i>textons</i>	32
3.3. Representaciones holísticas y aproximación BoW	33
3.4. Descomposición de la imagen en blobs	34
3.4.1. Blobs perceptuales	34
3.4.2. Blobs en color	37
3.4.3. <i>Textons</i> : atributos de los blobs	40
3.5. Descriptor de <i>Textons</i> (TD)	41
3.5.1. Generación del vocabulario	42
3.5.2. Representación de la imagen	42

3.5.3.	Evaluación de rendimiento	43
3.5.4.	Experimento 1: Evaluación del <i>TD</i>	47
3.6.	Descriptor Perceptual Conjunto de <i>Textons</i> (<i>JPTD</i>)	49
3.6.1.	Representación de las características	50
3.6.2.	Generación del Vocabulario	51
3.6.3.	Representación de la imagen	54
3.6.4.	Experimento 2: Evaluación del <i>JPTD</i>	55
3.7.	Descriptor Perceptual de <i>Textons</i> (<i>PTD</i>)	59
3.7.1.	Vocabulario y Representación de la imagen	59
3.7.2.	Experimento 3: Evaluación del <i>PTD</i>	60
3.8.	Aplicación a la clasificación de texturas	66
3.8.1.	Resultados experimentales y evaluación	68
4.	Hacia una representación conceptual de texturas	71
4.1.	Introducción	71
4.2.	Representación conceptual básica	71
4.2.1.	Agrupación por similitud	72
4.2.2.	Similitud entre representaciones conceptuales	76
4.2.3.	Vocabulario básico de texturas en color	78
4.3.	Aplicación: Recuperación de imágenes de texturas	80
4.3.1.	Resultados experimentales	81
4.3.2.	Discusión sobre los resultados experimentales	81
4.3.3.	Evaluación cualitativa	86
4.4.	Aplicación: <i>Browsing</i>	86
4.5.	Aplicación: Detección de texturas en imágenes	88
4.5.1.	Agrupación por proximidad y similitud	89
4.5.2.	Integración de componentes texturales	94
4.5.3.	Experimento 1: Recuperación de imágenes en BD de Mosaicos	95
4.5.4.	Experimento 2: Recuperación de imágenes en BD con escenas naturales	97
5.	Conclusiones	101
5.1.	Aportaciones	102
5.2.	Líneas futuras	103
A.	Tablas con resultados experimentales	105
A.0.1.	Resultados experimentales con el descriptor <i>JPTD</i>	105
A.0.2.	Resultados experimentales con el descriptor <i>PTD</i>	111
B.	BD de imágenes	117
	Referencias	135

Capítulo 1

Introducción

En este capítulo se explican las motivaciones que han dado lugar a la realización de este trabajo, así como los objetivos que se han perseguido. Al final del capítulo se proporciona un resumen del contenido de los capítulos que componen este documento.

1.1. Motivación

Uno de los objetivos de la Visión por Computador es obtener la información relevante de un entorno físico que permita automatizar algunas de las tareas que realiza el ser humano, como identificar elementos en una escena, buscar imágenes digitales por su contenido, etc. Para ello se utilizan tecnologías computacionales y se desarrollan representaciones matemáticas apropiadas para facilitar el procesado automático de la información, capturada en forma de imagen.

Dentro del ámbito de la Visión por Computador, unos de los estímulos visuales más utilizados como fuentes de información para el procesamiento de las imágenes son la textura y el color. En la bibliografía se encuentran multitud de trabajos donde se ha estudiado la textura, utilizando para ello la información acromática de la imagen. Son menos los estudios en los que se ha utilizado también la información cromática. Ambos estímulos visuales, la textura y el color, tienen características diferentes. Mientras la textura es una propiedad espacial, el color es una propiedad esencialmente del punto de la imagen. Esta diferencia básica ha dificultado la integración de ambos estímulos visuales, siendo hoy en día un problema todavía abierto.

La información textural de una imagen en algunos casos puede ser suficiente para identificar o discriminar los elementos presentes en una imagen, mientras que en otros casos la información de color puede ser determinante. Lo que parece claro, y muchos investigadores coinciden en ello, es que la suma de ambos estímulos visuales proporciona más información y por tanto ha de ofrecer mejores resultados en las aplicaciones derivadas que el uso exclusivo de cada uno de ellos. Todavía no está claro como combinar ambas informaciones, si tratando de manera separada ambos estímulos y finalmente combinándola o realizando un tratamiento conjunto. Aún cuando ya se han realizado algunos estudios empíricos para determinar que método es mejor (Drimbarean y Whelan, 2001;

Mäenpää y Pietikäinen, 2004), no existe una conclusión clara al respecto.

En relación a los descriptores utilizados para representar el contenido textural, quedan pendientes por resolver algunos aspectos. El primero de ellos hace referencia a la falta de un descriptor general de texturas. La mayor parte de los descriptores desarrollados hasta ahora son específicos y representan de manera adecuada subconjuntos de texturas. El segundo aspecto a considerar es que los descriptores de texturas no tienen una semántica asociada. Los conceptos como la regularidad-aleatoriedad, homogeneidad-heterogeneidad o rugosidad relacionados con las características perceptuales de las texturas, están alejados de las características locales obtenidas por los descriptores, aunque en muchos casos se intentan correlacionar. Esta problemática no existe para el caso de los descriptores utilizados para describir el color, ya que se ha encontrado la relación entre el estímulo físico producido por el color con el nombre del color que le corresponde.

1.2. Objetivos

El objetivo principal de este trabajo es desarrollar un descriptor computacional de texturas en color con las siguientes cualidades:

Que integre la textura y el color combinando ambos aspectos en un único concepto.

Generalidad que sea aplicable a cualquier tipo de texturas.

Perceptual que tenga una correlación con la percepción humana.

Con cierta semántica cuyo significado tenga una cierta traducción a palabras del lenguaje natural.

Eficiente con un coste computacional bajo para que sea factible su uso en aplicaciones prácticas.

Partiendo de estos objetivos en el capítulo 2 se ha realizado una extensa revisión de los trabajos publicados relacionados con la textura y el color tratados como estímulos independientes, tanto desde el punto de vista computacional como desde el punto de vista psicofísico. Así mismo se hace una pequeña recopilación de los pocos trabajos donde se utilizan ambos estímulos, color y textura, como base para representar el contenido de las imágenes. Se finaliza el capítulo analizando la problemática derivada de la combinación de estas características.

En el capítulo 3 se aborda con más profundidad el estudio psicofísico de Julesz y Bergen (1983) que ha servido de inspiración para el desarrollo del descriptor computacional de texturas en color que se propone. A continuación se explica la aproximación que permite la descomposición de la imagen en sus partes elementales denominadas *blobs perceptuales*. Estos constituyen la base de dos modelos diferentes para representar las texturas. Por un lado se desarrolla un modelo general para representar texturas en color, la llamada representación *holística*. Ésta representación utiliza

la aproximación *Bag-of-words* (*BoW*) que ha sido aplicada mayoritariamente en el contexto de categorización de escenas y objetos (G. Wang y cols., 2006; Sivic y cols., 2005) y en descripción de materiales (Burghouts y Geusebroek, 2009; Varma y Zisserman, 2005; Sivic y Zisserman, 2003; Leung y Malik, 2001). La representación holística se ha evaluado en dos aplicaciones similares, en la recuperación y en la clasificación de imágenes, utilizando para ello diversas y variadas bases de datos de imágenes. Para evaluar la eficacia de la representación se ha repetido la experimentación usando conocidos descriptores de textura y color con los que se han comparado los resultados.

En el capítulo 4 sin embargo, se han utilizado los *blobs perceptuales* en la construcción de un segundo modelo de representación con mayor contenido semántico. Éste cumple en gran medida todos los objetivos mencionados al inicio de este apartado y por tanto constituye un modelo de representación adecuado. La viabilidad del descriptor derivado de esta representación se ha evaluado en aplicaciones de recuperación de imágenes que permiten realizar una valoración cuantitativa. Las propiedades perceptuales de la representación se han evaluado de manera cualitativa en una segunda aplicación que ha consistido en el *Browsing* de texturas. Así mismo se ha estudiado la viabilidad de utilizar esta representación para segmentar imágenes con el objetivo de detectar texturas en escenas naturales.

Finalmente en el capítulo 5 se resumen las aportaciones del trabajo y las posibles vías de continuación de éste.

Capítulo 2

Antecedentes

2.1. Introducción

Cualquier descripción computacional de una imagen parte siempre de un conjunto de características locales. Las características locales utilizadas dependen de la tarea a realizar, son de muy diversa tipología y con diversos niveles de complejidad (contornos, blobs, líneas, color y textura entre otros). Una misma característica se puede representar de diferentes maneras, la más común es a través de una medida u operador que proporciona un valor. A la representación de una característica se le ha atribuido el nombre de *descriptor* (Pereira(Ed.), 1999). Por tanto una manera de representar computacionalmente una imagen es mediante el conjunto de descriptores que la caracterizan. Así mismo la adecuada resolución de problemas computacionales pasa siempre por utilizar una buena representación de la imagen.

A lo largo de este capítulo se ofrece una revisión bibliográfica de los principales trabajos que realizan un análisis del color, la textura y la combinación de ambos estímulos visuales desde disciplinas relacionadas como la visión por computador, la ciencia del color y la psicología experimental. Al finalizar el capítulo se analizan los problemas abiertos en el tratamiento combinado del color y la textura que constituyen el punto de partida de este trabajo.

2.2. Color

El color ha demostrado ser una característica importante para la realización de diversas tareas visuales, sobre todo en la identificación y en la discriminación de objetos. Al contrario que en la textura, existe una definición clara del color y se han definido diversas maneras para representar este estímulo visual.

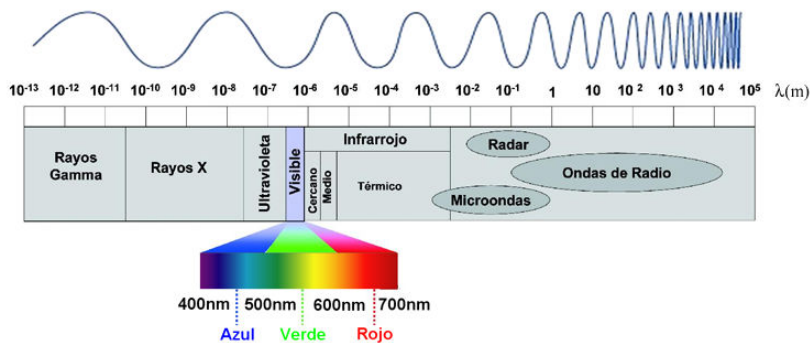


Figura 2.1: Espectro electromagnético.

2.2.1. Física del color

La luz es un flujo de energía compuesto por ondas electromagnéticas de diversas frecuencias. El ojo humano es sensible sólo a un subconjunto de radiaciones electromagnéticas (longitudes de onda) que definen el espectro visual. Como ejemplo en la figura 2.1 se muestra todo el rango del espectro electromagnético.

Cuando la luz incide en un objeto, la superficie de éste absorbe la energía de cada longitud de onda con diferente grado dependiendo de su reflectancia. La energía que no es absorbida por el objeto se refleja de diferente manera dependiendo de las propiedades del material del objeto y del ángulo de incidencia de la luz, todo ello junto con la posición del observador determina la percepción del color del objeto. Dentro de los modelos que realizan una aproximación del modelo real de reflexión se encuentra el *modelo dicromático de reflexión* (Shafer, 1985). Según este modelo, cuando el material de un objeto es ópticamente no homogéneo y opaco, se produce un fenómeno de reflexión especular y un fenómeno de reflexión difusa. La reflexión especular corresponde a la porción de la luz incidente que es reflejada por la superficie del objeto, cuya dirección de reflexión se concentra alrededor de la dirección simétrica a la dirección de la luz incidente con respecto a la superficie normal. Por otro lado el fenómeno de la reflexión difusa se produce debido a la porción de luz que entra en el material. Esta luz es absorbida por las partículas colorantes o pigmentos que se encuentran dentro del material y es re-emitida en parte hacia el exterior en todas direcciones. El color de la luz reflejada por la reflexión difusa es determinado por las longitudes de onda que no son absorbidas por los pigmentos. En la figura 2.2.(a) se encuentra un gráfico mostrando estos fenómenos.

De acuerdo con la geometría mostrada en la figura 2.2.(b), la formulación matemática del modelo dicromático es la siguiente:

$$L(\lambda, i, e, g) = L_i(\lambda, i, e, g) + L_b(\lambda, i, e, g) = m_i(i, e, g) c_i(\lambda) + m_b(i, e, g) c_b(\lambda) \quad (2.1)$$

donde la radiación total de la luz reflejada L es la suma de dos partes independientes: la radiación L_i de la reflexión especular y la radiación L_b provocada por la difusión del material. A su vez

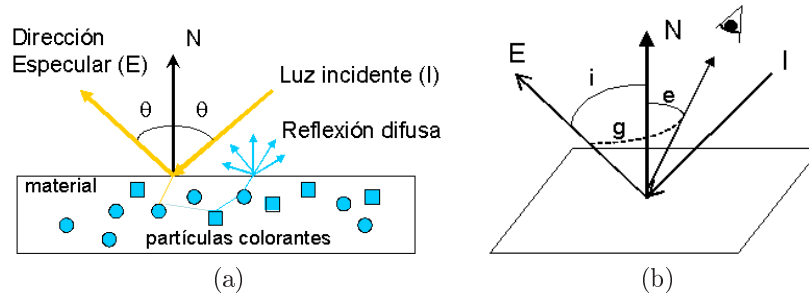


Figura 2.2: (a) Reflexión especular y difusa.(b) Ángulos fotométricos.

cada una de estas componentes se puede dividir en dos partes, la distribución espectral relativa (c_i o c_b) que depende sólo de la longitud de onda y es independiente de la geometría, y la segunda componente es el factor de escala geométrica (m_i o m_b) que depende sólo de la geometría y es independiente de la longitud de onda.

A parte de las restricciones de este modelo comentadas anteriormente, otras asunciones son que la superficie está uniformemente coloreada, sólo hay una fuente de luz (no difusa) y que la función espectral de la iluminación es constante. De otra manera el modelo de reflexión es más complejo.

Para el caso de materiales con reflexiones especulares muy altas (tales como metales) en la fórmula 2.1 sólo se considera el término L_i y en superficies mates (llamadas dieléctricos) el término L_b .

Hay que mencionar que en la mayoría de trabajos dentro del ámbito de la visión por computador, al igual que en éste, se asume que las superficies son Lambertianas. Una superficie Lambertiana produce una reflexión perfectamente difusa e isotrópica (igual en todas las direcciones), esto implica una simplificación del modelo de reflexión dicromático al considerar sólo la componente L_b de este modelo, siendo en este caso la componente m_b independiente del ángulo de visión del observador. Entonces la fórmula 2.1 queda de la siguiente manera:

$$L(\lambda, i) = L_b(\lambda, i) = m_b(i) c_b(\lambda) \quad (2.2)$$

donde el primer término depende exclusivamente de la iluminación y el segundo del objeto.

2.2.2. Percepción humana del color

En la visión humana la percepción del color se produce por etapas, teniendo cada una de ellas una ubicación diferente en el sistema visual. En la primera fase interviene la retina del ojo, en esta zona se encuentran millones de células especializadas de varios tipos, llamadas *fotoreceptores*, que absorben la luz y la convierten en señales neuronales. Estas señales son procesadas por otras células de la retina antes de que pasen al cerebro a través del nervio óptico (en la figura 2.3 se muestra la estructura del ojo). Hay cuatro tipos de fotoreceptores, tres de ellos son los denominados *conos* que son los responsables de la visión diurna y del color mientras que el cuarto tipo son los

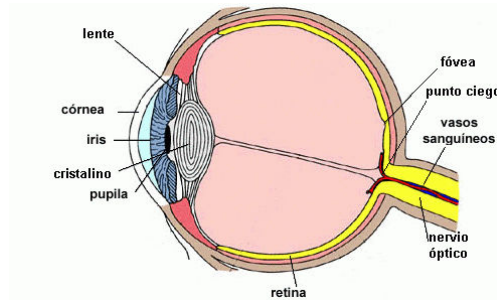


Figura 2.3: Estructura del ojo humano.

bastones. Estos últimos se activan en intensidades bajas de luz (visión nocturna) ya que son mucho más sensibles a la luz que los conos, pero no contribuyen a la visión del color (Robertson, 1992). Según la *teoría tricromática* (Wyszecki y Stiles, 1982) cada fotoreceptor es sensible a (absorbe) un conjunto de longitudes de onda concretas de la luz, de esta manera los conos se diferencian en: conos tipo L (del término inglés *Long Wavelength*), tipo M (*Medium Wavelength*) y tipo S (*Short Wavelength*). En la figura 2.4 se muestra en el espectro visual (rango de frecuencias visible) la sensibilidad de cada uno de los conos, reflejando la proporción de luz que absorben en cada longitud de onda. La señal de la excitación neuronal para cada tipo de fotoreceptor (C_L , C_M y C_S) producida por un estímulo genérico $L(\lambda)$ filtrado por cada sensibilidad espectral $l(\lambda)$, $m(\lambda)$ y $s(\lambda)$ dentro del espectro visible $[\lambda_1.. \lambda_2]$ se representa:

$$C_L = \int_{\lambda_1}^{\lambda_2} L(\lambda)l(\lambda) d\lambda, \quad C_M = \int_{\lambda_1}^{\lambda_2} L(\lambda)m(\lambda) d\lambda, \quad C_S = \int_{\lambda_1}^{\lambda_2} L(\lambda)s(\lambda) d\lambda \quad (2.3)$$

La percepción del color no sólo se determina a partir de factores psicofísicos, según Evans (1974) en la percepción del color están involucrados cinco atributos perceptuales: Tono, Saturación, Luminosidad, Viveza y Brillantez. Aunque para colores no fluorescentes el conjunto de atributos se reduce a los tres primeros. A continuación se definen estos conceptos (Robertson, 1992; Plataniotis y Venetsanopoulos, 2000):

- El tono está asociado con la longitud de onda dominante, por tanto representa el color dominante que es percibido por un observador (ej: rojo, amarillo, verde,...).
- La saturación es la pureza relativa del color. Cuanto menos saturado es un color, más proporción de blanco tiene. Los colores puros están completamente saturados y el grado de saturación disminuye a medida que son menos puros.
- La luminosidad se corresponde con la apreciación subjetiva de claridad y oscuridad (ej: un gris claro tiene más luminosidad que un gris oscuro).

Por otro lado, hay una serie de factores ambientales que repercuten en la percepción del color estos son: el tipo de luz y los colores del entorno. A continuación se explican los efectos que estos producen.

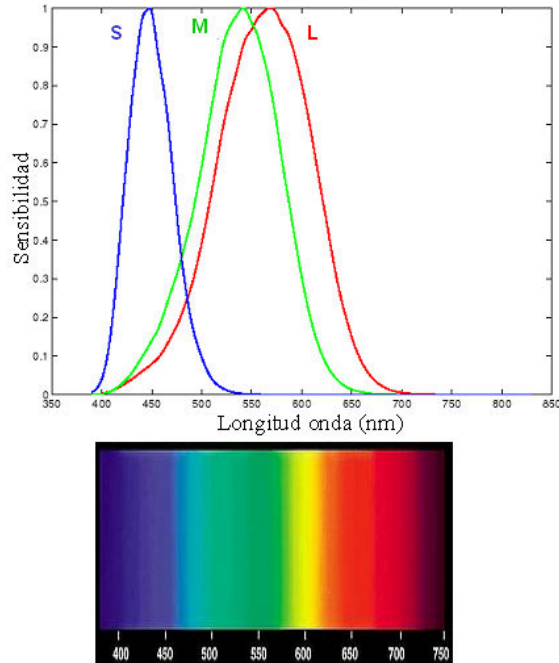


Figura 2.4: Sensibilidad de los conos.

Una misma superficie puede presentar diferentes apariencias bajo diferentes fuentes de luz o bajo diferentes intensidades del mismo iluminante. En la figura 2.5 se muestra un ejemplo de las variaciones que puede sufrir la misma escena cuando las características del iluminante cambian. Sin embargo, un observador humano es capaz de inferir el iluminante y percibir los colores de la escena de manera estable. Esto es debido a que el sistema visual humano posee un mecanismo adaptativo que permite separar las variaciones espectrales de la luz de una escena y preservar el aspecto de un objeto bajo una amplia gama de fuentes de luz. Esta habilidad perceptual se llama *constancia de color*. La constancia del color ha sido modelada en la visión por computador por diferentes métodos que se basan en obtener el iluminante de la escena y proporcionan como resultado la escena bajo un iluminante de referencia, simulando de esta manera el proceso de adaptación al iluminante (Maloney y Wandell, 1986; Forsyth, 1990; Finlayson y cols., 2001; Tous, 2006).

Los colores del entorno también afectan a la percepción humana del color, en la figura 2.6 se muestran unos ejemplos. A este fenómeno se conoce con el nombre de *inducción cromática* e incluye dos efectos diferentes: la *asimilación cromática* y el *contraste cromático*. La asimilación cromática se produce cuando la percepción de la cromacidad de un estímulo cambia hacia la cromacidad del color que le rodea. Esto normalmente ocurre cuando se observa una superficie que posee una elevada frecuencia espacial. Por otro lado, el contraste cromático tiene lugar cuando la cromacidad de un estímulo cambia dependiendo de la cromacidad del entorno. Este efecto aparece cuando se observa una superficie con baja frecuencia espacial. Ambos efectos han sido modelados en la visión por computador (Petrou y cols., 1998; Vanrell y cols., 2004; Otazu y cols., 2008), la



Figura 2.5: Dos imágenes con la misma escena bajo iluminantes diferentes.

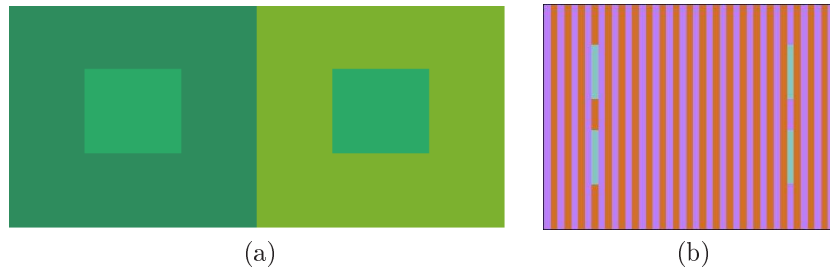


Figura 2.6: Fenómeno de inducción cromática. (a) Constaste cromático: los cuadrados interiores tienen el mismo color pero el color percibido en cada uno es diferente. (b) Asimilación cromática.

asimilación cromática como un efecto de *blurring* o suavizado, mientras que el contraste cromático se ha modelado como una operación de *sharpening*.

En este trabajo se asume que antes de aplicar los descriptores para representar el contenido de las texturas color, las imágenes deben de ser pre-procesadas con los algoritmos que modelan estos fenómenos.

2.2.3. Representación computacional del color

La colorimetría es la ciencia que se ocupa del estudio de como especificar numéricamente el color correspondiente a un estímulo físico y de como medir la diferencia entre colores. Dado que existen tres tipos de fotorreceptores responsables del color, parece adecuado representar al color con tres componentes numéricas. Dentro de la colorimetría éste es el enfoque más extenso y corresponde a la *generalización tricromática* (Wyszecki y Stiles, 1982). Según la definición de cada una de las componentes numéricas se derivan diferentes maneras de representar los colores, dando lugar a *espacios de color* diferentes. Dependiendo de la aplicación es más adecuado usar unos espacios u otros.

Los dispositivos utilizados en las aplicaciones de visión por computador suelen ser cámaras CCD cuyos sensores capturan los colores de una escena. Cada dispositivo tiene sus propios sensores de tres tipos (R, G y B) con sensibilidad específica. Por tanto, la información capturada por estos dispositivos es dependiente del dispositivo y da lugar a que los espacios donde se representan

también lo sean.

En las siguientes secciones se detallan algunos de los espacios de color más utilizados en las aplicaciones de visión por computador.

El espacio RGB

El color se representa con tres componentes: R (rojo), G (verde) y B (azul). El valor de éstas es la suma de sus correspondientes funciones de sensibilidad a la luz y se calculan de la siguiente manera:

$$R = \int_{300}^{830} L(\lambda) \bar{R}(\lambda) d\lambda, \quad G = \int_{300}^{830} L(\lambda) \bar{G}(\lambda) d\lambda, \quad B = \int_{300}^{830} L(\lambda) \bar{B}(\lambda) d\lambda \quad (2.4)$$

siendo $L(\lambda)$ el espectro de la luz capturada y $\bar{R}(\lambda)$, $\bar{G}(\lambda)$ y $\bar{B}(\lambda)$ las funciones de sensibilidad de los sensores. Dado que las componentes calculadas dependen de la función sensibilidad, este espacio es dependiente del dispositivo. Los colores en este espacio se representan dentro de un cubo situado en el espacio tridimensional de ejes R, G y B, en la figura 2.7.(a) se muestra este espacio.

Un factor a tener en cuenta en los dispositivos CCD, que afecta a los valores RGB, es que la sensibilidad de los sensores es logarítmica, por tanto da lugar a una representación no lineal del espacio RGB, conocido como espacio $R'G'B'$. Algunas cámaras disponen de un mecanismo llamado *corrección gamma* para linealizar las respuestas dando lugar al espacio RGB lineal.

Este espacio y todos los espacios derivados de éste, no son perceptualmente uniformes, es decir, las distancias geométricas entre dos puntos no se corresponden con las distancias perceptuales entre los colores correspondientes.

El espacio de colores oponentes

Este espacio fue inspirado en la *teoría de colores oponentes* propuesta por el fisiólogo Hering (1964), en ésta se postula que el sistema visual funciona a partir de un proceso de oposición de colores, existiendo seis colores primarios agrupados en tres pares: rojo-verde ($R-G$), amarillo-azul ($Y-B$) y los acromáticos blanco-negro (I) ya que los colores de cada par no coexisten en ningún color percibido. Se han definido diversas representaciones para el espacio oponente, entre ellas la de Valois y Valois (1993) que la definen a partir de una transformación lineal del espacio RGB:

$$\begin{pmatrix} I \\ RG \\ BY \end{pmatrix} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1 & -1 & 0 \\ -1 & -1 & 2 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (2.5)$$

siendo RG y BY las componentes cromáticas e I la componente acromática.

La principal propiedad de este espacio es la decorrelación que presentan todos sus canales si se compara con el espacio RGB.

La familia de espacios HSI

La familia de espacios HSI se deriva del espacio RGB y se obtiene realizando diversas transformaciones no lineales sobre las coordenadas de éste. La característica diferencial respecto al espacio RGB es que en esta familia de espacios la definición de las dimensiones está relacionada con los atributos perceptuales del color: tono, saturación y luminosidad.

La familia de espacios HSI usa coordenadas cilíndricas donde la componente S es proporcional a la distancia radial y la H es una función del ángulo, ambas por tanto en coordenadas polares. Estas componentes están relacionadas con los conceptos de saturación y tono respectivamente. La componente I es la distancia a lo largo del eje perpendicular al plano de coordenadas polares y está relacionada con el concepto luminosidad.

Estos espacios, a pesar de no ser perceptualmente uniformes, son más adecuados para representar la percepción del color humano que los espacios RGB o el espacio de colores oponentes, ya que son más intuitivos (Plataniotis y Venetsanopoulos, 2000). La separación entre componentes cromáticas (H y S) de la acromática (I) es útil en aplicaciones como la segmentación en donde el tono y la saturación tienen mayor importancia.

Existen diversas maneras computacionales de calcular el espacio HSI a partir de las componentes RGB dando lugar a la familia de espacios HSI. A continuación se muestran tres de ellas.

T. Carron y Lambert (1994) definió el espacio HSI de la siguiente manera:

$$I = \frac{1}{3}(R + G + B)$$

$$S = 1 - \frac{3}{(R + G + B)}[\min(R, G, B)]$$

$$H = \begin{cases} \theta & \text{si } B \leq G \\ 360 - \theta & \text{si } B > G \end{cases}$$

siendo θ

$$\theta = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R - G) + (R - B)]}{\sqrt{(R - G)^2 + (R - B)(G - B)}} \right\}.$$

Antes de efectuar, el cálculo las componentes R, G y B deben estar normalizadas entre [0, 1].

En la figura 2.7.(b) puede verse la distribución de los colores en este espacio.

A. Smith (1978) definió el espacio HSV en donde los límites del espacio se encuentran dentro de un cilindro. La formulación es la siguiente:

$$V = \max i$$

$$S = \frac{\max i - \min i}{\max i}$$

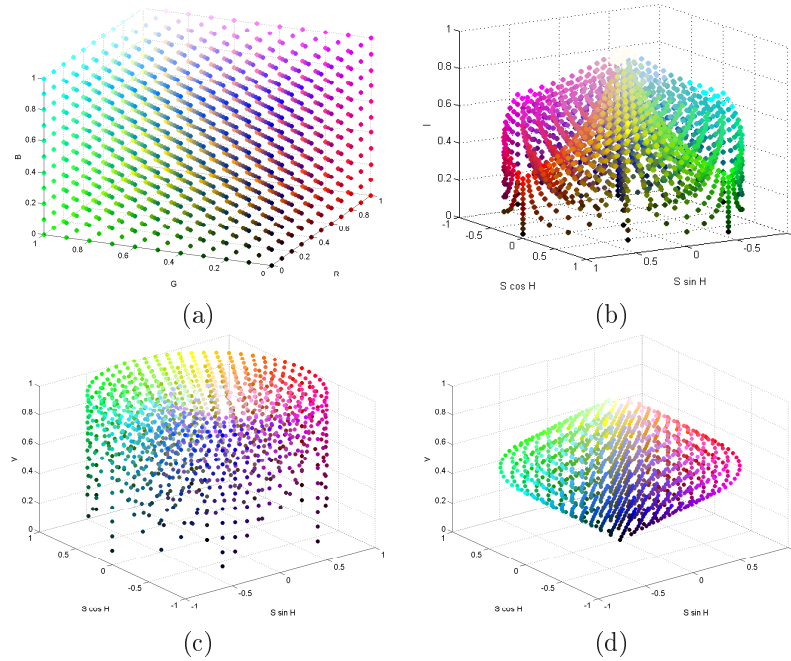


Figura 2.7: (a) Espacio de color RGB. (b) Espacio de color HSI-Carron. (c) Espacio de color HSV-Smith. (d) Espacio de color HSV-Yagi.

$$H = \begin{cases} 60 \times \frac{G-B}{S} & \text{si } \textit{maxi} = R \\ 60 \times (2 + \frac{B-R}{S}) & \text{si } \textit{maxi} = G \\ 60 \times (4 + \frac{R-G}{S}) & \text{si } \textit{maxi} = B \end{cases}$$

siendo

$$\textit{maxi} = \max(R, G, B), \quad \textit{mini} = \min(R, G, B).$$

El espacio HSV definido por Yagi y cols. (1992) puede considerarse el más perceptual de los tres ya que las distancias entre negros y entre blancos se ajusta mejor a la distancia percibida por el sistema visual humano (SVH). En esta nueva formulación los límites del espacio están definidos por un doble cono. La componente Hue se calcula como en el modelo de A. Smith (1978) pero cambia la definición de la saturación y la intensidad (o value):

$$V = \frac{\textit{maxi} + \textit{mini}}{2}$$

$$S = \textit{maxi} - \textit{mini}.$$

En la figura 2.7(c) y (d) se muestran estos dos últimos espacios.

Los espacios CIE

La *CIE* (Comisión Internacional de la Iluminación) con el objetivo de crear un sistema de representación del color estándar y que sirviera como sistema de referencia para la industria creó el espacio de representación *CIE-XYZ*. Este espacio se basa en las medidas de tres funciones de sensibilidad de un observador humano estándar obtenidas experimentalmente que se muestran en la figura 2.8. La integral de cada una de estos estímulos (de manera análoga a la formulación indicada en 2.4) da lugar a las componentes X , Y y Z del nuevo espacio de representación, donde cada tríada de valores representa un solo color visible. Normalizando las variables de este espacio se obtienen las nuevas componentes:

$$\begin{aligned} x &= \frac{X}{X+Y+Z} \\ y &= \frac{Y}{X+Y+Z} \\ z &= \frac{Z}{X+Y+Z}, \quad \text{también puede expresarse como } z = 1 - x - y \end{aligned} \quad (2.6)$$

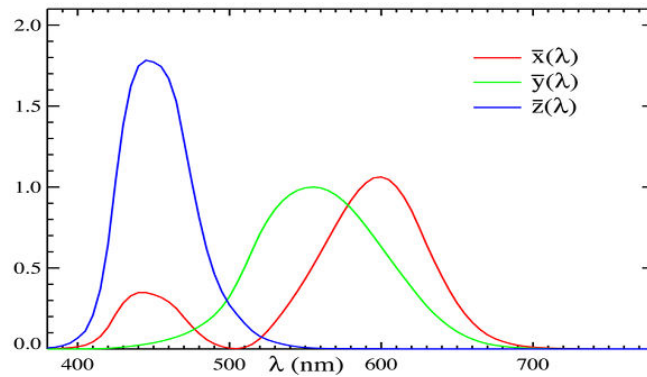


Figura 2.8: Funciones de la correspondencia del color (CMF: *Color Matching Functions*).

La representación de las componentes cromáticas sin tener en cuenta la intensidad se conoce con el nombre de *diagrama cromático* (ver figura 2.9). Este diagrama equivale a proyectar los puntos (X,Y,Z) en el plano $x + y + z = 1$.

Se pueden convertir las coordenadas X , Y y Z a cualquiera de los espacios derivados del RGB y al revés conociendo ciertos valores del dispositivo que ha producido las componentes RGB.

Por definición este espacio es independiente del dispositivo, puesto que es el espacio dependiente del observador estándar, pero continua siendo no perceptualmente uniforme. Por este motivo la *CIE* creó los espacios $L^*u^*v^*$ y $L^*a^*b^*$, en ambos casos la componente L^* es la luminosidad y el resto de componentes son las cromáticas. A continuación se muestra la ecuación que permite calcular las componentes del espacio $L^*a^*b^*$ a partir de los valores X , Y y Z :

$$L^* = \begin{cases} 116 \left(\frac{Y}{Y_0} \right)^{\frac{1}{3}} - 16 & \text{si } \frac{Y}{Y_0} > 0,008856 \\ 903,3 \left(\frac{Y}{Y_0} \right) & \text{si } \frac{Y}{Y_0} \leq 0,008856 \end{cases}$$

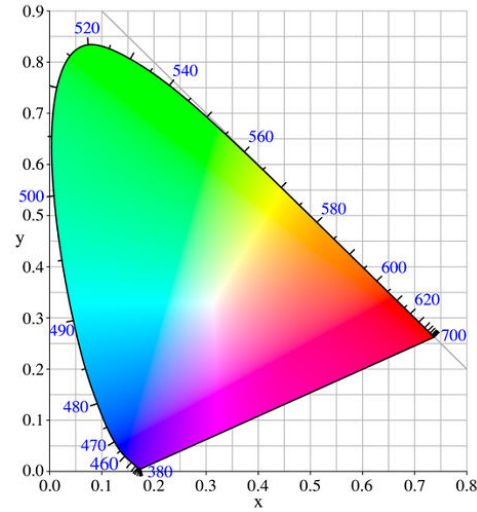


Figura 2.9: Diagrama cromático.

$$a^* = 500 \left[f \left(\frac{X}{X_0} \right) - \left(\frac{Y}{Y_0} \right) \right]$$

$$b^* = 200 \left[f \left(\frac{Y}{Y_0} \right) - \left(\frac{Z}{Z_0} \right) \right]$$

siendo (X_0, Y_0, Z_0) las coordenadas (X, Y, Z) del blanco de referencia y la función f como se define a continuación:

$$f(U) = \begin{cases} U^{\frac{1}{3}} & \text{si } U > 0,008856 \\ 7,787U + 16/116 & \text{si } U \leq 0,008856 \end{cases}$$

$$U(X, Y, Z) = \frac{4X}{X + 15Y + 3Z}.$$

En este espacio de color para obtener el concepto tono se utilizan las componentes a y b de la siguiente manera:

$$\text{tono} = \arctan \left(\frac{b^*}{a^*} \right).$$

Para poder calcular esta transformación se necesita conocer el blanco de referencia¹ y esto sólo es posible si se dispone de dispositivos calibrados.

2.2.4. Nombres de colores

Los humanos para describir un color generalmente utilizan nombres. Estos pueden ser nombres basados en objetos conocidos (chocolate, pistacho, oliva, coral,...) o pueden ser nombres básicos

¹Es el estímulo que el observador percibe como blanco.

aprendidos desde la infancia. A partir de un estudio intercultural de 98 lenguas Berlin y Kay (1969) determinó 11 términos básicos para describir las categorías de colores, estos son: blanco, negro, rojo, verde, amarillo, azul, marrón, gris, naranja, violeta y rosa. Sobre estos se suelen aplicar modificadores como “claro”, “oscuro” o también se pueden combinar de la forma “azul verdoso” (Robertson, 1992). Los experimentos de Berlin y Kay (1969) demostraron que los seres humanos son capaces de seleccionar para cada categoría un ejemplo representativo, pero tienen dificultades a la hora de establecer fronteras entre categorías. Esto llevó a la definición del concepto de *foco*, como el representante de cada categoría y a la hipótesis de que los colores tienen una pertenencia difusa a las categorías.

Si bien se ha demostrado la existencia de 11 colores básicos, el debate está aún abierto y existen trabajos que hablan de 6 colores básicos (rojo, verde, azul, amarillo, blanco, negro) que estarían relacionados con los *unique hues* descritos por Hering.

2.2.5. Descriptores computacionales del color

En este apartado se ha realizado un resumen de los principales descriptores computacionales utilizados en diversas aplicaciones dentro del ámbito de la visión por computador.

Histograma de color Caracteriza la distribución de colores en una imagen contabilizando la frecuencia de aparición de cada color. Para ello previamente hay que cuantificar el espacio donde se representan los colores para determinar los colores a considerar. Dado que los espacios de representación del color son tridimensionales existen dos opciones a la hora de caracterizar el color: mediante un histograma tridimensional o mediante tres histogramas unidimensionales (un histograma para cada dimensión). Jain y Vailaya (1996) utilizan esta última opción realizando la cuantificación del espacio de color *RGB*. Debido a que cada dimensión de los diferentes espacios de color tiene una contribución diferente en la percepción final del color, este método no parece ser igual de efectivo que el uso de histogramas 3D. El uso de histogramas 3D ha tenido una mayor aceptación, desde Flickner y cols. (1995) que lo utilizaron para cuantificar el espacio de color *Munsell* en el sistema *QBIC* para realizar búsquedas en bases de datos de imágenes, hasta Manjunath y cols. (2001) que incluyeron dos descriptores de este tipo dentro del estándar *MPEG-7* (Manjunath y cols., 2003). Dada la repercusión de estos últimos a continuación se resumen las principales características de estos:

SCD (Scalable Color Descriptor) el objetivo de este descriptor es capturar la distribución de colores de la imagen ignorando su disposición espacial, para ello utiliza un histograma 3D de 256 intervalos, estos se obtienen de la cuantización del espacio de color *HSI-Smith* dedicando 16 niveles para cuantizar la componente *H*, 4 niveles para la componente *S* y 4 niveles para la componente *V*.

CSD (Color Structure Descriptor) este descriptor pretende capturar el contenido de color y la estructura de este contenido, para ello contabiliza la frecuencia relativa de “zonas” que

contienen un color determinado (siendo una “zona” un vecindario de 8x8 pixels). Utiliza el espacio de color *HMMD* (HSI-Yagi) para representar los colores.

La problemática asociada al uso de este tipo de descriptores es bien conocida, dependiendo de la cuantificación del espacio de representación del color, puede ocurrir que colores perceptualmente diferentes estén cuantificados en el mismo intervalo o que colores perceptualmente similares estén cuantificados en diferentes intervalos.

Momentos de color Interpretando la distribución de colores en una imagen como una función de distribución de probabilidad, se caracteriza la primera por sus momentos. El descriptor está compuesto por la media (momento de primer orden), la varianza (momento de segundo orden) y la asimetría (momentos de tercer orden) calculados en cada una de las tres dimensiones del espacio de color escogido para representar el color. Es por tanto una manera muy compacta de representar el color de una imagen aunque suponga una pérdida en el poder de discriminación (Long y cols., 2003) y evita los problemas derivados de la cuantificación que tiene el uso de histogramas (Rui y cols., 1999). Stricker y Orengo (1995) utilizaron este descriptor para obtener las características dominantes de la distribución de colores en imágenes, sobre la base del espacio de color *HSV*. Shih y Chen (2002) utilizó este descriptor de manera local ya que previamente divide la imagen en trozos de tamaño fijo y sobre estos calcula el momento de primer orden. En este caso el descriptor está formado sólo por la media obtenida en cada trozo y en cada componente del espacio color *YIQ* (utilizado para codificar imágenes).

Color dominante El objetivo de este descriptor es caracterizar la imagen con los colores que predominan. Dentro del estándar *MPEG-7* (Manjunath y cols., 2001) se hallan dos descriptores de este tipo. El primero llamado *Dominant Color* está compuesto por un vector que contiene para cada color dominante: el porcentaje de pixels de ese color, la varianza de colores dentro del color dominante y la “coherencia espacial”. Ésta última representa la homogeneidad espacial del color dominante. Para encontrar los colores dominantes utiliza un método de “clustering” en el espacio de color *CIE LUV*. El segundo descriptor llamado *CLD* (Color Layout Descriptor) hace un tratamiento más local ya que divide la imagen en 64 bloques de tamaño fijo y en cada bloque se calcula el color representativo utilizando la *DCT* (transformada de Coseno discreta) sobre el espacio color *YCrCb* (utilizado para codificar y comprimir imágenes).

2.3. Textura

La textura es una característica de la mayoría de superficies naturales y es un estímulo visual importante en algunas tareas, al igual que el color, permite reconocer regiones de una imagen, segmentar, clasificar y reconocer objetos o superficies. Por tanto dentro de la visión artificial el análisis de la textura constituye un área de estudio importante.

Al contrario que el color, en la actualidad todavía no existe una definición única del concepto *Textura* y menos aún de un vocabulario estándar adecuado a sus propiedades que sea plenamente aceptado, así como tampoco existe una taxonomía estándar de éstas. Uno de los motivos de la falta de consenso puede ser el que apunta Jain y Farrokhnia (1991) al afirmar que la diversidad de texturas naturales y artificiales hace que sea imposible dar una definición universal. Sin embargo, la mayoría de investigadores consideran la textura como una propiedad inherente a las superficies.

A continuación se muestran algunas definiciones de este concepto:

- Richards y Polit (1974)** *Es un atributo de un campo que no tiene componentes enumerables.*
- Haralick (1979)** *Se puede describir por el número, tipo de sus primitivas (o elementos que contiene) y por la organización espacial de éstas (aleatoria o estructurada).*
- Sklansky (1978)** *Una región de la imagen tiene textura constante si un conjunto de estadísticos locales o otras propiedades locales de la imagen son constantes, varían lentamente, o son aproximadamente periódicas.*
- Vilnrotter y cols. (1986)** *Puede ser descrita como el patrón de la ordenación espacial de diferentes intensidades (o colores).*
- Unser (1986)** *Debe considerarse como una propiedad de la vecindad de un punto de la imagen... es la realización de un proceso estocástico ergódico estacionario bidimensional.*
- Blostein y Ahuja (1989)** *Está compuesta por unidades repetitivas, texels (elementos de textura).*
- Tuceryan y Jain (1998)** *Es una función de la variación espacial de la intensidad de los pixels.*
- Ojala, Pietikäinen, y Mäenpää (2002)** *Es un fenómeno de dos dimensiones que se caracteriza por dos propiedades ortogonales: la estructura espacial (patrón) y contraste (la cantidad de textura local).*

Esta colección de definiciones, algunas de ellas vagas e imprecisas, muestran que la definición de la textura está formulada en función de la aplicación particular. Algunas definiciones están planteadas desde la perspectiva perceptual y otras son totalmente impulsadas por la aplicación en la que se utiliza. Lo que está claro es que la textura es una propiedad de una región de la imagen y no una propiedad de un pixel. La textura puede ser el resultado de las propiedades físicas de una superficie, tales como rugosidad, que además tienen una cualidad táctil asociada, o puede ser el resultado de diferentes reflectancias tales como diferencias de color de una superficie (Tuceryan y Jain, 1998).

El principal problema es que las texturas en el mundo real son a menudo no uniformes debido a variaciones en la orientación de ésta, a la escala espacial y a cambios de la iluminación (Ojala, Pietikäinen, y Mäenpää, 2002). Estas consideraciones han dado lugar en diversos estudios a incorporar la invarianza de estas propiedades. En el presente trabajo se estudian las texturas que son el resultado de una determinada proyección de una superficie con unas condiciones de iluminación y escala fijas.

La textura además de haber sido estudiada en visión por computador también ha sido estudiada en múltiples disciplinas incluyendo la neurociencia y la psicofísica. En estas disciplinas, los estudios de textura se centran en analizar los procesos neuronales que participan en la percepción visual, con el objetivo de comprender los mecanismos que permiten la detección de la textura y su segregación. Dentro del ámbito de la visión por computador muchos de los trabajos se han inspirado en estudios psicofísicos para la realización de diferentes modelos computacionales (en la sección 2.3.2 se hace un revisión de estos), es por este motivo por el que a continuación se presentan algunos de los trabajos realizados dentro de estas disciplinas.

2.3.1. Antecedentes en la Ciencia Cognitiva

La ciencia cognitiva se dedica al estudio interdisciplinario de cómo la información es representada y transformada en el cerebro humano. Ésta abarca disciplinas como la psicología y la neurociencia. Dentro de la neurociencia, que intenta comprender cómo un sistema físico puede recibir, seleccionar y procesar información, se encuentra el trabajo de Neisser (1967). Éste lanzó la hipótesis bajo la cual el procesamiento de la información del sistema visual humano utiliza dos sistemas funcionales distintos: pre-atentivo y atento. El sistema pre-atentivo se realiza a nivel pre-consciente, abarca un gran campo visual y no requiere la ayuda de procesos cognoscitivos ni ningún esfuerzo, ya que se realiza de manera casi instantánea. Sin embargo, el sistema atento es un proceso cognitivo que requiere una atención focal y abarca un menor campo visual. El primero se encargaría de separar o segregar las regiones de la escena en figura y fondo, actuando como guía para el segundo que se encargaría de identificar los objetos.

Por otro lado, los mecanismos pre-ativos que utiliza el sistema visual humano son estudiados dentro de la psicofísica. Ésta es una rama de la psicología en donde se estudia la relación entre la magnitud de un estímulo físico y la respuesta sensorial que este estímulo causa en un observador. El objetivo es medir cuantitativamente las magnitudes perceptuales. En este ámbito de estudio, según Martínez (1995), las diferentes aproximaciones relacionadas con la discriminación de las texturas se pueden clasificar en:

- Modelos estadísticos.
- Modelos basados en la extracción de características.
- Modelos basados en la teoría multicanal.

Dentro de los modelos estadísticos se sitúan los trabajos de Julesz y cols. (1973), que han estudiado cómo el sistema visual humano (SVH) lleva a cabo la discriminación de texturas. Julesz argumenta que la discriminación de las texturas depende por completo de las propiedades estadísticas de la imagen. En concreto, inicialmente propuso la hipótesis bajo la cual pares de texturas con iguales estadísticos de segundo orden² no pueden discriminarse pre-ativamente (sin esfuerzo).

²Estadístico de segundo orden se define como la probabilidad de que pares de puntos (pixels) separados una cierta distancia y en una determinada orientación tengan determinados niveles de gris.

Más tarde, los mismos Julesz y cols. (1978) desmintieron esta conjetura al encontrar contraejemplos.

En los modelos basados en la extracción de características, se engloban los estudios que consideran que el proceso de discriminación pre-atentiva de texturas está basado en las diferencias entre los elementos geométricos simples contenidos en ellas. Siendo estímulos elementales o primitivas de la visión un conjunto de características o rasgos locales. La segunda etapa de los trabajos de Julesz y sus colaboradores (Julesz y Bergen, 1983; Julesz, 1986) se encuentran dentro de esta clasificación, donde postulan la *teoría de los Textons*. En ésta definen las primitivas de la textura bajo el nombre de *Textons* y formulan la hipótesis que la discriminación pre-atentiva de las texturas se explica como resultado de la diferencia en los estadísticos de primer orden³ de los *Textons*.

Treisman y Gelade (1980) desarrollaron la teoría de la *integración de características* en la que determinaron la lista de características reconocidas como primitivas en la segmentación de texturas, exponiendo que la escena visual se codifica a lo largo de un número de dimensiones separables tales como el color, orientación, frecuencia espacial, brillo y dirección del movimiento. La organización de estas primitivas es también una tarea importante en la discriminación de texturas, esto llevó a algunos investigadores (entre otros Beck y cols. (1983)) a concentrarse en las propiedades que conducen a la percepción de la agrupación de las primitivas. Según el modelo que desarrollaron, la discriminación de texturas está basada en la información sobre la pendiente, tamaño, color y contraste de las primitivas de la textura y de sus estructuras emergentes.

Por último, en los modelos basados en la teoría multicanal consideran al sistema visual humano (SVH) no como un analizador de características, sino como un analizador de frecuencias y orientaciones espaciales contenidas en las imágenes. Los autores (Richards y Polit, 1974; Harvey y Gervais, 1978) demostraron la existencia de una relación sistemática entre las frecuencias espaciales contenidas en los estímulos visuales y la similitud que entre ellas percibían los sujetos.

La mayoría de estudios psicofísicos son esencialmente cualitativos ya que su objetivo es determinar qué características perceptuales utiliza el SVH pero no como éstas se pueden obtener. En los estudios comentados, para estudiar los aspectos de la percepción de la textura se han utilizado diferentes tipos de imágenes de texturas sintéticas, esto ha sido motivo de crítica de algunos investigadores, entre ellos Bergen (1991), que argumenta que las diferentes explicaciones a los estímulos estudiados dependen de manera significativa del tipo de textura que han utilizado. Éste considera que las texturas tienen una complejidad con mayor orden de magnitud que las texturas utilizadas en los experimentos psicofísicos.

Otros investigadores sin embargo, han utilizado texturas naturales en su experimentación psicofísica. A continuación se resumen los trabajos de dos de ellos. Ambos han utilizado algunas de las texturas contenidas en el album fotográfico de Brodatz (1966), uno de los más utilizados para el análisis de texturas.

³ Estadístico de primer orden es la probabilidad de que un punto (pixel) tenga un determinado nivel de gris.

Tamura y cols. (1978) centran el objetivo de su experimentación en definir características texturales que se correspondan con la percepción del sistema visual humano. Para ello parten de la especificación de las que consideran características básicas de las texturas, éstas son:

Gruesa versus fina hace referencia al tamaño de los elementos que forman la textura (tamaño del grano).

Alto contraste versus bajo contraste se refiere al rango de los valores de gris de la imagen de textura.

Direccional versus no direccional se refiere tanto a la forma como a la organización de los elementos de la textura.

Forma de línea versus forma de gota (blob) hace referencia a la forma de los elementos que forman la textura.

Regular versus irregular hace referencia a las variaciones en la organización de los elementos de la textura.

Rugosa versus lisa hace referencia a una cualidad táctil.

Como resultado de la experimentación psicofísica para determinar la validez de las características propuestas realizan las siguientes conclusiones:

- Existe ambigüedad entre las características rugosidad-grosor, quizás debida a la ambigüedad en la definición de estas propiedades.
- La direccionalidad está muy relacionada con “forma de línea”.
- Las características que han demostrado ser más significativas son direccionalidad, contraste y especialmente el grosor.

De forma similar Rao y Lohse (1996) en su experimentación intentan identificar las dimensiones del espacio de representación de las texturas. Inicialmente partieron de un espacio de 12 dimensiones, siendo cada una de las dimensiones las siguientes características: contraste, repetitividad, granularidad, aleatoriedad, rugosidad, densidad, direccionalidad, complejidad estructural, grosor, regularidad, orientación y uniformidad. Después de realizar diversos análisis de los datos obtenidos en la experimentación, y con el objetivo de obtener un espacio de representación de baja dimensión, llegaron a la conclusión que la textura se puede representar en un espacio tridimensional. En este espacio cada una de las dimensiones corresponde a varias de las características que utilizaron inicialmente, estas dimensiones son:

1. Repetitiva (correlacionada positivamente con regularidad, uniformidad y negativamente con aleatoriedad) versus no repetitiva.
2. Alto contraste y no direccional versus bajo contraste y direccional.

3. Granular, grosor y baja complejidad versus no granular, fina y alta complejidad.

Los estudios de Tamura y cols. (1978) y Rao y Lohse (1996) parecen coincidir en las tres características básicas para caracterizar las texturas: la regularidad/repetitividad, la rugosidad-grosor/granularidad y el contraste. En otras palabras, la textura se puede caracterizar por los elementos que contiene (granularidad), la organización de estos (regularidad) y por el contraste.

2.3.2. Descriptores computacionales de la textura

Debido a la extensa variedad de texturas naturales y artificiales no existe un único método adecuado para representar todas las texturas. Para modelar tipos particulares de texturas se han utilizado diferentes enfoques. Todos ellos calculan cuantitativamente las características asociadas a propiedades de la textura y, dependiendo de la aplicación, solucionan diferentes problemas. El estudio de las texturas se ha llevado a cabo de forma paralela según varios enfoques, de acuerdo con Shapiro y Stockman (2001) los dos principales son:

Estadístico se asume que la textura es una medida cuantitativa de la disposición de las intensidades y se describe la relación espacial entre pixels individuales.

Estructural se asume que la textura es un conjunto de primitivas (elementos texturales o texels) relacionadas entre ellas de acuerdo con una disposición más o menos regular y se describen las relaciones espaciales entre primitivas tales como puntos (dots) o gotas (blobs).

Estos enfoques tienen que ver con la resolución de la textura, que Azencott y cols. (1990) definieron como “el tamaño del cuadrado más pequeño necesario para caracterizar la textura”. Así el enfoque puramente estadístico es adecuado para texturas de grano fino (microtexturas) mientras que el enfoque puramente estructural es más adecuado para texturas de grano grueso (macrotexturas).

Dentro de los descriptores computacionales desarrollados hasta ahora sólo algunos de ellos se pueden clasificar como puramente estadísticos o estructurales, la mayoría podrían clasificarse dentro de una categoría híbrida ya que independientemente de la resolución de la textura la modelizan estadísticamente o caracterizan sus texels. A continuación se resumen los descriptores computacionales que han tenido mayor trascendencia.

Haralick y cols. (1973), posiblemente inspirados en la experimentación psicofísica de Julesz y cols. (1973), utilizaron la definición de estadísticos de segundo orden (ver 2.3.1) para calcular las matrices de co-ocurrencia, las cuales les permiten evaluar ciertas relaciones espaciales de los niveles de gris en una imagen. A partir de estas matrices y mediante el cálculo de diversos estadísticos (entre ellos entropía, correlación, etc) obtienen hasta 14 características. Estas proporcionan información sobre propiedades como la homogeneidad, contraste y la presencia de alguna organización estructurada. Pero las características no son visualmente obvias (Tamura y cols., 1978), es decir, no se aproximan a la percepción visual. Por otro lado el cálculo de las matrices de co-ocurrencia comporta un gran coste computacional y, además, requiere un conocimiento a priori de la textura para

determinar los parámetros de los que depende (distancia y ángulo). Las matrices de co-ocurrencia han sido utilizadas por Lohmann (1994) para clasificar imágenes de satélite.

Tsuji y Tomita (1973) fueron los pioneros en ofrecer un enfoque estructural puro. En un trabajo más reciente (Tomita y Tsuji, 1990) obtienen los *atomic texture elements* o *texels* de la textura utilizando una técnica de segmentación. Ésta les proporciona regiones homogéneas (con intensidad similar) de las que extraen la forma. Luego analizan la organización espacial de los grupos de texels que tienen propiedades similares (forma e intensidad).

Laws (1980) fue el precursor del enfoque textura-energía, que consiste en aplicar filtros sobre la imagen para detectar diferentes tipos de texturas. El filtrado se realiza en el dominio espacial a través del uso de máscaras de convolución de tamaño fijo. Estas máscaras detectan la presencia de determinados elementos (bordes, puntos (spots), ondas). Pietikäinen y cols. (1983) utilizaron estos operadores para clasificar texturas y Lovell y cols. (1992) los usaron junto a otros tipos de operadores para segmentar imágenes.

Voorhees y Poggio (1987) implementan un detector de los elementos de la textura (blobs, la primitiva definida por Julesz y Bergen (1983)) y utilizan las diferencias de la distribución de los atributos de blobs (largo, ancho y orientación) para segmentar imágenes con texturas naturales. La detección de blobs se realiza en una escala dada y utilizando la función Laplaciana con umbral mayor que cero. El umbral de detección es determinado en función del ruido de la imagen. Parten del supuesto que la textura en escenas naturales puede ser representada utilizando blobs y sus atributos.

Blostein y Ahuja (1989) implementan un detector multiescala de los elementos de la textura (blobs). El método de identificación de blobs simultáneamente detecta la orientación de la superficie, ésta es determinada analizando las variaciones del área de los blobs. Utilizan como detector de blobs la función Laplaciana.

L. Wang y He (1990) lanzan la hipótesis de que una textura se puede descomponer en un conjunto de pequeñas unidades elementales (llamadas *Unidades de Textura*). Cada unidad elemental está formada por 8 elementos con tres posibles valores (0,1,2) obtenidos en una vecindad de 3x3 pixels. Caracterizan una textura por su espectro, definido como la función de distribución de probabilidad de las Unidades de Textura. Representan de esta manera la textura con $3^8 = 6561$ elementos (unidades de textura).

Ojala y cols. (1994) desarrollan una nueva versión de las *Unidades de Textura* (L. Wang y He, 1990) al que bautizan con el nombre de *Local Binary Pattern (LBP)*, ya que en esta ocasión los valores de los 8 elementos del vecindario de 3x3 pixels sólo tienen dos posibles valores. Esto permite reducir el número de unidades elementales a $2^8 = 256$. Caracteriza la textura por la función de distribución de probabilidad del *LBP* (histograma) y del *LBP/C* (aporta una medida de contraste sobre el LBP). Se puede considerar que este descriptor combina los enfoques estructural y el estadístico porque el LBP detecta microtexturas y su distribución es estimada por el histograma. Recientemente Ojala, Pietikäinen, y Mäenpää (2002) han realizado la versión multiescala $LBP_{P,T}^{riu2}$ que es invariante a la rotación, caracterizando también la textura por el histograma de $LBP_{P,T}^{riu2}$.

F. Liu y Picard (1996) realizan la distinción entre texturas estructuradas y aleatorias. Dependiendo de esta clasificación la caracterizan con operadores diferentes. Modelizan la componente estructurada de una textura a partir de su transformada de Fourier y la no estructurada o aleatoria con un modelo autoregresivo multiresolución (*MRSAR*). La función de autocovarianza de la imagen les sirve de test para determinar si la textura es o no estructurada. Han aplicado esta idea en aplicaciones de recuperación y segmentación de imágenes.

Manjunath y Ma (1996) desarrollan un nuevo descriptor multiescala de texturas utilizando filtros de Gabor. Estos filtros pueden considerarse como detectores de líneas y contornos sintonizables según la orientación y la escala. Caracterizan la textura mediante la media y la desviación estándar (estadísticos de primer orden) de cada una de las respuestas de un banco de filtros al ser aplicados a la imagen. La versión ampliada de este descriptor se presentó más tarde (Manjunath y cols., 2001) con el nombre de *HTD (Homogeneous Texture Descriptor)* dentro de la lista de descriptores que componen el *MPEG-7*⁴. Este descriptor es adecuado para representar texturas homogéneas.

Manjunath y cols. (2000) desarrollan el descriptor *PBC* formado por tres características perceptuales: regularidad, direccionalidad y escala. Cada característica es representada mediante valores acotados. Estas características están relacionadas con las dimensiones perceptuales que Rao y Lohse (1996) identificó en su experimentación psicofísica, éstas son la repetitividad, direccionalidad y granularidad. Este descriptor fue incluido dentro del *MPEG-7* con el nombre de *TBC (Texture Browsing Component)* ya que su utilidad se centra en aplicaciones de “browsing”.

Manjunath y cols. (2001) presentan un descriptor adecuado para texturas no homogéneas, el *EHD (Edge Texture Descriptor)*, que captura la distribución espacial de los contornos. Estos se clasifican en 5 categorías: verticales, horizontales, diagonal 45°, diagonal 135° e isotrópicos. El descriptor contabiliza en un histograma la ocurrencia de las 5 categorías de contornos de una imagen, previa partición de ésta en 16 sub-imágenes. Forma parte de los descriptores incluidos en el *MPEG-7*.

Leung y Malik (2001) presentan un modelo estadístico para representar las superficies de materiales. Éste consiste en modelizar las texturas con un histograma de *Textons*. Donde el término *textons* hace referencia a un vector de respuestas de un banco de filtros. Utilizan un algoritmo de clustering para determinar el conjunto de *textons* que compone el vocabulario. Varma y Zisserman (2005) utiliza el mismo modelo estadístico pero usan un banco de filtros diferente con el que consiguen invarianza a la rotación.

Renninger y Malik (2004) proponen un modelo simple de reconocimiento de texturas para la identificación temprana de escenas. Re-formulan el concepto de *Textons* identificándolo con las características obtenidas con un banco de filtros. El vocabulario de *textons* es aprendido en un conjunto de escenas y se modela la textura a partir del histograma de *textons*.

Lazebnik y cols. (2005) desarrollan una representación no densa⁵ de la imagen “invariante” a transformaciones geométricas. Se basa en obtener regiones significativas de la imagen (con esquinas,

⁴ estándar para describir el contenido de datos multimedia que ofrece descriptores para diversos tipos de aplicaciones.

⁵ No todos los pixels de la imagen están representados.

contornos o blobs) y su contenido se caracteriza utilizando el descriptor *RIFT*, que es la versión invariante del descriptor *SIFT* (Lowe, 1999). Este descriptor captura la distribución de las intensidades de los pixels y la distribución de las orientaciones del gradiente. Se representa la imagen con los descriptores representativos, estos son el resultado de aplicar un método de clustering en el espacio de los descriptores. Utilizan la representación para realizar búsquedas de texturas en una base de datos y para clasificar texturas.

Este trabajo parte del realizado por Salvatella y Vanrell (2007), en donde re-definen el concepto blob al englobar dentro de este término los propios blobs de Julesz y Bergen (1983) (elementos quasi redondos) y las barras (elementos alargados). Bajo la misma suposición que Voorhees y Poggio (1987) desarrollan una representación de la textura compuesta por los atributos (area, relación de aspecto, orientación y contraste) de los blobs que contiene, pero utilizando un detector de blobs multiescala. Proponen además un espacio perceptualmente uniforme donde representan los atributos de los blobs.

Pocos descriptores computacionales han sido ajustados mediante experimentos psicofísicos para que su comportamiento sea el mismo que el humano. A continuación se resume el trabajo de uno de ellos.

Benke y cols. (1988) partiendo de los operadores de Laws (1980), ajustan los coeficientes de las máscaras de convolución de manera que proporcionen un resultado en la clasificación de texturas similar al realizado por seres humanos. Para ello plantean un experimento psicofísico donde los individuos deben clasificar un pequeño subconjunto de texturas (del album de (Brodatz, 1966)) de acuerdo con tres criterios: direccional versus no direccional, regular versus aleatorio y forma de línea versus forma de gota. El objetivo es doble, por un lado determinar si los criterios establecidos corresponden a los ejes de representación de las texturas que proponen y segundo, demostrar la posibilidad de ajustar un determinado operador de acuerdo con estos ejes. Los resultados obtenidos les llevan a concluir que los ejes propuestos no son ortogonales ya que existe una correlación elevada entre la dimensión forma de línea y regularidad.

2.4. Textura y Color

Aunque la mayoría de texturas que aparecen en la naturaleza presentan estas dos características simultáneamente, éstas han sido normalmente estudiadas por separado. En los trabajos donde se ha analizado la textura se han limitado a tratar la información de la intensidad y pocos han incorporado la información cromática en los descriptores de texturas. Quizás porque atendiendo a la noción de textura se considera que dos imágenes con los mismos colores pero diferente patrón textural son diferentes texturas de color, pero no está claro si dos imágenes teniendo el mismo patrón textural y diferentes colores son diferentes texturas.

Desde diversas disciplinas, en la neurofísica Yoe y Essen (1988) y en la psicofísica Poirson y Wandell (1996), sugieren la hipótesis que el ser humano procesa de manera separada la componente cromática (color) de la componente acromática (textura). Pocos trabajos se centran en estudiar

como interaccionan estas dos componentes en el sistema visual humano a la hora de realizar determinadas tareas. De entre estos figura Finlayson y Tian (1999) que realizan un experimento psicofísico para conocer como los humanos determinan la similitud entre patrones textiles de colores. A partir de éste concluyen que la interacción entre la textura (patrón) y el color depende fuertemente del observador, para unos la similitud entre los patrones textiles depende de los dos aspectos conjuntamente, sin embargo para otros uno de los dos aspectos tiene mayor importancia. De este hecho deducen que las personas tienen intrínsecamente diferentes nociones de similitud y que ésta depende del dominio de aplicación.

Mojsilovic, Kovacevic, Kall, y cols. (2000) se plantean un objetivo parecido realizando un experimento similar. En éste pretenden determinar qué dimensiones perceptuales y qué reglas se aplican sobre estas dimensiones cuando el ser humano determina la similitud entre patrones textiles de color. Realizando un análisis más profundo que Finlayson y Tian (1999), obtienen que las cuatro dimensiones más importantes son: la ausencia/presencia de colores dominantes, la dirección dominante en la repetición de los elementos estructurales, la regularidad (de la organización de los elementos estructurales) y la pureza del color (hace referencia a la saturación del color). Por otro lado determina que las reglas usadas por los humanos para decidir la similitud entre patrones son las siguientes:

1. Dos texturas con el mismo patrón (igual regularidad y dirección dominante), aunque tengan diferente color, son similares.
2. Dos texturas con las mismas direcciones dominantes o con regularidad similar, aunque tengan diferente color, son similares.
3. Dos texturas con los mismos colores dominantes y la misma dirección dominante son similares.
4. Dos texturas con las mismas distribuciones de colores dominantes se perciben como similares.

A partir de estas conclusiones se puede extraer que, aunque la naturaleza de textura y color son diferentes, en su tratamiento no se pueden separar totalmente.

2.4.1. Descriptores computacionales de la textura en color

Los métodos que analizan la textura en color se pueden clasificar esencialmente en dos. En el primero se procesa separadamente el color y la textura combinando luego la información, mientras que en el segundo se considera el color y la textura como un fenómeno conjunto. En la figura 2.10 se muestran los esquemas correspondientes a estos dos métodos.

- *Métodos que tratan de manera separada la textura y el color:* En estos métodos se utiliza un descriptor de color, obtenido a partir de los tres canales de color, y un descriptor de textura, obtenido al procesar únicamente la intensidad. La combinación entre textura y color se hace a nivel de cálculo de disimilitudes, combinando las distancias obtenidas con cada descriptor por separado o combinando sus vectores de características. Aplicando esta metodología Flickner

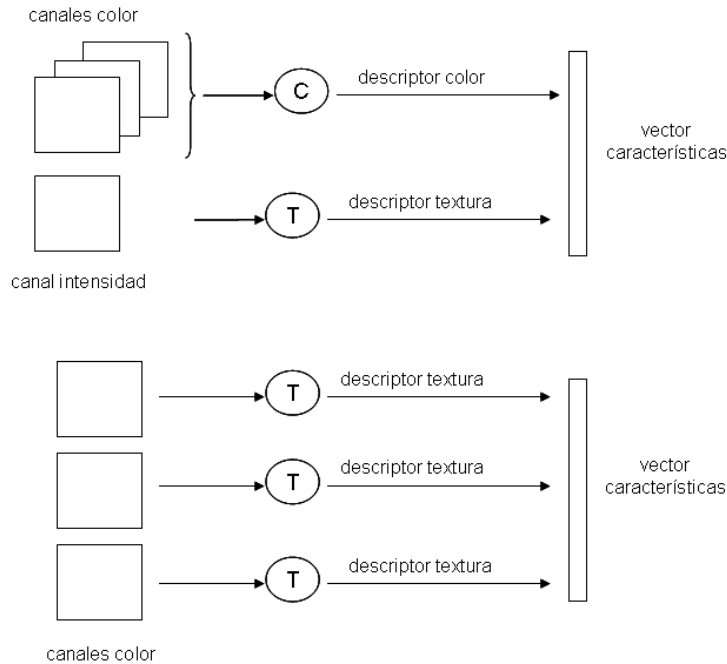


Figura 2.10: Arriba: esquema del modelo de procesamiento del color y la textura de manera separada. Abajo: esquema del modelo de procesamiento conjunto.

y cols. (1995) suman las distancias obtenidas con el descriptor de color (histograma) y con el descriptor de textura basado en el descriptor de Tamura y cols. (1978). P. Liu y cols. (2007) en cambio desarrollan un método automático para determinar los pesos que ponderan las dos distancias en una aplicación de búsqueda de imágenes. El peso depende de un estadístico que evalúa el contenido textural.

Mojsilovic, Kovacevic, Hu, y cols. (2000) no combinan la información de color y textura, sino que utiliza cada una de ellas dependiendo de los requerimientos de la aplicación (búsqueda de patrones texturales en bases de datos). En este trabajo desarrollan dos descriptores (uno para la textura y otro para el color) relacionados con las dimensiones perceptuales que obtuvieron en un estudio previo (Mojsilovic, Kovacevic, Kall, y cols., 2000). El descriptor de color está compuesto por los colores dominantes de la imagen y por su frecuencia de aparición, mientras que el descriptor de la textura, calculado a partir de la *achromatic pattern component*⁶, contiene las medias y desviaciones de la distribución de contornos en diferentes escalas y orientaciones. Con este descriptor estima conjuntamente regularidad y direcciones dominantes.

- *Métodos que tratan conjuntamente la textura y el color:* Se analiza la textura en todas las componentes del espacio de color, y dependiendo del número de procedimientos aplicados en

⁶mezcla de la componente acromática y las componentes acromáticas.

cada componente se diferencian entre dos tipos de enfoques: enfoque unicanal y multicanal. En la mayoría de casos el vector de características resultante se obtiene concatenando los vectores de características obtenidos al aplicar el mismo procedimiento sobre los tres canales del espacio color escogido.

Dentro del enfoque unicanal, Yu y cols. (2003) sobre cada una de las tres componentes del espacio HSV aplican un filtrado con un conjunto de 8 máscaras y calculan los momentos de color (media y varianza), el vector de características resultante tiene $16 \times 3 = 48$ dimensiones. Mäenpää y Pietikäinen (2004) aplican descriptor LBP sobre cada uno de los canales del espacio de color, siendo el vector de características resultante el formado por la concatenación de los histogramas obtenidos en cada canal. Zhong y Jain (2000) utilizan la DCT sobre los 3 canales del espacio color $YCrCb$ para obtener el descriptor de textura y el descriptor de color (es el CLD). En este caso, se combinan ambos descriptores en el nivel del cálculo de la disimilitud, se suman las distancias obtenidas con cada uno de los descriptores. Huang y cols. (1999) desarrollaron el *Correlograma* como una extensión a color de las *matrices de co-ocurrencia* desarrolladas por Haralick y cols. (1973) para caracterizar las texturas. En este caso con el correlograma se contabiliza la probabilidad de que dos colores se encuentren a una cierta distancia. Al igual que en los histogramas se requiere cuantificar el espacio de color donde se representan los colores para determinar el número de colores a considerar. En este descriptor se utilizó una cuantificación de 64 colores sobre el espacio de color RGB .

Dentro del enfoque multicanal, Paschos (2001) y Hoang y cols. (2005) aplican un banco de filtros multiescala de Gabor sobre cada una de las tres componentes del espacio de color para obtener un vector de características. La diferencia es que Hoang y cols. (2005) utilizan el espacio de colores oponente y obtiene un vector de $20 \times 3 = 60$ dimensiones, mientras que Paschos (2001) usan diversos espacios de color $L^*a^*b^*$, RGB , HSV y obtienen un vector de $12 \times 3 = 36$ dimensiones. Burghouts y Geusebroek (2009) utilizan un banco de filtros (variantes del MR8 Varma y Zisserman (2005)) que aplican sobre cada uno de los canales del espacio de colores Oponente y obtienen un vector de $8 \times 3 = 24$ dimensiones. A diferencia del resto de trabajos dentro del enfoque multicanal, Burghouts y Geusebroek (2009) diferencian los vectores de características de cada canal a la hora de utilizarlos en una aplicación de clasificación de materiales. La integración de la información obtenida en cada canal la realizan en el nivel del clasificador.

Drimbarean y Whelan (2001) y Mäenpää y Pietikäinen (2004) evalúan empíricamente los dos métodos en un experimento de clasificación. Los primeros concluyen que con el método conjunto se obtienen mejores resultados que con el método paralelo mientras que Mäenpää y Pietikäinen (2004) obtienen resultados muy similares en los dos métodos. Hay que comentar que los procedimientos que utilizan ambos son diferentes y que en la evaluación, mientras Drimbarean y Whelan (2001) utilizan únicamente un pequeño subconjunto de 16 texturas de la BD Vistex ⁷ Mäenpää y Pietikäinen (2004) lo hacen con 54 texturas de esta BD y con una BD creada por ellos mismos con

⁷Vision Texture-database (MIT Media Lab).

68 texturas.

A tenor de estos estudios no parece claro qué método es el mejor, ni siquiera que el utilizar la información de textura y el color siempre ofrezca mejores resultados que la obtenida utilizando sólo una de estas características (Mäenpää y Pietikäinen (2004) en algunos casos obtienen mayor discriminación utilizando sólo la información de color). Las conclusiones de estos estudios son altamente dependientes del conjunto de imágenes de test y de los descriptores que han utilizado, sería necesario realizar un estudio mucho más extenso para que estas conclusiones se pudieran extender a todas las texturas.

2.5. Problemas abiertos

Como se ha visto en la revisión histórica tanto de la textura como del color, ha existido una tendencia a desarrollar descriptores que extraen información “relacionada” de alguna manera con la que utiliza el sistema visual humano, de ahí que muchos estén inspirados en los resultados de la psicología experimental que se ha hecho al respecto. Lo cierto es que todavía no se ha resuelto el problema de como representar las texturas en color dado que existen grandes lagunas en el conocimiento de los niveles de procesamiento de la información en el sistema visual humano.

Hasta ahora se han desarrollado pocos métodos computacionales que combinen la información propiamente textural y el color, constituyendo éste todavía un problema abierto. El hecho que estas características posean diferentes propiedades, mientras el color es una propiedad de un punto de la imagen la textura es una propiedad de una región, quizás es uno de los motivos por los cuales todavía no está claro cual es la mejor manera de fusionar ambas características.

Con respecto a las diferentes maneras de combinar la textura y el color, hasta ahora no hay un enfoque definitivo y algunos de los enfoques presentan ciertos problemas: cuando la fusión de ambas características se realiza en el nivel del vector de características, concatenando los diferentes vectores en un único vector, es necesario escalar de manera adecuada los valores del vector de características y si los vectores a concatenar tienen diferente dimensión deben aplicarse pesos a cada uno. Por otro lado, si la fusión se realiza a la hora de calcular la disimilitud, esto requiere definir una medida de disimilitud para cada descriptor y escalar las medidas para que ambas sean comparables. En las aplicaciones de clasificación estos problemas se pueden evitar pasando el problema al clasificador, siendo la clasificación final el resultado de combinar las clasificaciones obtenidas con cada uno de los descriptores. En este caso el problema es definir el criterio del clasificador final.

Actualmente, en la mayoría de estudios centrados en el desarrollo de descriptores computacionales para la extracción de características, el objetivo principal es obtener mejoras cuantitativas en las medidas utilizadas para evaluar las diferentes tareas (clasificación, recuperación de imágenes, segmentación). Por ello, la batalla está centrada en la obtención de descriptores que ofrezcan mejores resultados cuantitativos y que sean invariantes al mayor número de aspectos posible (a transformaciones afines como la escala, la inclinación de la cámara, la perspectiva y a cambios de

iluminación en la escena). A todo esto, el aspecto cualitativo (relacionado con la percepción) ha sido el gran abandonado, seguramente motivado por la falta de procedimientos adecuados, sobretodo cuando el número de imágenes a procesar es elevado.

Por otro lado, mientras en el ámbito del color los descriptores pueden ser traducidos a palabras, mediante la técnica de *Color Naming*, aspecto que facilita la compresión de la información obtenida con el descriptor, en el ámbito de las texturas los descriptores suelen carecer de contenido semántico y los valores de los vectores de características no son mapeados a palabras. Aunque existen excepciones, el descriptor *PBC* (Manjunath y cols., 2000) cuyo vector de características tiene una traducción a los términos regularidad, direccionalidad y escala.

En este trabajo se abordan los problemas planteados en el desarrollo de descriptores de texturas en color, dejando de lado las corrientes investigadoras en vigor y partiendo de la psicología experimental, con el objetivo de obtener descriptores perceptuales que permitan una mayor comprensión de la información que estos extraen de la imagen, a la vez que combinan la información textural y de color.

Capítulo 3

Representación holística de texturas en color¹

3.1. Introducción

El modelo computacional para representar la texturas en color que se presenta en este trabajo está inspirado en el modelo perceptual de Julesz (1975) que se ha introducido en el capítulo anterior (2.3.1). Así pues, en este trabajo se propone una implementación de la definición original de los *Textons*. Según Julesz y Bergen (1983) los atributos de los blobs (*Textons*) forman parte de las características locales utilizadas por el SVH para percibir y discriminar las texturas. A nivel pre-atentivo, son las diferencias en los estadísticos de primer orden las responsables de la discriminación de las texturas.

Partiendo de esta hipótesis, en este capítulo se propone una representación de texturas en color basada en el modelo perceptual de los *textons* cuya definición es similar a la aproximación *Bag-of-words* (*BoW*), utilizada mayoritariamente en aplicaciones de categorización de escenas y objetos. Este modelo, de acuerdo con la segunda heurística propuesta por Julesz (apartado 3.2), ignora las relaciones de posición entre los elementos que constituyen las texturas.

Se comienza este capítulo resumiendo la Teoría de los *Textons*² en donde se define este concepto, se presenta su implementación computacional y se detalla la construcción del modelo holístico sobre el que, a través de diversos experimentos, se demuestra su viabilidad como modelo de representación general de texturas color.

Utilizando como características de partida los *blobs perceptuales* y sus atributos, se han investigado diversas maneras para representar estas características y diferentes formas de construir el vocabulario.

¹Contenido parcialmente publicado en Alvarez, Salvatella, Vanrell, y Otazu (2010b).

²Término en inglés que a lo largo del documento no se ha traducido y por ello se referencia en cursiva.

3.2. Teoría de los *textons*

La *teoría de los Textons* fue desarrollada por Julesz y Bergen (1983). En ésta se definen los *Textons* como las características locales sobresalientes responsables de la percepción y discriminación pre-atentiva de las texturas, son por tanto las unidades básicas de las texturas.

La teoría se plantea en base a las dos siguientes heurísticas:

- Heurística 1: *Textons* son:
 - blobs alargados (elipses, segmentos de línea) con propiedades específicas (color, orientación, anchura y longitud).
 - terminaciones (finales de línea).
 - cruces de segmentos de línea.

- Heurística 2: La visión pre-atentiva dirige a la visión atenta hacia el lugar donde existen diferencias en la densidad (número) de *textons*, pero ignora las relaciones de posición entre *textons*.

En esta teoría se afirma que en el sistema visual humano la discriminación pre-atentiva de las texturas está basada en las diferencias de *textons* (características locales) o en la diferencia en los estadísticos de primer orden de los *textons* (características globales).

Pocos años más tarde Julesz (1986) revisa esta teoría heurística y afirma que en la fase pre-atentiva del proceso visual el sistema pre-consciente sabe dónde se producen los gradientes de *textons* (diferencias de *textons* o diferencias de su densidad), pero no tiene conocimiento de lo que son los *textons*. Es en el nivel consciente, por tanto en la fase atenta, cuando los reconoce.

Una de las críticas de esta teoría es que en el estudio se han utilizado texturas sintéticas muy concretas y que por tanto es un estudio parcial. Se cuestiona su aplicabilidad a texturas naturales ya que éstas tienen una complejidad con un orden de magnitud mayor. Además, no se propone como obtener los *textons*.

Un intento de implementación computacional de los *textons* la llevaron a cabo Voorhees y Poggio (1988). Estos detectaban blobs y definieron un estadístico de primer orden utilizando los atributos de los blobs para determinar las fronteras entre texturas. Más tarde Leung y Malik (2001); Varma y Zisserman (2005) retomaron esta teoría y proponen un modelo holístico de representación. Pero estos últimos, al igual que Burghouts y Geusebroek (2009), parten de una definición de *textons* diferente. Consideran que los *textons* son las respuestas de un banco de filtros aplicado sobre la imagen.

En este trabajo se parte de la definición original de los *textons* y se propone una representación perceptual de las texturas color basada únicamente en los blobs de las imágenes y sus atributos (color, orientación, anchura y longitud).

3.3. Representaciones holísticas y aproximación BoW

El término holístico es usado en la psicología cognitiva para definir una de las formas en que se produce el aprendizaje, significa una concepción basada en la integración total frente a un concepto. Expresa la idea de que todas las propiedades de un sistema dado no pueden ser determinadas o explicadas por las partes que lo componen por sí solas. El sistema como un todo determina cómo se comportan las partes.

La implementación computacional directa de los *Textons* conduce a un modelo estadístico en donde las texturas se representan por el número de ocurrencias de los atributos de los *blobs perceptuales*. Renninger y Malik (2004) fueron unos de los primeros en utilizar el concepto holístico en el ámbito de la visión computacional para identificar una manera de representar las texturas. El modelo estadístico propuesto por Renninger y Malik (2004) coincide con la aproximación *bag-of-words* (*BoW*) que recientemente ha sido aplicada con éxito en aplicaciones de categorización de objetos (Sivic y cols., 2005; G. Wang y cols., 2006). Este modelo también ha sido utilizado en otras aplicaciones, Leung y Malik (2001) y Burghouts y Geusebroek (2009) lo emplean en el reconocimiento de materiales, Sivic y Zisserman (2003) en la recuperación de objetos y Varma y Zisserman (2005) para clasificar texturas.

La construcción del modelo de representación de la imagen en enfoques *BoW* y dentro del contexto de la categorización de objetos está formada por tres etapas principales:

1. Detección de características.
2. Representación de las características.
3. Construcción del vocabulario.

En la primera etapa se obtienen las regiones de la imagen que son significativas (*patches*). En la segunda se representan las características de las regiones con un vector numérico para obtener así su descriptor. La tercera etapa consiste en la construcción del vocabulario en donde se requiere el uso de técnicas de aprendizaje, esto es, a partir de un subconjunto de imágenes se encuentran los valores de las características que son representativos. Los representantes se definen como *visual words* (Sivic y Zisserman, 2003) que constituyen los componentes del vocabulario. El método más extendido para obtener los representantes es el algoritmo de clustering *K-means*. Finalmente la imagen es representada por el histograma de las *visual words*. Una característica a destacar de este modelo es que no importa donde están localizadas las características sino su frecuencia.

En la representación holística que aquí se presenta se utiliza la aproximación *BoW* para representar las texturas color, en donde las características utilizadas son los atributos de los blobs perceptuales y el modelo de construcción del vocabulario es menos costoso. El vocabulario se obtiene realizando una cuantificación de los atributos de los blobs perceptuales, por tanto sin aprendizaje previo. La imagen es representada por la función de distribución de probabilidad de los términos o *palabras visuales*³ del vocabulario. Como se verá a continuación la representación holística es

³traducción del inglés *visual words*.

simple pero se ajusta a la propuesta de Julesz y Bergen (1983) en donde afirman que “el sistema visual pre-atentivo utiliza globalmente sólo los *textons* de la manera más simple posible, contando su número (densidad)”.

3.4. Descomposición de la imagen en blobs

Las características utilizadas en esta aproximación *BoW* son los blobs de la imagen. En este apartado se detalla el procedimiento computacional para obtener los blobs percibidos en una textura color además de sus atributos. Definimos un blob como una región convexa homogénea en su contraste o color.

3.4.1. Blobs perceptuales

Partiendo de la asunción que todos los blobs de las imágenes tienen forma Gaussiana, estos pueden ser isotrópicos (forma circular) y no isotrópicos (forma elipsoide). Los blobs se caracterizan por su forma (anchura, altura, orientación) y color.

Para detectar los blobs perceptuales y obtener los atributos que definen su forma se ha utilizado el procedimiento desarrollado por Salvatella y Vanrell (2007) para texturas monocromas. En éste se usa la representación de la imagen en el espacio-escala propuesto por Lindeberg (1994). En el espacio multiescala, la imagen I se representa a partir de una familia de imágenes con diferentes niveles de detalle. Éstas se obtienen al convolucionar la imagen con una función de suavizado con diferentes parámetros σ :

$$L(x, y; \sigma) = I(x, y) * G(x, y; \sigma) \quad (3.1)$$

siendo $G(\cdot; \sigma)$ la función de suavizado Gaussiana:

$$G(x, y; \sigma) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{x^2 + y^2}{2\sigma^2}\right\}. \quad (3.2)$$

Los blobs de una imagen se obtienen aplicando el operador diferencial Laplaciano de una Gaussiana (LoG) normalizado en un subconjunto de escalas. Este operador se define de la siguiente manera:

$$\nabla_{norm}^2 L(\cdot; \sigma) = \sigma^2 \nabla^2 L(\cdot; \sigma) \quad (3.3)$$

siendo el conjunto de escalas utilizado $\sigma = \{1, 284^n, n \in [1 \dots 11]\}$.

Para desestimar los blobs de la imagen debidos al ruido se impone la restricción:

$$\nabla_{norm}^2 L(\cdot; \sigma) \geq \eta_{det} \quad (3.4)$$

donde η_{det} vendrá determinada por el nivel de ruido de la imagen y supone el único parámetro necesario para la detección de los blobs.

La utilización de este banco de filtros permite obtener la escala y la localización espacial de los blobs: los centros de los blobs (x, y) corresponden a la localización de los máximos de la función $\nabla_{norm}^2 L(\cdot; \sigma)$ a lo largo de todas las escalas, mientras que la escala del blob corresponde a la escala en la cual se ha obtenido el máximo, que denotaremos como s_{LoG} .

La orientación y las dimensiones de los blobs se obtienen aplicando el operador tensor estructural 2D, éste es la matriz de momentos de segundo orden, sobre la imagen. Este operador evalúa la estructura de los niveles de gris alrededor de cada punto de la imagen y se define de la siguiente manera:

$$\mu_L(\cdot; t, s) = G(\cdot; s) * ((\nabla L)(\cdot; t)(\nabla L)(\cdot; t)^T) \quad (3.5)$$

siendo ∇L el gradiente de la imagen evaluado para $t = s_{LoG}$ (escala del blob) y G una función gaussiana con $s = \gamma s_{LoG}$ que constituye la escala de integración del tensor estructural. Típicamente el valor de γ es 2.

A partir de la matriz de momentos se obtiene la orientación del blob (θ) de la siguiente manera:

$$\theta = \arctan(v_2/u_2) \quad (3.6)$$

siendo (v_2, u_2) las coordenadas del vector propio de menor valor propio de la matriz de momentos.

Respecto a las dimensiones del blob, la anchura (w) corresponde a la escala en la que el filtro se ajusta mejor a esta dimensión, y la longitud del blob (l) se calcula a partir de λ_1 y λ_2 que son los valores propios de la matriz de momentos en orden decreciente:

$$\begin{aligned} w &= s_{LoG} \\ l &= \sqrt{\lambda_1/\lambda_2} \cdot w = \sqrt{\lambda_1/\lambda_2} \cdot s_{LoG}. \end{aligned} \quad (3.7)$$

Por tanto a partir del procedimiento descrito se han localizado los blobs en la imagen, con sus coordenadas (x, y) , y se han obtenido los atributos que definen su forma: ancho, largo y orientación (w, l, θ) .

Para ilustrar el resultado de la detección de blobs, en la figura 3.1 se muestran algunos ejemplos sobre imágenes monocromas. Para una mejor visualización del resultado en imágenes diferentes se muestran los blobs con contraste positivo y los blobs con contraste negativo.

Respecto a la detección de la forma de los blobs, ésta no es exacta y cuando estos son alargados el error cometido es mayor. En la figura 3.1.(f) donde se muestra un ejemplo de detección de blobs alargados se puede observar, en la imagen de la derecha, que un blob alargado se compone por varios blobs pero ninguno de ellos se ajusta correctamente a la forma. Este error es provocado por dos motivos, el primero es debido al suavizado que provoca la Gaussiana (ecuación 3.2) utilizada para estimar el ancho del blob (w), y el segundo a la hora de calcular sus dimensiones con el tensor estructural (ecuación 3.5). El parámetro de éste (s), que es la escala de integración del tensor estructural, para escalas grandes produce un excesivo suavizado de la imagen y esto provoca el error en la estimación de la longitud del blob (l). La inexactitud en la detección de la forma de

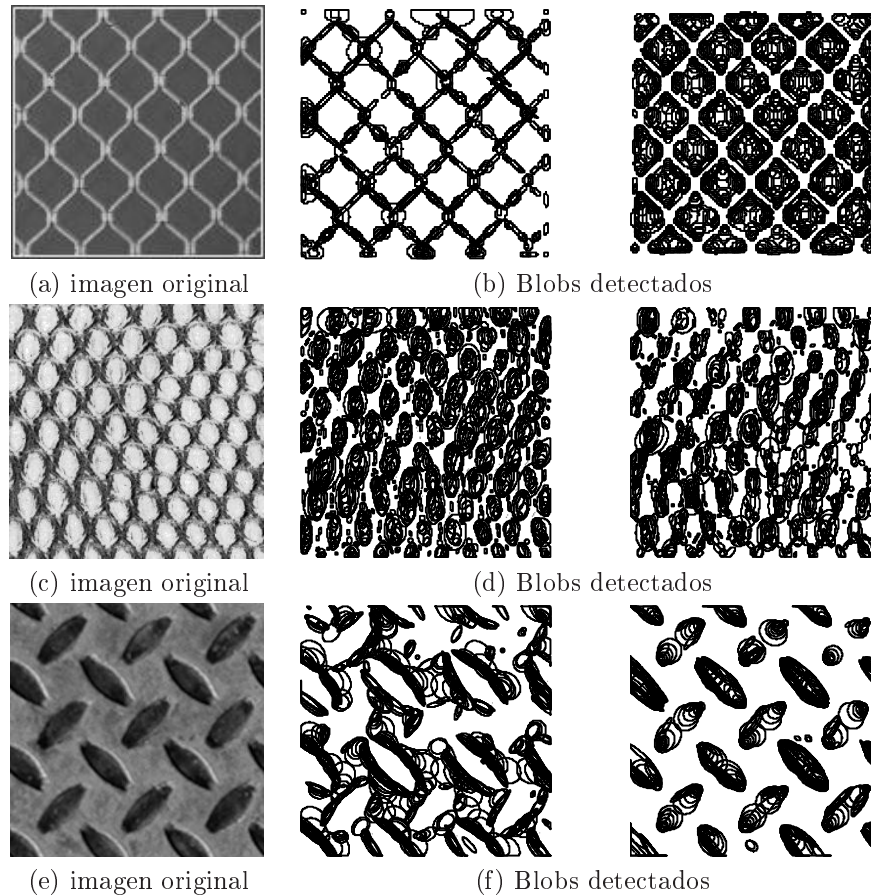


Figura 3.1: (a), (c) y (e) Imágenes monocromas. (b), (d) y (f) Blobs detectados, con contraste positivo a la izquierda y con contraste negativo a la derecha.

los blobs va a provocar que la representación que se derive no sea del todo fiel al contenido de la imagen.

Por otro lado, en las mismas figuras se puede observar que la detección de blobs genera multitud de blob redundantes, es decir, blobs con formas similares que se solapan en el espacio (x, y) y solamente algunos de ellos corresponden a los realmente percibidos. Cada uno de las formas que aparecen en las imágenes deberían estar englobadas por un único blob, de esta manera sí que se corresponderían a la percepción humana. Con el objetivo de eliminar los blobs redundantes y obtener solamente los realmente percibidos se ha desarrollado un *filtrado perceptual* que se desglosa en los siguientes pasos:

1. Reconstrucción de los blobs detectados a partir de su localización y sus atributos.
2. Localización de blobs solapados en el espacio.
3. Eliminación de blobs redundantes. De entre los blobs solapados mantener el que tiene mayor

energía (mayor valor de la función $\nabla_{norm}^2 L(\cdot; \sigma)$ en la escala σ donde se ha detectado) y filtrar el resto.

El resultado de este filtrado se puede observar en la figura 3.2 en donde los blobs perceptuales obtenidos ahora se ajustan en mayor medida a los blobs realmente percibidos.

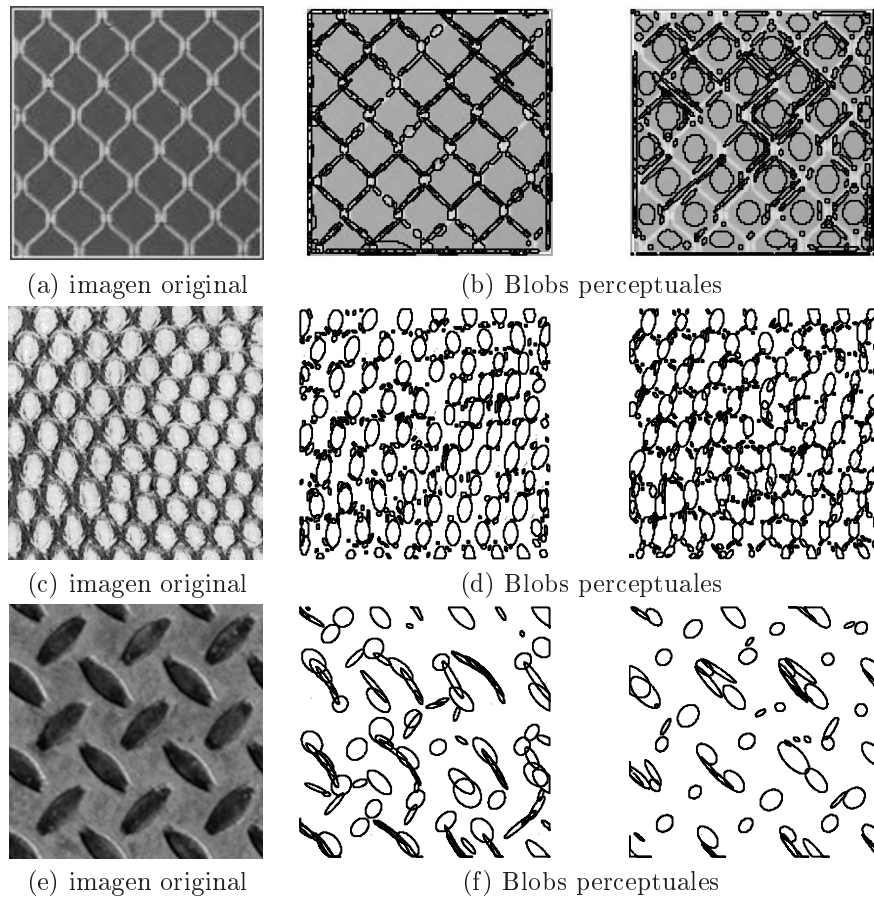


Figura 3.2: (a), (c) y (e) Imágenes monocromas. (b), (d) y (f) Blobs perceptuales, con contraste positivo a la izquierda y con contraste negativo a la derecha.

3.4.2. Blobs en color

En una imagen color los blobs percibidos pueden emerger debido a: diferencias en la intensidad (provocadas por la geometría de la superficie, ej: rugosidad), diferencias de las propiedades cromáticas (provocadas por la reflectancia) o por diferencias de ambas propiedades. En la figura 3.3.(a)-(c) se muestran tres ejemplos de texturas color en donde la percepción de la textura es provocada por diferentes propiedades de la escena.

Teniendo en cuenta estas propiedades, para extender el concepto de blob perceptual en texturas color es necesario utilizar una representación del espacio de color que separe las dimensiones

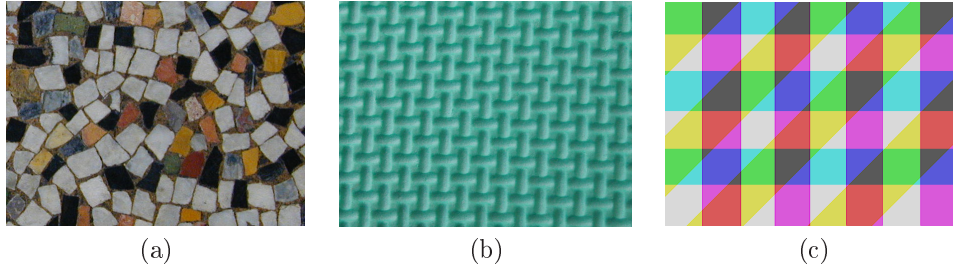


Figura 3.3: (a) y (c) textura percibida por las diferencias de las componentes cromáticas. (b) textura percibida por los cambios de la intensidad.

cromáticas de la intensidad. El espacio de colores oponentes ha sido él escogido por este motivo. En la figura 3.4 se muestra la descomposición de las imágenes de la figura 3.3 en los canales del espacio RGB y en los canales oponentes. En la figura 3.4 se demuestra como los blobs producidos por las propiedades cromáticas de la imagen 3.3.(c) no aparecen en la clásica descomposición RGB mientras que en los canales oponentes sí.

La componente intensidad del espacio oponente ha sido definida de acuerdo con la formulación expresada en el apartado 2.2.3 ($I = R + G + B$), y las dimensiones cromáticas rg y by en esta ocasión han sido normalizadas para hacerlas independientes de la intensidad. Éstas se obtienen a partir de la siguiente transformación:

$$\begin{pmatrix} rg \\ by \end{pmatrix} = \begin{pmatrix} 1/2 & -1/2 & 0 \\ 1/3 & 1/3 & -2/3 \end{pmatrix} \cdot \begin{pmatrix} r \\ g \\ b \end{pmatrix} + \begin{pmatrix} 1/2 \\ 2/3 \end{pmatrix} \quad (3.8)$$

donde r , g y b son las componentes cromáticas obtenidas normalizando por la intensidad, esto es:

$$\begin{aligned} r &= R/I \\ g &= G/I \\ b &= B/I \end{aligned} \quad (3.9)$$

y donde R , G y B son respectivamente las componentes rojo, verde y azul del espacio RGB de la imagen original. En esta transformación se halla incluida una corrección para que los valores rg y by estén en el rango $[0..1]$ al igual que la componente I .

Los blobs perceptuales de una imagen color se obtienen aplicando el operador diferencial (ecuación 3.3) y el filtrado perceptual sobre cada una de las componentes del espacio de colores oponentes. Este procedimiento además de detectar los blobs debidos a cambios en la intensidad de la imagen, también permite detectar los blobs que presentan un bajo contraste en el canal intensidad y que se discriminan mejor en cualquiera de los dos canales cromáticos. De esta manera, se obtiene una descomposición completa de una imagen en sus blobs.

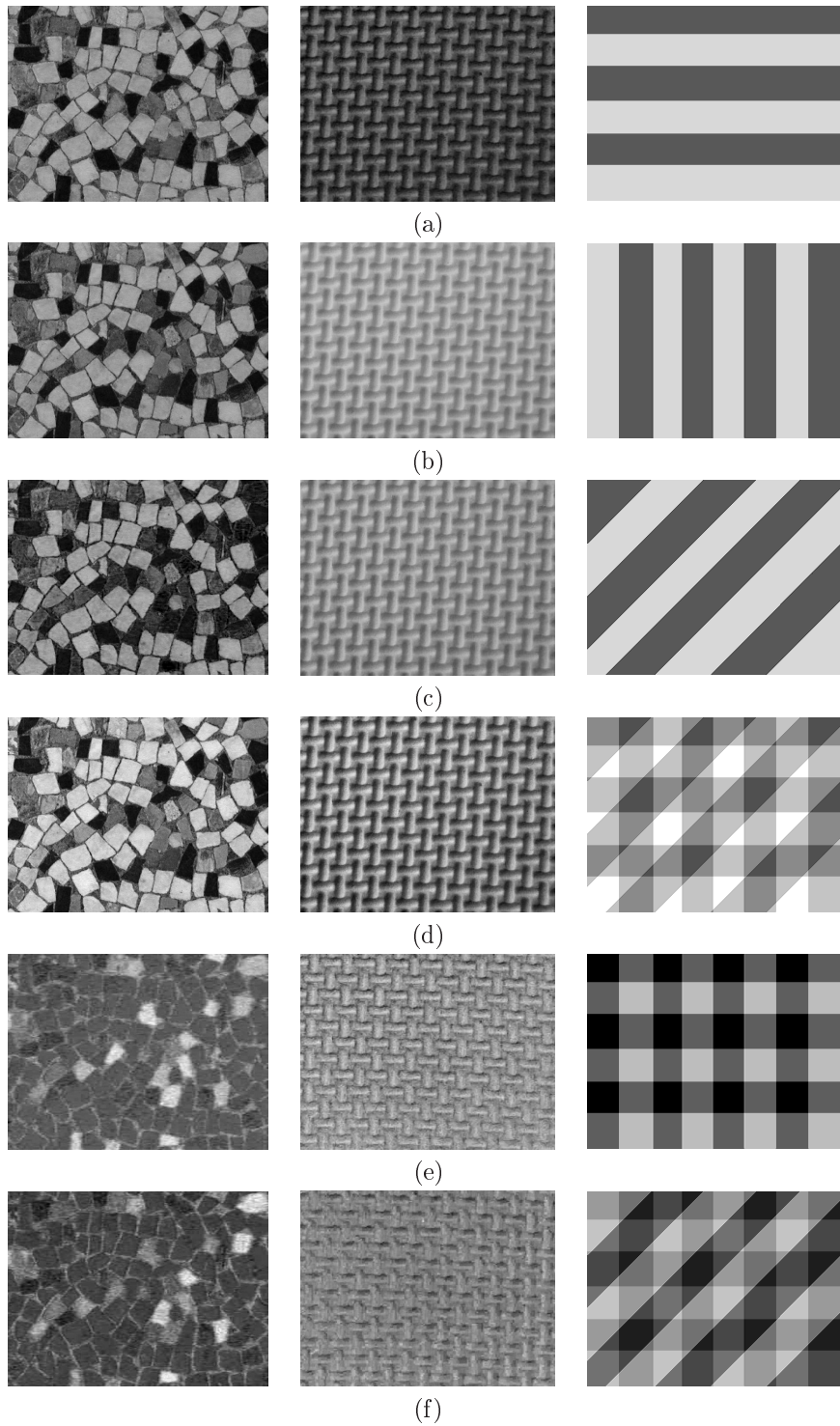


Figura 3.4: En las filas (a)-(c) canales R, G y B respectivamente. En las filas (d)-(f) canales espacio oponente, I, RG y BY respectivamente para las imágenes de la figura 3.3.

3.4.3. *Textons*: atributos de los blobs

Los blobs perceptuales de una imagen se caracterizan por medio de su localización y de sus atributos forma y color, en forma matricial esto se expresa de la siguiente manera:

$$\begin{aligned}\mathbf{T}_{\text{loc}} &= [\mathbf{XY}] \\ \mathbf{T}_{\text{sha}} &= [\mathbf{WL}\Theta] \\ \mathbf{T}_{\text{col}} &= [\mathbf{IRG BY}]\end{aligned}\tag{3.10}$$

donde \mathbf{T}_{loc} contiene la localización de los centros de los blobs, \mathbf{T}_{sha} contiene los atributos que caracterizan la forma de los blobs w, l, θ (ancho, largo y orientación) calculados en las ecuaciones 3.6 y 3.7 y \mathbf{T}_{col} contiene los atributos de color. Los componentes de cada una son las siguientes:

$$\begin{aligned}\mathbf{X}^T &= [x_1 \dots x_n] & \mathbf{Y}^T &= [y_1 \dots y_n] \\ \mathbf{W}^T &= [w_1 \dots w_n] & \mathbf{L}^T &= [l_1 \dots l_n] & \mathbf{\Theta}^T &= [\theta_1 \dots \theta_n] \\ \mathbf{I}^T &= [\bar{i}_1 \dots \bar{i}_n] & \mathbf{RG}^T &= [\bar{r}\bar{g}_1 \dots \bar{r}\bar{g}_n] & \mathbf{BY}^T &= [\bar{b}y_1 \dots \bar{b}y_n]\end{aligned}\tag{3.11}$$

siendo n el número de blobs perceptuales detectados y T indica la versión transpuesta de la matriz.

El atributo de color del blob j -ésimo $(\bar{i}_j, \bar{r}\bar{g}_j, \bar{b}y_j)$ corresponde a la intensidad y la cromaticidad del blob respectivamente. Este atributo se estima a partir de los atributos color de todos los pixels que forman el blob, utilizando para ello el operador estadístico *mediana*. Esta estimación en el color del blob no se ve afectada por los valores atípicos (*outliers*) que pudieran aparecer como consecuencia de la inexactitud en la detección de la forma de los blobs.

La matriz formada por la lista de blobs obtenidos en una imagen además de sus atributos se define bajo el concepto de *textons*, de acuerdo con la teoría de Julesz (1986), y se expresa a través de la matriz:

$$\mathbf{T} = [\mathbf{T}_{\text{loc}}\mathbf{T}_{\text{sha}}\mathbf{T}_{\text{col}}].\tag{3.12}$$

Para ilustrar la descomposición de una imagen en los blobs perceptuales obtenidos en cada uno de los canales del espacio oponente, en la figura 3.5 se muestran dos ejemplos. Los blobs perceptuales obtenidos se han pintado de acuerdo con los atributos que poseen (forma y color). Hay que mencionar que esta descomposición depende únicamente del parámetro η_{det} y del número total de escalas considerado (en ambos casos $\eta_{det} = 0,05$ y $n = 11$, introducidas en la ecuación 3.4).

Los blobs perceptuales en sí mismos constituyen una combinación de dos características perceptuales diferentes, la forma y el color, ya que los atributos relacionados con la forma de los blobs han sido obtenidos en el espacio del color. Por tanto ambas características están relacionadas bajo el concepto de *blob perceptual*.

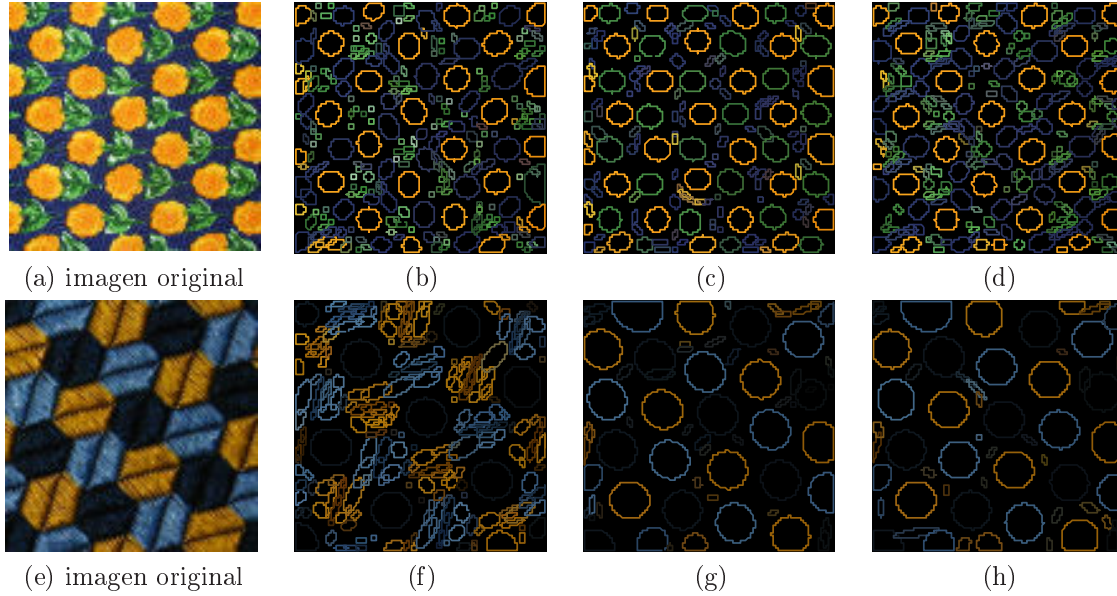


Figura 3.5: (a) y (e) Imágenes con textura en color. (b) y (f) Blobs perceptuales en el canal I. (c) y (g) Blobs perceptuales en el canal RG. (d) y (h) Blobs perceptuales en el canal BY.

3.5. Descriptor de *Textons* (TD)

Este primer descriptor es el resultado de la implementación directa de los *Textons* Julesz (1986) aplicando el modelo *BoW* sobre los *textons* detectados en el apartado 3.4. Es un descriptor muy simple que puede entenderse como la actualización del descriptor usado por Voorhes-Poggio pero extendido a color.

Las características utilizadas en el modelo son los atributos de los *blobs perceptuales* (indicados en la ecuación 3.10) que definen su forma y su color. Se ha representado la forma con los atributos: ancho, largo y orientación, y el color mediante cada una de las componentes del espacio de color oponente. Por tanto los atributos de los blobs utilizados son:

$$\begin{aligned}
 \mathbf{b} &= \{\mathbf{b}_s, \mathbf{b}_c\} \\
 \mathbf{b}_s &= \{w, l, \theta\} \\
 \mathbf{b}_c &= \{i, rg, by\}
 \end{aligned} \tag{3.13}$$

y cada atributo se representa en su propio espacio.

En los siguientes apartados se presenta este descriptor y su evaluación. El procedimiento empleado será el mismo que más adelante será usado para presentar y evaluar otras propuestas de descriptores.

3.5.1. Generación del vocabulario

Para obtener el vocabulario se realiza la cuantificación de cada uno de los espacios donde se representan los atributos de los blobs perceptuales. Para ello se define Q_{Δ} como la función de cuantificación:

$$Q_{\Delta} : \mathbb{R}^k \rightarrow \mathbb{N}^k \quad (3.14)$$

donde k corresponde a la dimensión del espacio en el cual se realiza la cuantificación y el subíndice Δ indica el modelo de cuantificación utilizado en cada caso.

Dado que cada espacio es unidimensional y acotado se utiliza la función Q_{Δ} con $k = 1$. Cada espacio se ha cuantificado con un modelo donde los intervalos tienen igual longitud, que se identifica bajo el nombre M_{\parallel} y por tanto la función de cuantificación que utiliza este modelo se denota como Q_{\parallel} .

Si se cuantifica cada atributo en v intervalos se obtiene su correspondiente vocabulario:

$$VV^X = \{x_1, \dots, x_v\} \quad (3.15)$$

donde x_j una *palabra visual* y X es la variable aleatoria que toma valores en los diferentes atributos, $X \in \{W, L, \Theta, I, RG, BY\}$ que representa respectivamente los atributos: ancho, largo, orientación, intensidad, componente *rg* y componente *by*.

El vocabulario del descriptor *TD* se compone de la unión de los vocabularios obtenidos en cada uno de los espacios o ejes donde se representan los atributos de los blobs y su tamaño total es V , siendo $V = 6 \times v$.

3.5.2. Representación de la imagen

Una vez definido el vocabulario para representar una imagen, se obtienen sus palabras visuales aplicando la función Q_{\parallel} sobre los atributos de sus blobs perceptuales, esto es:

$$\{Q_{\parallel}(\mathbf{W}), Q_{\parallel}(\mathbf{L}), Q_{\parallel}(\mathbf{\Theta}), Q_{\parallel}(\mathbf{I}), Q_{\parallel}(\mathbf{RG}), Q_{\parallel}(\mathbf{BY})\} \quad (3.16)$$

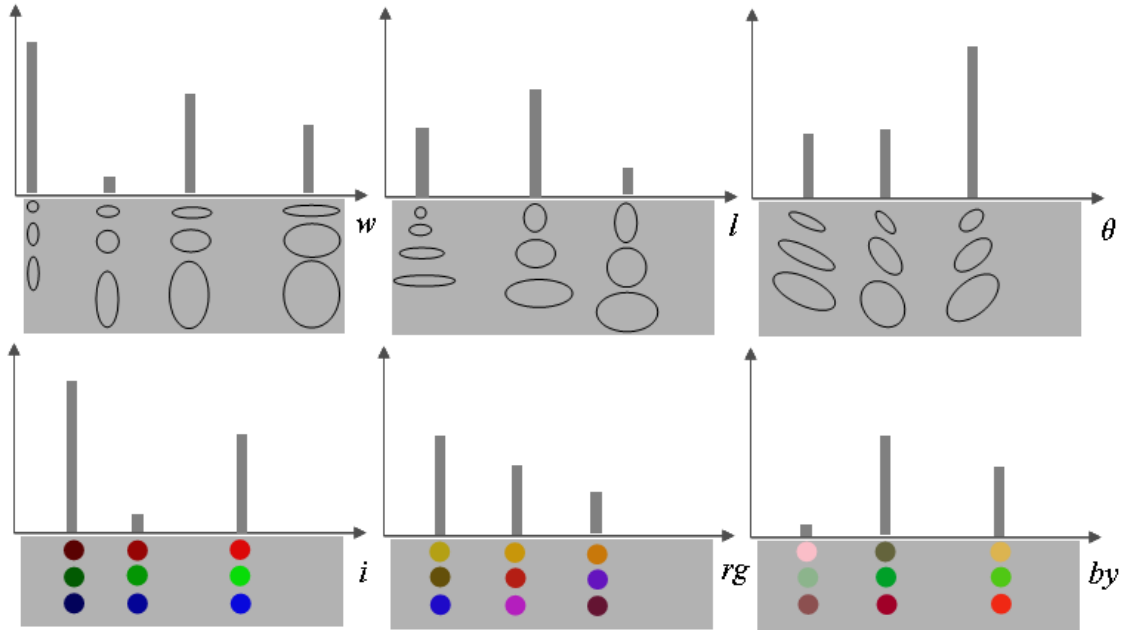
La representación holística de una textura se define a continuación:

Siendo X cualquiera de las variables aleatorias discretas obtenidas aplicando la función de cuantificación Q_{\parallel} , se define la componente i de la representación holística de una textura con la función de densidad de probabilidad, ésta es:

$$P_X(x_i) = P[X = x_i] \quad (3.17)$$

que define la probabilidad de que aparezca un determinado atributo en la imagen.

Entonces se define el *descriptor de textons (TD)* de esta representación como la distribución de probabilidad de las 6 variables aleatorias. El descriptor se expresa de la siguiente manera:

Figura 3.6: Descriptor de *Textons* (TD).

$$TD = [P_W, P_L, P_\theta, P_I, P_{RG}, P_{BY}] \quad (3.18)$$

En la figura 3.6 se muestra un ejemplo del descriptor en donde se puede observar que un mismo término del vocabulario, en el caso de los atributos de forma, hace referencia a diferentes geometrías de blobs. Lo mismo ocurre con los atributos que identifican el color de los blobs perceptuales, una mismo término identifica colores diferentes.

3.5.3. Evaluación de rendimiento

Con el objetivo de comprobar la eficacia de la representación holística y evaluar su rendimiento cuantitativamente, se han realizado experimentos mediante una aplicación de recuperación de imágenes sobre diversas bases de datos de texturas color. La necesidad de encontrar imágenes semánticamente relevantes en grandes bases de datos ha provocado desde hace más de una década el desarrollo de sistemas de recuperación de imágenes basados en su contenido (CBIR). El objetivo de estos sistemas es representar e indexar bases de datos de imágenes usando el contenido visual de éstas. La mayoría de sistemas CBIR utilizan características de bajo nivel (color, forma, textura y/o localización espacial) para describir el contenido de las imágenes.

Diseño del experimento

Para evaluar los sistemas de recuperación de imágenes es necesario determinar el conjunto de texturas similares a una textura dada que el sistema de recuperación debería de proporcionar. Por

tanto, hay que determinar para cada textura el conjunto de imágenes similares, también conocido como *conjunto relevante*. Dado que la definición exacta de los conjuntos relevantes es altamente subjetiva, prácticamente en toda la experimentación publicada al respecto se construye el conjunto relevante de una textura sub-dividiendo ésta en porciones disjuntas. Éste ha sido el procedimiento empleado a la hora de construir los conjuntos relevantes en las diferentes BD utilizadas en esta experimentación.

Para valorar cuantitativamente el resultado de la recuperación de imágenes se han utilizado la medida *recall* (J. Smith, 1998). El *recall* mide la habilidad del sistema de encontrar todas las imágenes relevantes. Esta medida depende del número de imágenes a buscar (r) y se define de la siguiente manera:

$$recall(r) = \frac{\text{Retrieval}}{\text{Rellevant}} \quad (3.19)$$

siendo *Retrieval* el número de imágenes relevantes encontradas dentro de la r buscadas y *Rellevant* es el número de imágenes relevantes en la BD.

En todas las BD utilizadas el número de imágenes relevantes corresponde al número de sub-imágenes en que se han dividido las texturas.

Dado que para cada BD se realiza la recuperación de todas sus imágenes, se calcula el *recall promedio* en forma de porcentaje, que se denota como \overline{recall} , de la siguiente forma:

$$\overline{recall}(r) = \frac{1}{P} \sum_{i=1}^P recall_i(r) \times 100 \quad (3.20)$$

siendo $recall_i$ la medida de recall de la imagen i y P el número de imágenes de la BD. Idealmente, si r es igual al número de imágenes relevantes el *recall promedio*(r) debería de ser 100.

Un factor a tener en cuenta a la hora de evaluar los resultados es que la manera de construir el conjunto de imágenes relevantes, sub-dividiendo una textura en trozos, puede conducir a errores en la recuperación de imágenes. Estos son provocados si hay una falta de homogeneidad en el conjunto relevante. En la figura 3.7 se muestran dos ejemplos en donde el conjunto relevante de una textura no es homogéneo ya que las sub-imágenes que los forman no tienen las mismas propiedades.

Dado que las texturas se representan con un modelo estadístico y que dos texturas color con la misma o similar representación corresponden a la misma textura. Para comparar distribuciones de densidad de probabilidad se ha utilizado la función χ^2 que proporciona una medida de similitud entre dos distribuciones de probabilidad. Dadas las distribuciones de probabilidad H_1 y H_2 la distancia entre éstas se define de la siguiente manera:

$$\chi^2(H_1, H_2) = \frac{1}{2} \sum_{t=1}^T \frac{[H_1(t) - H_2(t)]^2}{H_1(t) + H_2(t)}. \quad (3.21)$$

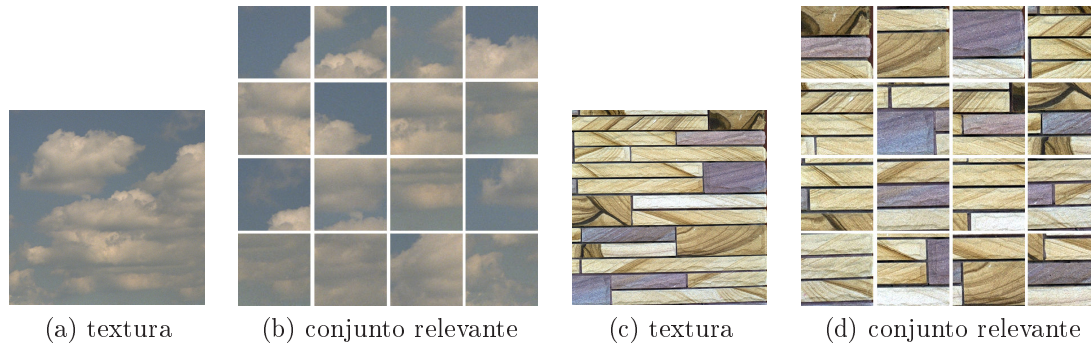


Figura 3.7: Ejemplos de conjuntos relevantes no homogéneos de la BD *VisTexL*. (b) y (d) muestran las subdivisiones de las texturas (a) y (c) que componen respectivamente sus conjuntos relevantes.

Bases de datos de imágenes

A continuación se indica el proceso de formación de las bases de datos (BD) de texturas color que se han utilizado:

- Outex: base de datos creada por Ojala, Mäenpää, y cols. (2002) (TC-00013). Ésta contiene 68 texturas (resolución 640×512 pixels) y cada una de ellas está dividida en 20 sub-imágenes de resolución 128×128 pixels.
- VisTex⁴: base de datos formada por texturas y escenas en el exterior (resolución 512×512 pixels). Normalmente se han utilizado diversos subconjuntos de ésta que incluyen sólo las texturas. Se han considerado dos de estos subconjuntos, el utilizado por Mäenpää y Pietikäinen (2004) con 54 texturas que se identificará como *VisTexP* y el utilizado por Liapis y Tziritas (2004) con 55 texturas que se identificará como *VisTexL*. Ambos dividen cada textura en 16 sub-imágenes de 128×128 pixels.
- Colección Corel⁵: está formada a su vez por 11 bases de datos de texturas de diferentes tipos, a continuación se indican los nombre de éstas, entre paréntesis consta el número de la referencia y el nombre bajo el cual se han identificado: Textures (137000, *CorelTex*), Textures II (404000, *CorelTex2*), Various Textures I (593000, *CorelV1Tex*), Various Textures II (594000, *CorelV2Tex*), Textile Patterns (192000, *CorelTexPat*), Sand & Pebble Textures (390000, *CorelSand*), Bark Textures (399000, *CorelBark*), Colors & Textures (403000, *CorelCol*), Marble Textures (349000, *CorelMarb*), Painted Textures (265000, *CorelPain*), Shell Textures (355000, *CorelShel*). Todas las BD contienen 100 texturas excepto la BD *CorelCol* que tiene 85. La resolución de éstas es de 1536×1024 pixels, pero para la experimentación se ha reducido a la mitad (768×512 pixels). Cada textura se ha dividido en 6 sub-imágenes de 240×240 pixels.

⁴Vision Texture-VisTex database (MIT Media Lab) <http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html>.

⁵BD Corel se distribuye en <http://www.emsps.com/photocd/corelclds.htm>.

- Arenas⁶: contiene 50 imágenes con diferentes tipos de arena (con resolución 3080×1452 pixels). Cada textura se ha dividido en 60 sub-imágenes de 256×256 pixels.
- BDHomo: es una BD que contiene exclusivamente texturas homogéneas⁷, su composición es la siguiente: 59 texturas obtenidas del album *Mayang*⁸, 16 texturas de la BD Outex (TC-00013) y 5 texturas de la BD VisTex, éstas son 80 texturas diferentes. De cada textura se han obtenido 6 muestras de 100×100 pixels.

En el apéndice B se encuentran las imágenes que componen las diferentes bases de datos.

Dado que la falta de homogeneidad en el conjunto relevante provoca errores, esta fuente de errores se ha evitado con la construcción de la BD *BDHomo*. Ésta es la única las 16 BD utilizadas en la que los conjuntos relevantes de todas las texturas son homogéneos, de manera que con esta BD se puede valorar exactamente la eficacia del descriptor propuesto. El resto de BD presentan diferentes grados de no homogeneidad en sus conjuntos relevantes.

Comparación con otros métodos

Con el objetivo de poder contrastar los resultados se ha realizado el mismo experimento con conocidos descriptores de texturas y color. Para ello se han utilizado los dos modelos de fusión de textura y color referenciados en el apartado 2.4.1. Para el modelo donde se trata por separado la textura y el color se han utilizado descriptores MPEG-7 (Manjunath y cols., 2001) y para el modelo donde se tratan conjuntamente estas características se ha utilizado el descriptor *LBP RGB* (Mäenpää y Pietikäinen, 2004).

Dentro del estándar MPEG-7, para describir el contenido de datos multimedia se han utilizado dos de los descriptores para describir el contenido de color, concretamente el *CSD* y el *SCD*. Para describir la textura se han utilizado dos descriptores, el *HTD* (adecuado para texturas homogéneas) y el *EHD* (adecuado para texturas no homogéneas). En los apartados 2.2.5 y 2.3.2 se encuentran, respectivamente, la descripción de estos dos descriptores. El estándar MPEG-7 no define como combinar descriptores, por lo que se ha utilizado la combinación utilizada por Dorairaj y Namuduri (2004). Estos autores realizan la combinación de descriptores durante el cálculo de disimilitudes. La disimilitud entre dos texturas es obtenida simplemente sumando las disimilitudes de cada descriptor.

Respecto al descriptor *LBP RGB*, en la bibliografía se encuentran diversas maneras de aplicarlo: con diferentes parámetros y combinaciones de los propios descriptores, por este motivo se ha creído oportuno utilizar tres modelos diferentes (Mäenpää y cols., 2002; Mäenpää y Pietikäinen, 2004; Takala y cols., 2005). Los modelos utilizados son: $LBP(RGB)_{8,1}$ referenciado como LBP_1 , $LBP(RGB)_{8,1}^{u^2} + LBP(RGB)_{16,3}^{u^2} + LBP(RGB)_{24,5}^{u^2}$ referenciado como LBP_2 y $LBP(RGB)_{8,1}^{u^2} + LBP(RGB)_{16,2}^{u^2} + LBP(RGB)_{24,3}^{u^2}$ referenciado como LBP_3 .

⁶BD cedida por Coastal Image Lab.

⁷Aquellas que tienen las mismas propiedades en toda la superficie que ocupa la imagen.

⁸BD de W. Smith y A.M. Murni, 2001. <http://www.mayang.com/textures/>.

Para calcular la distancia entre descriptores LBP se ha utilizado la medida propuesta por Mäenpää y Pietikäinen (2004) expresada en la fórmula 3.22, mientras que para calcular la distancia entre los diferentes descriptores MPEG-7 según se indica en (Manjunath y cols., 2003) se ha utilizado la norma $l1$ indicada en la fórmula 3.23.

$$dist(s, m) = - \sum_{i=0}^{N-1} s_i \log m_i \quad (3.22)$$

$$dist(s, m) = \sum_{i=0}^{N-1} |s_i - m_i| \quad (3.23)$$

3.5.4. Experimento 1: Evaluación del TD

Para evaluar el descriptor TD se han realizado diversas pruebas variando su vocabulario. Cada vocabulario se ha obtenido utilizando la función de cuantificación $Q_{||}$ (especificada en el apartado 3.5.1) variando el tamaño de los intervalos. Dado un vocabulario para cada imagen se calcula la similitud entre su representación y la del resto de las imágenes de la BD (utilizando la medida 3.21), obteniendo así la lista de imágenes similares ordenada por similitud (ranking).

Resultados experimentales

Para cada vocabulario y en cada BD de texturas se ha calculado el $\overline{recall}(r)$, siendo r el número de imágenes relevantes en cada BD. Los resultados de la recuperación de imágenes se muestran en la tabla 3.1. En la fila superior de ésta se indica, para cada vocabulario, el número de intervalos en que se han cuantificado cada uno de los atributos de los *textons* (v) y el número total de términos del vocabulario derivado (V). Los mejores resultados para cada BD se han destacado en negrita.

Comparando todos los vocabularios que se muestran en la tabla 3.1, se puede observar como el número de intervalos utilizados en la cuantificación de los *textons* puede hacer variar, en algunos casos significativamente, la eficacia del descriptor TD . En 8 de las 16 BD el mejor resultado corresponde al obtenido al cuantificar cada uno de los atributos de los *textons* en 128 intervalos dando lugar a un vocabulario de 768 términos (última columna de la tabla 3.1), mientras que en 6 BD el mejor resultado corresponde a un vocabulario con la mitad de términos (384) respecto al anterior (penúltima columna de la tabla 3.1). Teniendo en cuenta el coste y la eficacia de los diferentes vocabularios, el vocabulario de 384 términos tiene una eficacia de representación prácticamente la misma que con el vocabulario de 768 pero con un coste mucho menor. Esto se confirma en la media de las eficiencias de todas las BD (última fila de la tabla).

Comparación con otros descriptores

La eficacia del descriptor TD se ha comparado con la obtenida por el descriptor $LBP\ RGB$ y diversas combinaciones de descriptores MPEG-7 tal y como se ha indicado en el apartado 3.5.3.

Tabla 3.1: Recall promedio con el descriptor *TD* utilizando diferentes cuantificaciones y diferentes BD.

BD \ (v, V)	$(9, 54)$	$(11, 66)$	$(16, 96)$	$(32, 192)$	$(64, 384)$	$(128, 768)$
<i>Outex</i>	61.41	65.04	68.84	73.35	75.39	75.47
<i>VisTexL</i>	80.01	81.83	82.02	84.30	84.01	84.11
<i>VisTexP</i>	76.51	78.20	78.44	80.58	80.35	80.60
<i>CorelTex</i>	73.47	74.53	76.14	77.92	78.78	78.47
<i>CorelTex2</i>	76.69	77.81	81.78	84.11	84.58	84.53
<i>CorelV1Tex</i>	86.75	87.83	88.72	90.94	91.03	91.28
<i>CorelV2Tex</i>	86.61	88.00	89.36	90.03	90.58	90.53
<i>CorelTexPat</i>	88.92	89.19	90.33	91.31	91.56	91.72
<i>CorelSand</i>	67.00	67.61	71.53	75.08	75.75	75.28
<i>CorelBark</i>	57.25	58.06	61.42	63.11	63.69	63.89
<i>CorelCol</i>	73.07	72.97	74.28	75.65	76.24	76.44
<i>CorelMarb</i>	66.69	69.39	73.36	76.44	78.33	78.47
<i>CorelPain</i>	67.06	68.03	69.69	72.42	72.28	72.25
<i>CorelShel</i>	42.39	43.64	45.64	48.08	48.78	48.75
<i>Arenas</i>	71.83	74.07	73.41	74.86	75.70	73.95
<i>BDHomo</i>	95.28	95.73	96.84	97.29	97.40	97.57
Media	73.18	74.47	76.36	78.46	79.02	78.95

Para medir la calidad de los descriptores se ha calculado el $\overline{recall}(r)$, siendo r el número de imágenes relevantes en cada BD.

Se ha incluido dentro de la comparativa los resultados obtenidos por Liapis y Tziritas (2004) sobre la BD *VisTexL* con el descriptor *CDWF* y los obtenidos por Kandaswamy y Adjeroh (2005) sobre la BD *Outex* con el descriptor *CCTM*.

En la tabla 3.2 se encuentran los resultados obtenidos en las 16 BD, en donde para cada BD se ha destacado en negrita el mejor resultado obtenido. Los resultados del descriptor *TD* corresponden a los obtenidos con el vocabulario de 384 términos.

Analizando los resultados de la tabla comparativa se comprueba que el descriptor *TD* sólo es superior al resto de descriptores en 3 BD, sin embargo en 8 de las 13 BD restantes la diferencia entre la eficiencia de este descriptor y el que produce el mejor resultado es apenas del 3%. Con respecto a la eficiencia media (última fila de la tabla), la del descriptor *TD* se haya muy cercana a la del descriptor *EHD+CSD* y es superior al resto de descriptores.

Conclusiones

A tenor de los resultados de la experimentación llevada a cabo con el descriptor *TD*, aunque en media el comportamiento de éste es similar al resto de los descriptores, en algunas BD no parece que la información capturada por el descriptor sea suficiente para poder discriminar entre todas las texturas. Sin embargo, se ha demostrado que con un descriptor muy simple, basado en un espacio de baja dimensión y con un método sencillo de cuantificación, se llega a un buen nivel dentro del estado del arte, aunque no se superan los resultados obtenidos con algunos de los descriptores de

Tabla 3.2: Comparación del descriptor TD con otros descriptores.

BD	TD	$HTD+SCD$	$HTD+CSD$	$EHD+CSD$	LBP_1	LBP_2	LBP_3	$CCTM$	$CDWF$
<i>Outex</i>	75.39	61.10	66.02	67.64	56.87	46.82	60.80	70.00	-
<i>VisTexL</i>	84.01	87.41	83.45	85.25	73.20	50.36	66.11	-	91.3
<i>VisTexP</i>	80.35	83.44	80.24	81.81	70.38	45.85	64.69	-	-
<i>CorelTex</i>	78.78	67.33	72.36	75.31	61.89	60.83	62.22	-	-
<i>CorelTex2</i>	84.58	76.11	85.89	87.33	72.50	68.89	72.42	-	-
<i>CorelV1Tex</i>	91.03	85.94	91.89	92.78	77.53	72.14	74.56	-	-
<i>CorelV2Tex</i>	90.58	88.53	85.89	92.28	81.47	79.81	81.44	-	-
<i>CorelTexPat</i>	91.56	93.69	89.86	90.50	86.89	82.92	86.03	-	-
<i>CorelSand</i>	75.75	72.00	85.06	85.28	55.54	60.17	59.89	-	-
<i>CorelBark</i>	63.69	64.83	69.11	70.36	48.39	51.69	51.67	-	-
<i>CorelCol</i>	76.24	73.43	72.78	75.52	68.89	68.73	70.62	-	-
<i>CorelMarb</i>	78.33	70.67	80.11	82.03	59.78	63.72	65.19	-	-
<i>CorelPain</i>	72.28	62.31	72.61	73.94	50.67	49.11	51.69	-	-
<i>CorelShel</i>	48.78	41.50	47.06	49.97	32.83	35.94	35.67	-	-
<i>Arenas</i>	75.70	56.99	80.09	80.21	52.06	63.29	61.47	-	-
<i>BDHomo</i>	97.40	97.85	97.78	98.21	91.11	85.28	84.93	-	-
Media	79.02	73.94	78.76	80.52	64.99	61.59	68.58	-	-

referencia MPEG-7. Por otro lado, es necesario mejorar la integración de la textura y el color ya que el descriptor no discrimina correctamente entre diferentes texturas en color. En la figura 3.8 se muestra un ejemplo que ilustra este hecho, dos texturas que perceptualmente son diferentes pero tienen la misma representación con el descriptor TD .

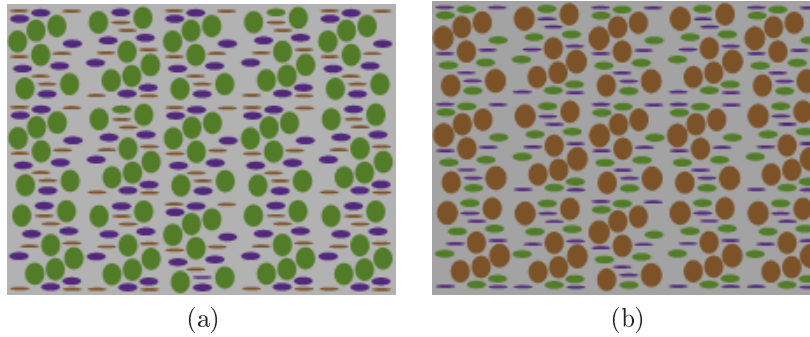


Figura 3.8: Las texturas (a) y (b) son perceptualmente diferentes pero tienen el mismo descriptor TD .

3.6. Descriptor Perceptual Conjunto de *Textons* (JPTD)

Teniendo en cuenta que la teoría de los *textons* Julesz (1986) fue obtenida a través de una experimentación sobre imágenes sintéticas (monocromas) con texturas simples, esto podría ser el motivo que ha provocado que la implementación de esta teoría mediante el descriptor TD no haya dado los resultados esperados. La experimentación en este caso se ha realizado con texturas

naturales en color y ésta última es una de las características que Julesz no tuvo en cuenta en su experimentación. En los siguientes apartados se proponen diferentes mejoras de la representación holística básica derivada de la teoría de Julesz.

3.6.1. Representación de las características

Para mejorar la representación de las características de textura y color se parte de dos observaciones sobre la representación con el descriptor *TD*:

1. La textura y el color están deslocalizadas ya que no se tiene en cuenta la co-ocurrencia de estas características.
2. La generación de los vocabularios no considera relaciones de apariencia perceptual.

Los atributos de los blobs perceptuales están relacionados entre si porque definen dos propiedades de la texturas, las formas de los elementos que contiene y los colores que poseen estos elementos. La propia naturaleza de estas propiedades ha motivado la utilización de un nuevo espacio de representación que mantenga las relaciones entre los atributos de los blobs perceptuales, perdidas en la representación holística de la textura mediante el descriptor *TD*.

Se han utilizado dos espacios diferentes, uno para representar los atributos de forma (\mathbf{T}_{sha}) y otro para representar los atributos del color de los blobs (\mathbf{T}_{col}). Estos espacios se han definido de manera que sean perceptuales y uniformes, es decir, las distancias entre los puntos deben ser proporcionales a su diferencia perceptual, y por tanto similares a los espacios uniformes de color (Wyszecki y Stiles, 1982).

El espacio de representación de los atributos de forma utilizado es similar al definido por Salvatella y Vanrell (2007), en donde proponen un espacio cuatridimensional cilíndrico caracterizado por dos ejes que representan la forma del blob (relación de aspecto y área), el tercer eje representa su orientación y el cuarto el signo del blob (contraste). En el espacio propuesto se ha prescindido es la cuarta dimensión y se ha realizado la siguiente transformación no lineal U ,

$$\begin{aligned} U : \quad \mathbb{R}^3 &\rightarrow \mathbb{R}^3 \\ (w, l, \theta) &\rightarrow (r, z, \phi) \end{aligned} \tag{3.24}$$

donde $r = \log_2(ar)$, $z = \log_2(\log_2(A))$ y $\phi = 2\theta$, siendo ar la relación de aspecto del blob ($ar = w/l$), A su área ($area = w \cdot l$) y θ su orientación.

Hay que mencionar que en este espacio los blobs válidos están localizados dentro del cono $\alpha_{max} = \pi/4$ mostrado en la figura 3.9, dado que por definición la anchura del blob es la más pequeña de las dos longitudes que caracterizan un blob.

Para representar el color de los blobs, el espacio de color más oportuno debería de ser cualquiera de los derivados del CIE ($L^*u^*v^*$ y $L^*a^*b^*$) ya que estos son perceptualmente uniformes. Dado que su uso está limitado al procesado de imágenes calibradas y que actualmente esto constituye

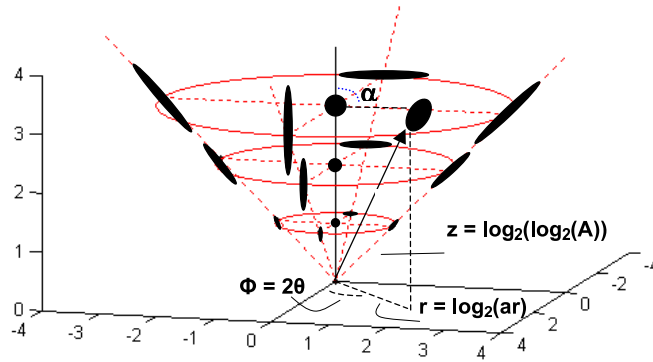


Figura 3.9: Espacio de representación de forma de los blobs en coordenadas cilíndricas.

una restricción de cara a su aplicabilidad, en su lugar se puede utilizar cualquiera de las variantes de la familia de espacios color HSI. Estos espacios son similares a los espacios de color uniformes por poseer cierta correlación con la percepción humana del color.

Dado que el color de los blobs \mathbf{T}_{col} (ecuación 3.10) está expresado en el espacio de colores oponentes, será necesaria realizar una transformación para obtener las coordenadas en el espacio color HSI:

$$\begin{aligned}
 HSI : \quad \mathbb{R}^3 &\rightarrow \mathbb{R}^3 \\
 (i, rg, by) &\rightarrow (h, s, i)
 \end{aligned}
 \tag{3.25}$$

Las nuevas características que definen los *textons* de una textura son entonces (r, z, ϕ) que identifican las formas de los blobs perceptuales y (h, s, i) que identifican el color de estos.

3.6.2. Generación del Vocabulario

El hecho de que los espacios de representación de los *textons* estén acotados, permite la construcción del vocabulario de *textons* simplemente realizando una cuantificación de estos espacios. El vocabulario obtenido es por tanto general, ya que no depende de las texturas que se quieran representar, y su tamaño vendrá determinado por el número de intervalos (niveles) utilizados en la cuantificación de estos espacios.

En el ámbito de investigación del color se ha demostrado que el uso de un espacio de color apropiado y la óptima cuantificación de éste son una clave importante en el rendimiento de su aplicación (Lee y cols., 2005). Por este motivo se han explorado diversas maneras de construir el vocabulario:

- Utilizando diferentes espacios para representar el color de los blobs perceptuales.
- Cuantificando de diferente forma los espacios de *textons*.

Para representar el color de los blobs perceptuales se han utilizado tres espacios de la familia

HSI introducidos en el apartado 2.2.3, estos son: HSI-Carron, HSV-Smith y HSV-Yagi. Respecto a su cuantificación, se han diseñado dos modelos:

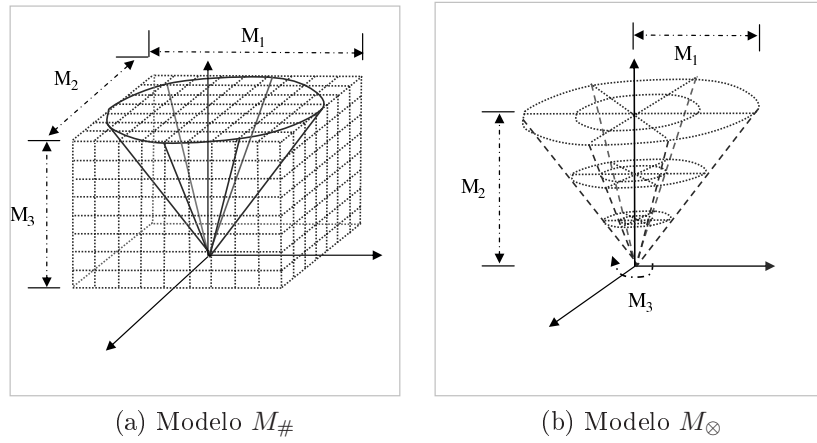


Figura 3.10: Diferentes modelos de cuantificación.

- modelo $M_{\#}$: cuantificación en celdas con forma de paralelepípedo (en la figura 3.10.(a) se muestra un ejemplo). En este caso cada uno de los tres ejes del espacio puede ser cuantificados con el mismo número de niveles (Lee y cols., 2005) (cuantificación uniforme) o como proponen Swain y Ballard (1991) realizar una cuantificación más fina en los ejes cromáticos y una cuantificación más burda en el eje acromático. La cuantificación se realiza en los ejes cartesianos definidos por $(s \cos(h), s \sin(h), i)$.
- modelo M_{\otimes} : cuantificación en sectores. Dado que los ejes de los espacios de la familia HSI tienen un significado perceptual, se pueden cuantificar cada uno de ellos de manera independiente dando lugar a la cuantificación que se muestra en la figura 3.10.(b). En este caso se realiza la cuantificación en coordenadas cilíndricas (s, h, i) .

Utilizando cualquiera de los modelos de cuantificación sobre el espacio de color si se cuantifica cada uno de los ejes en M_1 , M_2 y M_3 términos se obtiene su correspondiente vocabulario. Para ello se define Q_{Δ}^C , al igual que la función referenciada como 3.14 para $k = 3$, como la función de cuantificación aplicada sobre el espacio de color. Por tanto un término del vocabulario se obtiene de la siguiente manera:

$$Q_{\Delta}^C(HSI(\mathbf{b}_c)) \quad (3.26)$$

siendo HSI la transformación del espacio de atributos de color indicada en la ecuación 3.25 y \mathbf{b}_c representa un blob perceptual tal como se ha definido en la ecuación 3.13.

El tamaño total del vocabulario generado en el espacio de color es de $M = M_1 \times M_2 \times M_3$ palabras visuales.

Para referenciar la función de cuantificación con el modelo $M_{\#}$ se utilizará la notación $Q_{\#}^C$, y para el modelo M_{\otimes} la función Q_{\otimes}^C .

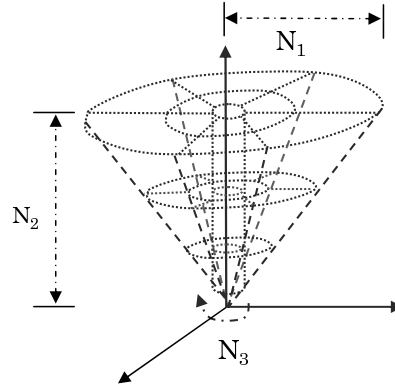


Figura 3.11: Cuantificación del espacio de forma con el modelo M_{\otimes} .

De la misma manera, el vocabulario correspondiente al espacio de los atributos de forma de los blobs ha sido construido teniendo en cuenta consideraciones perceptuales y se ha considerado la transformación propuesta en la ecuación 3.24 como la de un espacio uniforme. Por este motivo admite una cuantificación uniforme en celdas del mismo tamaño, pero la perceptualidad de sus ejes permite también una cuantificación independiente para cada uno de sus ejes. Por lo tanto los dos modelos de cuantificación descritos anteriormente también han sido utilizados en la cuantificación del espacio de forma. En el modelo $M_{\#}$ se realiza la cuantificación de los ejes cartesianos $(r \cos(\phi), r \sin(\phi), z)$ y en el modelo M_{\otimes} se hace en las coordenadas cilíndricas (r, ϕ, z) . Se ha añadido un tercer modelo, variante del modelo de cuantificación en sectores, en el cual se han diferenciado las formas correspondientes a blobs isotrópicos (los que se sitúan cercanos al eje vertical). En la figura 3.11 se muestra el ejemplo de esta cuantificación a la que se ha dado el nombre de *modelo M_{\otimes}* .

Utilizando cualquiera de los modelos de cuantificación sobre el espacio de atributos de forma si se cuantifica cada uno de los ejes en N_1 , N_2 y N_3 términos se obtiene su correspondiente vocabulario. Para ello se define Q_{Δ}^S , al igual que la función referenciada como 3.14 para $k = 3$, como la función de cuantificación aplicada sobre el espacio de atributos de forma. Por tanto un término del vocabulario se obtiene de la siguiente manera:

$$Q_{\Delta}^S(U(\mathbf{b}_s)) \quad (3.27)$$

siendo U la transformación perceptual del espacio de atributos de forma definida en la ecuación 3.24 y \mathbf{b}_s representa un blob perceptual tal como se ha definido en la ecuación 3.13.

El tamaño total del vocabulario generado en el espacio de forma es de $N = N_1 \times N_2 \times N_3$ palabras visuales.

Para referenciar la función de cuantificación con el modelo $M_{\#}$ se utilizará la notación $Q_{\#}^S$, para el modelo M_{\otimes} la función Q_{\otimes}^S y para el modelo M_{\circ} la función Q_{\circ}^S .

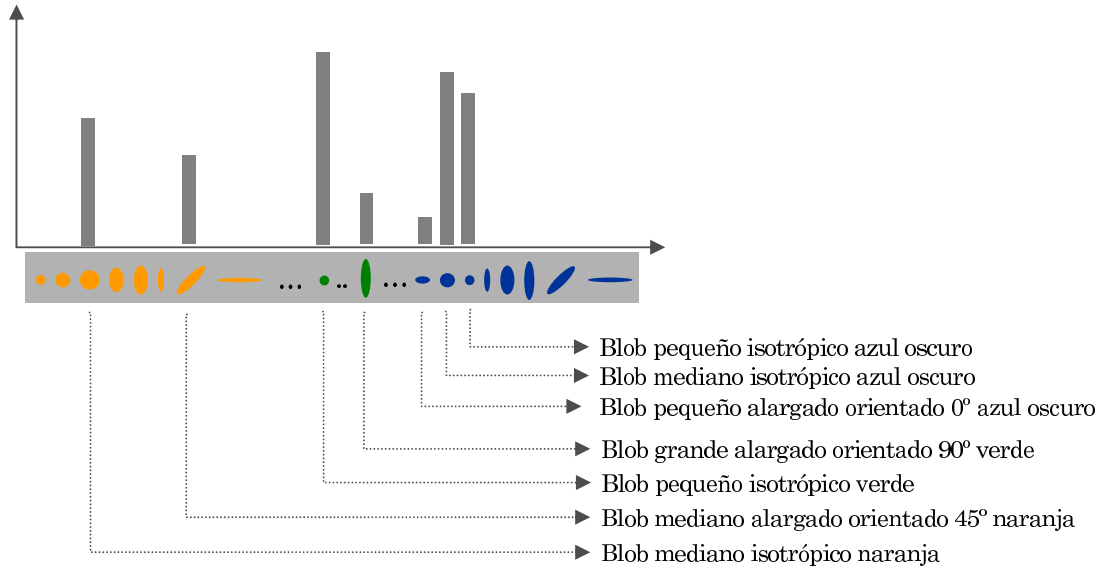


Figura 3.12: Descriptor Perceptual Conjunto de *Textons* (JPTD).

3.6.3. Representación de la imagen

En esta representación se fusiona en un único vocabulario los vocabularios que describen la forma y el color de los blobs. De esta manera, en esta representación holística se asume que cada clase de textura color se caracteriza por poseer determinadas formas con determinados colores. Desde el punto de vista estadístico se pueden considerar las características forma y color como variables aleatorias y a la conjunción de ambas como una variable aleatoria bidimensional. Así, se define C como la variable aleatoria discreta que pertenece al espacio cuantificado de color y S es la variable aleatoria discreta que pertenece al espacio cuantificado de forma, entonces (C, S) es la variable aleatoria bidimensional.

La representación holística en este caso modela estadísticamente las texturas con la función de densidad de probabilidad conjunta de los atributos forma y color, ésta se define como:

$$P_{C,S}(c_i, s_j) = P[C = c_i, S = s_j] \quad (3.28)$$

que representa la probabilidad de que aparezcan un conjunto de atributos en una imagen.

El descriptor que da lugar a esta representación, el *descriptor perceptual conjunto de textons* (JPTD), es la función de densidad de probabilidad de todos los valores de la variable aleatoria bidimensional (C, S) ,

$$JPTD = [P_{C,S}(c_1, s_1), \dots, P_{C,S}(c_M, s_N)] \quad (3.29)$$

De acuerdo con la función de cuantificación Q_{Δ}^C , la variable aleatoria C puede tener M posibles valores y la variable S , obtenida al cuantificar los atributos forma con la función Q_{Δ}^S , puede tener

N posibles valores. Por tanto la fusión de la forma y el color da lugar a un único vocabulario de tamaño $M \times N$.

En la figura 3.12 se muestra un ejemplo del descriptor en donde se puede observar que los términos del vocabulario tienen un significado claramente perceptual. Además, se pueden usar términos simples del lenguaje natural para explicar las palabras visuales.

3.6.4. Experimento 2: Evaluación del *JPTD*

Para comprobar el comportamiento del descriptor *JPTD* éste se ha aplicado para recuperar imágenes en BD. Se han utilizado las mismas BD y el mismo procedimiento que el descrito en 3.5.3. En este caso, sobre cada una de las imágenes de las BD se ha calculado el descriptor *JPTD* y a continuación, para cada imagen se calcula la similitud entre su representación y la del resto de las imágenes de la BD utilizando la distancia χ^2 (indicada en la ecuación 3.21).

Para valorar cuantitativamente el resultado de la recuperación de imágenes, dado que para cada BD se realiza la recuperación de todas sus imágenes, se calcula el *recall*.

Resultados experimentales

Se ha realizado diversas pruebas variando el vocabulario del descriptor *JPTD*. Cada vocabulario se obtiene al utilizar diferentes funciones de cuantificación Q_{Δ}^S y Q_{Δ}^C según los modelos descritos en el apartado 3.6.2. Para representar el color de los blobs perceptuales se han utilizado en pruebas diferentes los espacios de color *HSI-Carron* y *HSV-Smith*. Las pruebas realizadas se resumen en la tabla 3.3 en donde se indica la codificación de éstas, p.ej: la prueba q_2 corresponde al experimento utilizando el descriptor con el modelo de cuantificación $Q_{\#}^S$ para los atributos de forma y el modelo de cuantificación $Q_{\#}^C$ para los atributos de color, con un vocabulario total de 60750 palabras visuales.

Tabla 3.3: Codificación de los vocabularios testeados.

Modelos de cuantización	Tamaño del vocabulario						
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3).N, N \times M$	(5,5,5),125, 15625	(5,5,5),125, 60750	(6,6,6),216, 55296	(7,7,7),216, 55296	(7,7,7),343, 166698	(5,5,7),175, 105000
	$(M_1, M_2, M_3).M$	(5,5,5),125	(9,9,6),486	(8,8,4),256	(8,8,4),256	(9,9,6),486	(10,10,6),600
	<i>HSI-Carron</i> <i>HSV-Smith</i>	q_1 q_1	q_2 q_2	q_3 q_3	q_4 q_4	q_5 q_5	q_6 q_6
$Q_{\#}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3).N, N \times M$	(5,5,5),125, 32000	(7,7,7),343, 87808	(5,5,7),175, 100800			
	$(M_1, M_2, M_3).M$	(4,16,4),256	(4,16,4),256	(6,16,6),576			
	<i>HSI-Carron</i> <i>HSV-Smith</i>	q_7 q_7	q_8 q_8	q_9 q_9			
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3).N, N \times M$	(4,8,5),160, 20480	(4,8,5),160, 40960	(4,8,7),224, 57344	(3,8,7),168, 96768		
	$(M_1, M_2, M_3).M$	(4,8,4),128	(4,16,4),256	(4,16,4),256	(6,16,6),576		
	<i>HSI-Carron</i> <i>HSV-Smith</i>	q_{10} q_{10}	q_{11} q_{11}	q_{12} q_{12}	q_{13} q_{13}		
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3).N, N \times M$	(3,8,7),168, 43008	(3,8,7),168, 67200	(3,8,7),168, 96768			
	$(M_1, M_2, M_3).M$	(4,16,4),256	(5,16,5),400	(6,16,6),576			
	<i>HSI-Carron</i> <i>HSV-Smith</i>	q_{14} q_{14}	q_{15} q_{15}	q_{16} q_{16}			

En la figura 3.13 se encuentra representada de manera gráfica la eficiencia obtenida en cada prueba ($\overline{recall}(r)$, siendo r el número de imágenes relevantes de cada BD). Esta información se encuentra más detallada en las tablas A.1-A.16 que figuran en el apéndice A.

En las gráficas de la figura 3.13 para cada BD se han destacado, con un punto de mayor diámetro, el vocabulario donde se ha obtenido la eficiencia máxima mientras que las discontinuidades de las líneas que unen los resultados de las diferentes pruebas muestran un cambio de modelo de cuantificación.

En la mayoría de las BD, los modelos de cuantificación que dan lugar a los mejores resultados son Q_{\otimes}^S , Q_{\otimes}^C y $Q_{\otimes}^S, Q_{\otimes}^C$ con el mismo tamaño de vocabulario (96768 términos). Estos corresponden a los vocabularios codificados como $q16$ y $q13$ respectivamente. El vocabulario $q9$, que corresponde al modelo de $Q_{\#}^S, Q_{\otimes}^C$ de 100800 términos, supera con diferencia a todos los vocabularios en las BD *CorelSand*, *CorelShel* y *CorelPain* mientras que en algunas BD este vocabulario ofrece resultados similares. Esto se explica porque el vocabulario $q9$ tiene similar tamaño que los vocabularios $q13$ y $q16$, todos ellos utilizan la misma función para cuantificar los atributos de color y se diferencian únicamente en la función de cuantificación de los atributos de forma.

En esta representación se observa que el tamaño del vocabulario no es especialmente crítico. Tomando el vocabulario $q14$ con menos de la mitad de términos con respecto al vocabulario $q16$ (43008 y 96768 términos respectivamente) la eficacia baja del orden del 1 al 3 %. Con el vocabulario $q10$ con menos de la quinta parte de términos con respecto al vocabulario $q13$ (20480 y 96768 términos respectivamente) en este caso la eficacia baja en el peor de los casos hasta el 8 %, siendo la media un 4 %.

Respecto a los espacios de color utilizados, *HSI-Carron* y *HSV-Smith*, ofrecen resultados similares, siendo ligeramente superiores los obtenidos con el espacio *HSV-Smith*. Esto se puede observar en la gráfica de la figura 3.14.

Comparación con otros descriptores

Para realizar la comparativa se han utilizado los mismos descriptores de texturas que los referenciados en el apartado 3.5.3 y se ha calculado el descriptor *JPTD* utilizando el modelo de cuantificación Q_{\otimes}^S , Q_{\otimes}^C con 96768 términos (codificado como $q16$) con el espacio de color *HSV-Smith*. De nuevo para calcular la similitud entre representaciones holísticas se ha utilizado la medida χ^2 (ecuación 3.21).

En la tabla 3.4 se muestra la comparativa en la que se han incluido los resultados del descriptor *TD*. En negrita se ha destacado el mejor resultado obtenido en cada BD.

Comparando únicamente las representaciones holísticas, el descriptor *JTPD* supera al *TD* en 12 de las 15 BD, mientras que en 2 de las BD restantes (las *VisTex*) el resultado es muy similar. Estos resultados demuestran que la incorporación de las relaciones entre los atributos de *textons*, mediante el uso de los espacios perceptuales de los *Textons*, y que la integración de los vocabularios de forma y color consigue representar mejor las texturas que la representación que no integra los atributos (*TD*).

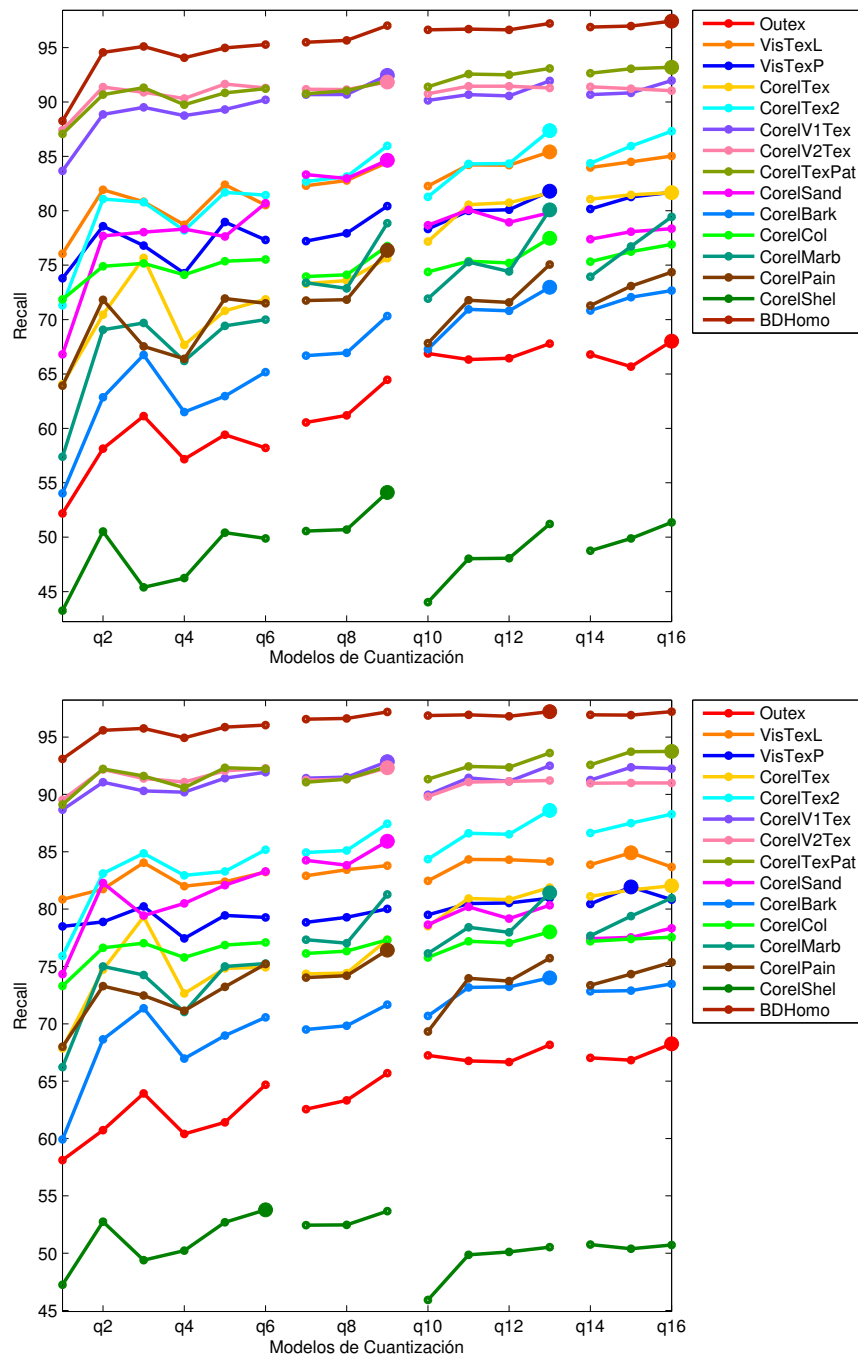


Figura 3.13: Resultados de la recuperación de imágenes con diferentes modelos de cuantificación. Arriba: con el espacio de color *HSI-Carron*. Abajo: con el espacio de color *HSV-Smith*.

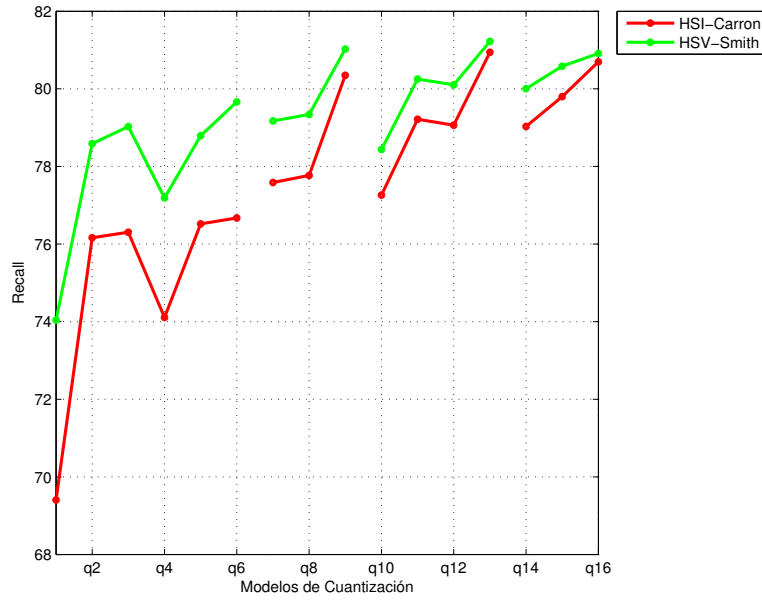


Figura 3.14: Media de la eficiencia de todas la BD para cada modelo de cuantificación.

Tabla 3.4: *Recall promedio* del descriptor *JPTD* y de otros descriptores.

BD	<i>JPTD</i>	<i>TD</i>	<i>HTD+SCD</i>	<i>HTD+CSD</i>	<i>EHD+CSD</i>	<i>LBP₁</i>	<i>LBP₂</i>	<i>LBP₃</i>	<i>CCTM</i>	<i>CDWF</i>
<i>Outex</i>	68.25	75.39	61.10	66.02	67.64	56.87	46.82	60.80	70.00	-
<i>VisTexL</i>	83.68	84.01	87.41	83.45	85.25	73.20	50.36	66.11	-	91.3
<i>VisTexP</i>	80.82	80.35	83.44	80.24	81.81	70.38	45.85	64.69	-	-
<i>CorelTex</i>	82.03	78.78	67.33	72.36	75.31	61.89	60.83	62.22	-	-
<i>CorelTex2</i>	88.28	84.58	76.11	85.89	87.33	72.50	68.89	72.42	-	-
<i>CorelV1Tex</i>	92.25	91.03	85.94	91.89	92.78	77.53	72.14	74.56	-	-
<i>CorelV2Tex</i>	91.00	90.58	88.53	85.89	92.28	81.47	79.81	81.44	-	-
<i>CorelTexPat</i>	93.75	91.56	93.69	89.86	90.50	86.89	82.92	86.03	-	-
<i>CorelSand</i>	78.33	75.75	72.00	85.06	85.28	55.54	60.17	59.89	-	-
<i>CorelBark</i>	73.47	63.69	64.83	69.11	70.36	48.39	51.69	51.67	-	-
<i>CorelCol</i>	77.55	76.24	73.43	72.78	75.52	68.89	68.73	70.62	-	-
<i>CorelMarb</i>	80.97	78.33	70.67	80.11	82.03	59.78	63.72	65.19	-	-
<i>CorelPain</i>	75.36	72.28	62.31	72.61	73.94	50.67	49.11	51.69	-	-
<i>CorelShel</i>	50.72	48.78	41.50	47.06	49.97	32.83	35.94	35.67	-	-
<i>BDHomo</i>	97.22	97.40	97.85	97.78	98.21	91.11	85.28	84.93	-	-
Media	80.91	79.25	75.07	78.67	80.54	65.85	61.47	64.85	-	-

Comparando todos los descriptores se puede observar que los descriptores holísticos propuestos superan al resto de descriptores en 8 BD y solamente en 2 BD (*VisTexL* y *CorelSand*) estos son superados significativamente por otros descriptores. En las BD donde el descriptor holístico *JPTD* es superado por la combinación de descriptores MPEG-7 (*EHD+CSD*) los resultados de ambos descriptores son muy similares. Sin embargo, observando la media de la eficiencia en todas las BD el descriptor *JPTD* supera a todos los descriptores

3.7. Descriptor Perceptual de *Textons* (PTD)

Con el descriptor que se presenta a continuación el objetivo es estudiar el efecto de eliminar la co-ocurrencia entre color y la forma a la hora de representar las texturas en color. Para ello, en esta nueva representación holística se utilizan las mismas características que en el descriptor *JPTD* pero la diferencia estriba en el vocabulario utilizado para representar la imagen. En este caso se utilizan dos vocabularios, uno para describir la forma y otro para describir el color de los blobs, eliminando así la co-ocurrencia entre la forma y el color. De esta manera se asume que una clase de textura color se puede caracterizar por las formas de los blobs que contiene y también por los colores que tienen los blobs.

3.7.1. Vocabulario y Representación de la imagen

A partir de los vocabularios construidos en el apartado 3.6.2 y utilizando la misma notación que en el apartado 3.6.3, la representación holística de una imagen se define a partir de las probabilidades marginales de la variable aleatoria bidimensional (C, S) que se calculan de la siguiente manera:

$$\begin{aligned} P_C(c) &= \sum_{s_j} P_{C,S}(c_i, s_j) \equiv \sum_{s_j} P[C = c_i, S = s_j] \\ P_S(s) &= \sum_{c_i} P_{C,S}(c_i, s_j) \equiv \sum_{c_i} P[C = c_i, S = s_j] \end{aligned} \quad (3.30)$$

siendo P_C la probabilidad marginal de variable discreta C y P_S es la probabilidad marginal de la variable discreta S .

El descriptor que da lugar a esta representación, el *descriptor perceptual de textons (PTD)*, es el resultado de concatenar las probabilidades marginales de la variable bidimensional (C, S) ,

$$\begin{aligned} PTD &= [PTD_s PTD_c] \\ PTD_s &= [P_S(s_1), \dots, P_S(s_N)] \\ PTD_c &= [P_C(c_1), \dots, P_C(c_M)] \end{aligned} \quad (3.31)$$

Dado que de acuerdo con la función de cuantificación Q_{Δ}^C , la variable aleatoria C puede tener M posibles valores y la variable S , obtenida al cuantificar los atributos forma con la función Q_{Δ}^S , puede tener N posibles valores, el vocabulario utilizado en esta representación está formado por la unión del vocabulario para describir el color y el vocabulario para describir la forma, siendo su tamaño de $M + N$ términos. En la figura 3.15 se muestra un ejemplo del descriptor *PTD* y los

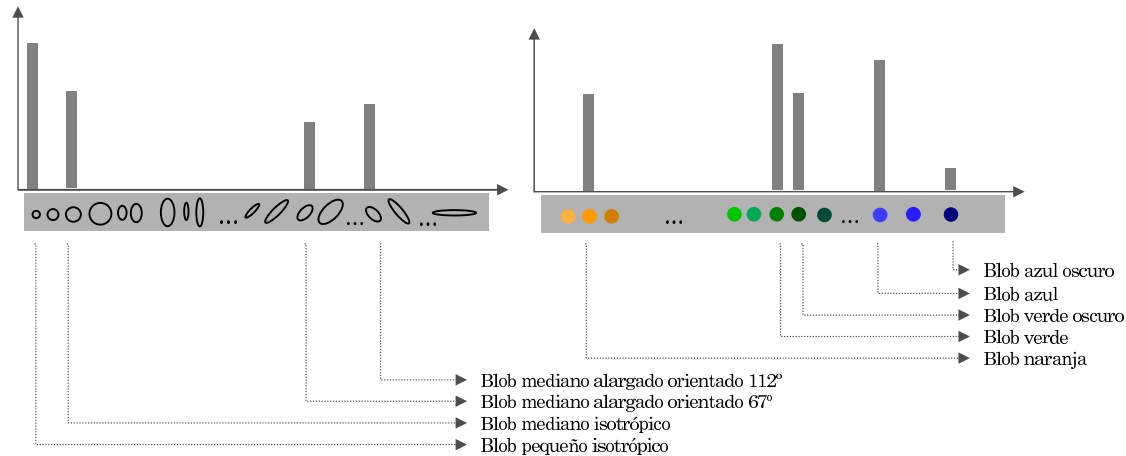


Figura 3.15: Descriptor Perceptual de *Textons* (PTD).

términos lingüísticos que están asociados a los términos del vocabulario. Como se puede observar, cada término tiene un significado perceptual ya que describe la apariencia de los blobs de acuerdo con su forma y su color.

3.7.2. Experimento 3: Evaluación del *PTD*

Para la evaluación del rendimiento de este descriptor se ha seguido el procedimiento descrito en el apartado 3.5.3. Los vocabularios probados se han obtenido de nuevo con diversas funciones de cuantificación y utilizando los espacios de la familia HSI para representar el color de los blobs (*HSI-Carron*, *HSV-Yagi* y *HSV-Smith*), tal y como se ha indicado en el apartado 3.6.2. A continuación, para cada imagen se calcula la similitud de su representación con las del resto de imágenes de la BD utilizando para ello la distancia χ^2 (indicada en la ecuación 3.21). Para valorar la calidad de la representación obtenida con el descriptor *PTD* de nuevo se ha calculado el \overline{recall} en todas las BD de texturas.

Resultados experimentales

Al igual que con el descriptor *JPTD* en esta experimentación se han realizado diversas pruebas utilizando diferentes vocabularios al variar las funciones de cuantificación Q_{Δ}^S y Q_{Δ}^C según los modelos descritos en el apartado 3.6.2. Las pruebas realizadas se muestran en la tabla 3.5 donde se indica la codificación de éstas. Así mismo en las figuras 3.16 y 3.17 se encuentran representadas de manera gráfica la eficiencia obtenida en cada prueba ($\overline{recall}(r)$, siendo r el número de imágenes relevantes de cada BD). Esta información se encuentra más detallada en las tablas A.17-A.32 que figuran en el apéndice A.

En las gráficas de la figura 3.16 para cada BD se han destacado con un punto mayor el vocabulario donde se ha obtenido la eficiencia máxima y las discontinuidades de las líneas que unen los resultados de las diferentes pruebas muestran un cambio de modelo de cuantificación.

Tabla 3.5: Codificación de los vocabularios testeados.

Modelos de cuantización	Tamaño del vocabulario						
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3).N, N+M$	(5,5,5),125, 611	(7,7,7),343, 829	(9,9,9),729, 1458	(6,6,6),216, 1512		
	$(M_1, M_2, M_3).M$	(9,9,6),486	(9,9,6),486	(9,9,9),729	(12,12,9),1296		
	<i>HSI-Carron</i>	q_1	q_2	q_3	q_4		
	<i>HSV-Yagi</i>	q_1	q_2	q_3	q_4		
<i>HSV-Smith</i>	q_1	q_2	q_3	q_4			
$Q_{\#}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3).N, N+M$	(5,5,5),125, 681	(7,7,7),343, 1639	(8,8,8),512, 1808			
	$(M_1, M_2, M_3).M$	(4,16,4),256	(9,16,9),1296	(9,16,9),1296			
	<i>HSI-Carron</i>	q_5	q_6	q_7			
	<i>HSV-Yagi</i>	q_5	q_6	q_7			
<i>HSV-Smith</i>	q_5	q_6	q_7				
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3).N, N+M$	(4,8,5),160, 288	(4,8,5),160, 416	(4,8,7),224, 480	(4,8,7),224, 1520		
	$(M_1, M_2, M_3).M$	(4,8,4),128	(4,16,4),256	(4,16,4),256	(9,16,9),1296		
	<i>HSI-Carron</i>	q_8	q_9	q_{10}	q_{11}		
	<i>HSV-Yagi</i>	q_8	q_9	q_{10}	q_{11}		
<i>HSV-Smith</i>	q_8	q_9	q_{10}	q_{11}			
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3).N, N+M$	(3,8,7),168, 424	(3,8,7),168, 744	(3,8,7),168, 1192	(3,8,7),168, 1464	(4,8,7),224, 1520	(3,8,7),168, 1768
	$(M_1, M_2, M_3).M$	(4,16,4),256	(6,16,6),576	(8,16,8),1024	(9,16,9),1296	(9,16,9),1296	(10,16,10),1600
	<i>HSI-Carron</i>	q_{12}	q_{13}	q_{14}	q_{15}	q_{16}	q_{17}
	<i>HSV-Yagi</i>	q_{12}	q_{13}	q_{14}	q_{15}	q_{16}	q_{17}
<i>HSV-Smith</i>	q_{12}	q_{13}	q_{14}	q_{15}	q_{16}	q_{17}	

Con respecto al uso de diferentes funciones de cuantificación, a raíz de los resultados se puede concluir que el mejor modelo de cuantificación prácticamente en todas las BD ha sido el codificado como q_{16} , que corresponde al modelo $Q_{\otimes}^S, Q_{\otimes}^C$ con 1520 términos (corresponde a una cuantificación cilíndrica para los atributos forma y color). Con este vocabulario en 13 de las 16 BD se consigue el mejor rendimiento y en las 3 BD *CorelTex*, *CorelSand* y *CorelMarb* con un vocabulario muy similar (q_{17} y q_{14} respectivamente). De todas formas la determinación del tamaño idóneo del vocabulario no es un aspecto crítico para la representación de las texturas porque los resultados no varían muy significativamente si se varía el tamaño de éste. A modo de ejemplo tomando el vocabulario q_{12} con menos de una cuarta parte de términos con respecto al vocabulario q_{17} (424 y 1768 términos respectivamente) el rendimiento baja una media del 3%.

Sobre el espacio de color utilizado para representar el color, en todas las BD *Corel* se obtiene un resultado ligeramente superior si se utiliza el espacio de color *HSV-Smith* en lugar del espacio *HSI-Carron*, justamente en estas BD el color parece ser el elemento más discriminatorio, véase la tabla 3.7, de ahí que sean más sensibles al espacio de color utilizado. Respecto al espacio de color *HSV-Yagi*, en muchos casos produce los peores resultados. La variación del rendimiento de la representación en función del espacio de color se puede observar en la gráfica de la figura 3.14, donde se confirma que en media el mejor espacio de representación es el *HSV-Smith*.

Conclusiones

Como era de esperar, la BD donde se ha conseguido un mejor resultado ha sido la BD *BDHomo* debido a la homogeneidad de sus conjuntos relevantes. En general todas las BD *Corel* contienen texturas que no son similares entre ellas, por tanto hay una gran variabilidad entre clases de

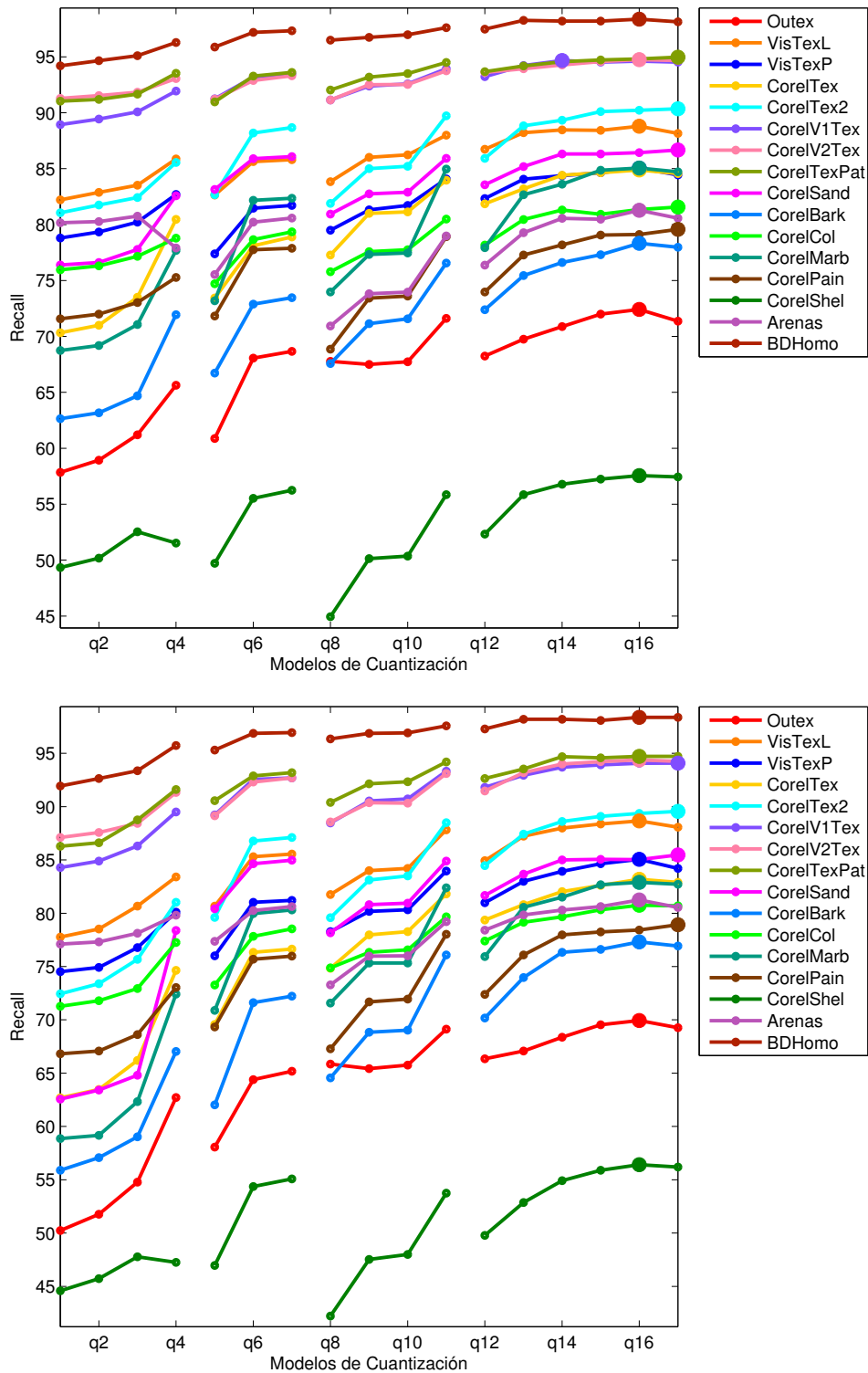


Figura 3.16: Resultados de la recuperación de imágenes con diferentes modelos de cuantización. Arriba: con el espacio de color *HSI-Carron*. Abajo: con el espacio de color *HSV-Yagi*.

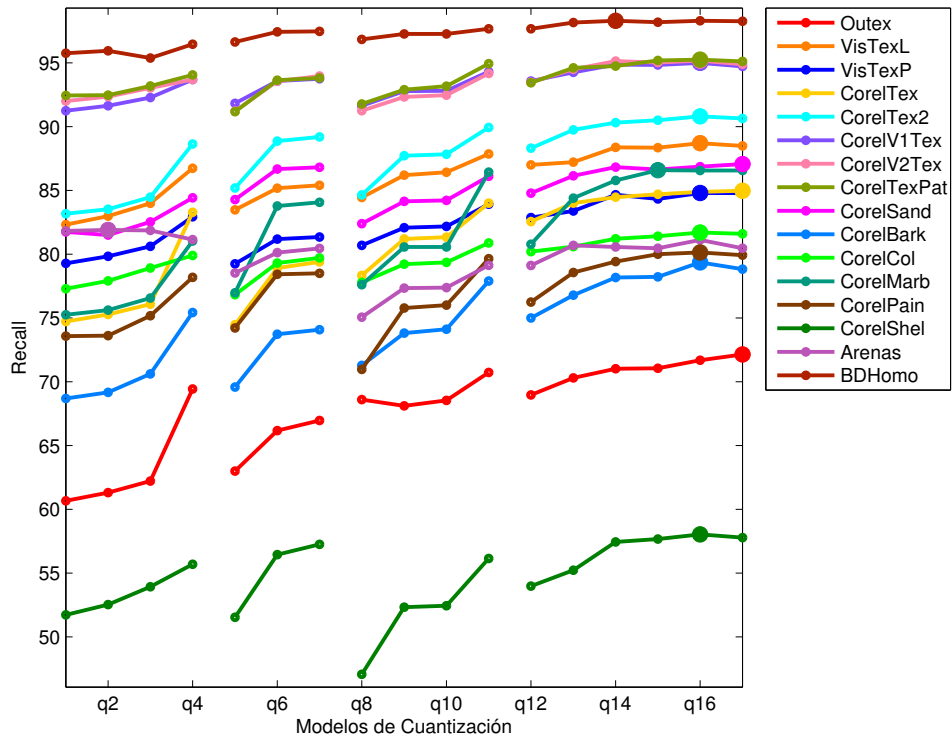


Figura 3.17: Resultados de la recuperación de imágenes utilizando el espacio de color *HSV-Smith* con diferentes modelos de cuantificación.

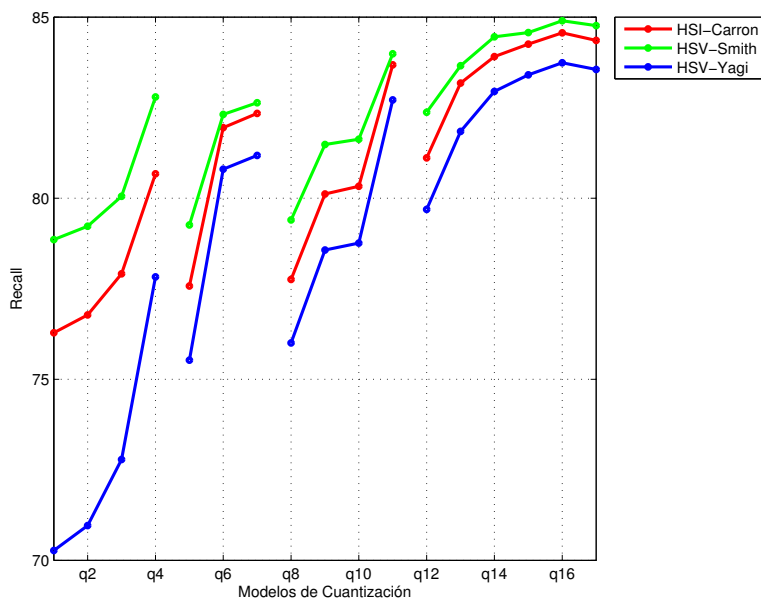


Figura 3.18: Media de la eficiencia de todas la BD para cada modelo de cuantificación.

Tabla 3.6: *Recall promedio* del descriptor *PTD* y de otros descriptores.

BD	<i>PTD</i>	<i>JPTD</i>	<i>TD</i>	<i>HTD+SCD</i>	<i>HTD+CSD</i>	<i>EHD+CSD</i>	<i>LBP₁</i>	<i>LBP₂</i>	<i>LBP₃</i>	<i>CCTM</i>	<i>CDWF</i>
<i>Outex</i>	71.69	68.25	75.39	61.10	66.02	67.64	56.87	46.82	60.80	70.00	-
<i>VisTexL</i>	88.70	83.68	84.01	87.41	83.45	85.25	73.20	50.36	66.11	-	91.3
<i>VisTexP</i>	84.80	80.82	80.35	<i>83.44</i>	80.24	81.81	70.38	45.85	64.69	-	-
<i>CorelTex</i>	84.89	<i>82.03</i>	78.78	67.33	72.36	75.31	61.89	60.83	62.22	-	-
<i>CorelTex2</i>	90.81	<i>88.28</i>	84.58	76.11	85.89	87.33	72.50	68.89	72.42	-	-
<i>CorelV1Tex</i>	95.00	92.25	91.03	85.94	91.89	<i>92.78</i>	77.53	72.14	74.56	-	-
<i>CorelV2Tex</i>	95.17	91.00	90.58	88.53	85.89	<i>92.28</i>	81.47	79.81	81.44	-	-
<i>CorelTexPat</i>	95.25	<i>93.75</i>	91.56	93.69	89.86	90.50	86.89	82.92	86.03	-	-
<i>CorelSand</i>	86.86	78.33	75.75	72.00	85.06	<i>85.28</i>	55.54	60.17	59.89	-	-
<i>CorelBark</i>	79.36	<i>73.47</i>	63.69	64.83	69.11	70.36	48.39	51.69	51.67	-	-
<i>CorelCol</i>	81.70	<i>77.55</i>	76.24	73.43	72.78	75.52	68.89	68.73	70.62	-	-
<i>CorelMarb</i>	86.56	80.97	78.33	70.67	80.11	<i>82.03</i>	59.78	63.72	65.19	-	-
<i>CorelPain</i>	80.14	<i>75.36</i>	72.28	62.31	72.61	73.94	50.67	49.11	51.69	-	-
<i>CorelShel</i>	58.03	<i>50.72</i>	48.78	41.50	47.06	49.97	32.83	35.94	35.67	-	-
<i>Arenas</i>	81.12	-	75.70	56.99	80.09	<i>80.21</i>	52.06	63.29	61.47	-	-
<i>BDHomo</i>	98.30	97.22	97.40	97.85	97.78	<i>98.21</i>	91.11	85.28	84.93	-	-
Media	84.90	<i>80.91</i>	79.03	73.94	78.76	80.52	64.99	61.59	65.38	-	-

texturas, esto se ve reflejado en los buenos resultados conseguidos por el descriptor. Las BD donde los resultados son peores son la BD *Outex* y *CorelShel*. En la primera BD esto es debido a que contiene texturas perceptualmente muy similares: las etiquetadas con el nombre *barleyrice* y algunas *granite*. Sin embargo, en la BD *CorelShel* la confusión viene provocada por la enorme falta de homogeneidad de sus conjuntos relevantes.

Comparación con otros descriptores

Con el objetivo de poder contrastar los resultados obtenidos con el descriptor *PTD* en la tabla 3.6 se encuentran los resultados obtenidos con los descriptores de texturas referenciados en el apartado 3.5.3 entre otros. También se encuentran de nuevo los resultados de los descriptores *TD* y *JPTD* ya mostrados en apartados anteriores. En la citada tabla figuran los resultados obtenidos en las 16 BD, donde para cada BD se ha destacado en negrita el mejor resultado obtenido y en cursiva el segundo mejor resultado. El vocabulario del descriptor *PTD* se ha obtenido con la cuantificación $q16$ correspondiente al modelo $Q_{\otimes}^S Q_{\otimes}^C$ con 1520 términos y el espacio de color *HSV-Smith*.

Comparando únicamente las representaciones holísticas se observa que el descriptor *PTD* es el mejor en todas las BD, excepto en la BD *Outex* donde es superado por el descriptor *TD*. El descriptor *PTD* es más eficiente que el *JPTD* ya que ofrece mejores resultados a costa de un vocabulario mucho menor (1520 términos el *PTD* y 96768 términos el *JPTD*). Comparando todos los descriptores, en todas las BD excepto en las BD *VisTexL* y *Outex* el descriptor holístico *PTD* supera a todos los descriptores. El segundo mejor descriptor (indicado en cursiva) es el descriptor MPEG-7 *EHD+CSD* y el descriptor holístico *JPTD* cuyos resultados son similares en muchos casos al descriptor *EHD+CSD*.

Considerando la eficiencia media (última columna de la tabla) el descriptor *PTD* supera en un 4% al segundo mejor descriptor *JPTD*.

Se puede concluir que el descriptor *PTD* tiene un buen comportamiento en todos los tipos de

Tabla 3.7: Aportación de cada componente del descriptor PTD en el \overline{recall} .

BD	PTD	PTD_s	PTD_c
<i>Outex</i>	72.42	48.18	63.19
<i>VisTexL</i>	88.79	63.49	81.61
<i>VisTexP</i>	85.05	62.37	78.44
<i>CorelTex</i>	84.89	56.78	75.86
<i>CorelTex2</i>	90.22	59.58	84.78
<i>CorelV1Tex</i>	94.61	55.72	91.61
<i>CorelV2Tex</i>	94.75	67.39	90.39
<i>CorelTexPat</i>	94.81	73.89	92.14
<i>CorelSand</i>	86.44	36.50	84.78
<i>CorelBark</i>	78.33	46.33	71.06
<i>CorelCol</i>	81.34	58.04	74.28
<i>CorelMarb</i>	85.06	44.72	80.50
<i>CorelPain</i>	79.11	40.30	76.33
<i>CorelShel</i>	57.76	32.00	52.50
<i>Arenas</i>	81.28	37.41	76.67
<i>BDHomo</i>	98.37	86.88	95.73
Media	84.57	54.34	79.36

texturas utilizados en la experimentación y por lo tanto se puede considerar un descriptor general de texturas en color.

Experimento 4

De manera adicional se ha creído interesante medir la aportación individual de cada una de las componentes del descriptor PTD en el cálculo del \overline{recall} promedio, la componente que modela la distribución de las formas de los blobs, PTD_s y la parte que modela la distribución del color de los blobs, PTD_c . Para ello se ha repetido la experimentación utilizando el descriptor PTD_c y en otro experimento el descriptor PTD_s . Para el descriptor PTD se ha utilizado la cuantificación $q16$, correspondiente al modelo $Q_{\otimes}^S Q_{\otimes}^C$ con 1520 términos y el espacio de color $HSI-Carron$, ya que es uno de los que produce los mejores resultados. El desglose de este descriptor da lugar al descriptor PTD_c con el modelo Q_{\otimes}^C de 1296 términos y al descriptor PTD_s con el modelo Q_{\otimes}^S de 224 términos. En la tabla 3.7 se muestran los resultados obtenidos.

Comparando la aportación de cada una de las componentes (PTD_c y PTD_s), en la tabla 3.7 se ha destacado en negrita la que es mayor. Como se puede observar, la componente PTD_c proporciona una eficiencia siempre superior que la obtenida con la componente PTD_s en todas las BD. En algunas BD (*Arenas*, *CorelSand*, *CorelMarb* y *CorelPain*) la aportación de la componente PTD_c es mucho mayor que la de PTD_s y por tanto tiene mayor poder discriminatorio, esto se explica porque estas BD se caracterizan por tener imágenes cuyas formas son similares pero con colores diferentes.

Experimento 5

Dado que las texturas se representan con modelos estadísticos, en este experimento se ha estudiado el efecto de usar diferentes funciones para calcular la similitud entre representaciones a la hora de aplicar el descriptor *PTD* para recuperar imágenes. Las medidas utilizadas que permiten comparar distribuciones de densidad de probabilidad son las siguientes:

1. La función χ^2 proporciona una medida de similitud entre dos distribuciones de probabilidad. Dadas las distribuciones de probabilidad H_1 y H_2 la distancia se define:

$$\chi^2(H_1, H_2) = \frac{1}{2} \sum_{t=1}^T \frac{[H_1(t) - H_2(t)]^2}{H_1(t) + H_2(t)}. \quad (3.32)$$

2. La distancia Bhattacharyya. Considerando cada distribución como un vector n-dimensional tiene una interpretación geométrica directa. Es el coseno del ángulo que forman los dos vectores que representan las distribuciones. Ésta se calcula:

$$\text{distancia } B(H_1, H_2) = \sum_{t=1}^T \sqrt{H_1(t) \times H_2(t)}. \quad (3.33)$$

3. Divergencia Jeffrey, es una divergencia empíricamente derivada de la divergencia Kullback-Leibler, ésta se puede interpretar como la ineficiencia de transformar una distribución en otra. Se calcula de la siguiente manera:

$$\text{divergencia } J(H_1, H_2) = \sum_{t=1}^T H_1(t) \log \frac{2H_1(t)}{H_1(t) + H_2(t)} + H_2(t) \log \frac{2H_2(t)}{H_1(t) + H_2(t)}. \quad (3.34)$$

Para obtener el vocabulario del descriptor *PTD* se ha utilizado la cuantificación *q16* correspondiente al modelo $Q_{\otimes}^S Q_{\otimes}^C$ con 1520 términos y el espacio de color *HSI-Carron*. En la tabla 3.8 se muestran los resultados obtenidos sobre todas las BD donde se ha destacado en negrita el mejor resultado obtenido en cada BD.

En la tabla 3.8 se demuestra que el uso de diferentes funciones de similitud a la hora de comparar las representaciones apenas afecta a la eficiencia obtenida. La divergencia Jeffrey es la función de similitud que por poco margen supera a la distancia χ^2 en la mayoría de los casos.

3.8. Aplicación a la clasificación de texturas

La clasificación de texturas constituye una de las aplicaciones más utilizadas para estudiar la viabilidad de los descriptores. Esta aplicación requiere la realización de dos fases: aprendizaje y clasificación. En la fase de aprendizaje o entrenamiento se utiliza un subconjunto de imágenes del conjunto relevante de cada clase de textura para entrenar el sistema. El conjunto relevante de una textura, al igual que en las aplicaciones de recuperación de imágenes, se ha construido

Tabla 3.8: Uso de diferentes funciones de similitud.

BD	Distancia χ^2	Divergencia Jeffrey	Distancia Bhattacharyya
<i>Outer</i>	72.42	72.59	72.10
<i>VisTexL</i>	88.79	89.01	88.90
<i>VisTexP</i>	85.05	85.38	85.50
<i>CorelTex</i>	84.89	85.14	84.83
<i>CorelTex2</i>	90.22	90.50	90.81
<i>CorelV1Tex</i>	94.61	94.81	94.31
<i>CorelV2Tex</i>	94.75	94.89	94.50
<i>CorelTexPat</i>	94.81	94.81	94.56
<i>CorelSand</i>	86.44	86.61	86.47
<i>CorelBark</i>	78.33	78.19	77.75
<i>CorelCol</i>	81.34	81.44	80.72
<i>CorelMarb</i>	85.06	85.56	85.69
<i>CorelPain</i>	79.11	80.03	80.08
<i>CorelShel</i>	57.76	57.97	57.69
<i>Arenas</i>	81.28	81.40	81.33
<i>BDHomo</i>	98.37	98.44	98.37
Media	84.57	84.79	84.6

sub-dividiendo la textura en porciones disjuntas. El resto de imágenes del conjunto relevante que no se ha usado en el aprendizaje es utilizado para llevar a cabo la clasificación o test.

Tanto para el descriptor *JPTD* como para el descriptor *PTD* se han utilizado dos métodos para realizar la clasificación:

- Clasificador con vecino más cercano: dada una imagen del conjunto de test representada por su descriptor, se calcula la distancia de éste a todos los descriptores de las imágenes de entrenamiento. La clase donde se clasificará la imagen corresponde a la clase a la cual pertenece la imagen de entrenamiento que se encuentra más cercana.
- Clasificador con modelo más cercano: para cada clase de textura se construye su modelo estadístico. En éste, se contabilizan de manera conjunta las frecuencias de los términos del vocabulario obtenidas en todas las imágenes que forman parte de la clase de entrenamiento de la textura, dando lugar al descriptor del modelo de la clase. Para realizar la clasificación de una imagen de test se calculan las distancias de ésta a todos los modelos de texturas y se clasifica en la clase de textura cuyo modelo se encuentra más cercano.

A la hora de determinar las imágenes que formaran parte del conjunto de entrenamiento se ha reproducido la misma metodología utilizada por Lazebnik y cols. (2005), para compensar los efectos de la variabilidad dentro de las clases repite la clasificación 20 veces, escogiendo cada vez las imágenes de entrenamiento de forma aleatoria. Para evaluar la clasificación se calcula el porcentaje de imágenes correctamente clasificadas y se calcula el promedio de las 20 clasificaciones.

Tabla 3.9: Clasificación con diversos descriptores.

BD	<i>JPTD-mc</i>	<i>PTD-mc</i>	<i>JPTD-vc</i>	<i>PTD-vc</i>	<i>EHD+CSD</i>	<i>HTD+CSD</i>	<i>HTD+SCD</i>	<i>LBP₁</i>	<i>LBP₃</i>	<i>MM</i>	<i>WDHMM</i>	<i>ML</i>
<i>Outex</i>	83.62	87.77	84.71	90.32	90.04	86.71	75.78	71.51	76.07	94.1	85.15	79.86
<i>VisTexL</i>	98.01	99.08	97.35	99.25	99.13	99.56	95.69	94.12	87.67	-	-	-
<i>VisTexP</i>	94.88	97.50	98.52	98.89	98.24	98.53	94.66	92.11	95.95	97.9	96.80	-
<i>CorelTex</i>	92.07	94.25	92.02	93.73	86.88	80.90	65.45	73.78	75.85	-	-	-
<i>CorelTex2</i>	95.57	96.12	96.88	97.78	95.38	88.97	75.38	84.57	83.32	-	-	-
<i>CorelV1Tex</i>	96.27	98.07	96.13	97.63	96.83	94.15	84.04	88.12	87.15	-	-	-
<i>CorelV2Tex</i>	96.53	98.47	96.08	98.77	96.98	95.07	86.45	90.90	90.57	-	-	-
<i>CorelTexPat</i>	97.77	97.68	97.42	98.48	95.28	97.87	94.98	96.35	94.47	-	-	-
<i>CorelSand</i>	88.83	94.47	80.88	92.83	92.10	78.13	63.32	61.05	67.42	-	-	-
<i>CorelBark</i>	84.78	91.07	85.08	91.10	84.95	70.57	56.58	55.78	64.17	-	-	-
<i>CorelCol</i>	85.61	89.43	86.35	89.37	85.57	86.12	73.81	82.43	81.72	-	-	-
<i>CorelMarb</i>	90.38	84.07	90.38	95.45	91.73	80.78	65.70	70.53	76.33	-	-	-
<i>CorelPain</i>	84.95	90.32	86.03	97.98	85.97	72.00	54.42	57.13	59.77	-	-	-
<i>CorelShel</i>	63.88	72.70	59.62	72.32	61.02	38.47	32.82	31.68	32.27	-	-	-
<i>Avenas</i>	90.77	94.54	83.13	94.59	94.81	87.52	78.12	77.44	91.09	-	-	-
<i>BDHomo</i>	99.27	99.54	99.65	99.69	99.60	99.98	96.88	95.98	91.85	-	-	-
Media	90.19	92.81	89.38	94.36	90.90	84.70	74.63	76.46	78.47	-	-	-

3.8.1. Resultados experimentales y evaluación

Se han utilizado los descriptores holísticos *PTD* y *JPTD* para clasificar las 16 BD utilizando los dos modelos de clasificación descritos en el apartado anterior, *PTD-m* identifica el descriptor *PTD* con clasificador por modelo más cercano y *PTD-vc* identifica el descriptor *PTD* con clasificador por vecino más cercano, de manera análoga se identifica *JPTD-m* y *JPTD-vc*. En ambos descriptores se ha usado el modelo de cuantificación Q_{\otimes}^S Q_{\otimes}^C y el espacio de color *HSV-Smith*. La similitud entre representaciones se ha calculado con la distancia χ^2 . El descriptor *PTD* se ha utilizado el vocabulario *q16* con 1520 términos y para el descriptor *JPTD* el vocabulario *q16* con 96768 términos.

Para contrastar los resultados obtenidos se han utilizado los descriptores utilizados en los anteriores experimentos, los descriptores MPEG-7 y los descriptores *LBP(RGB)*. Para estos descriptores se ha utilizado como clasificador el vecino más cercano.

En la tabla 3.9 se muestran los resultados de todos los descriptores sobre las diferentes BD. En ésta se han añadido los resultados de tres descriptores más, el *MM* (Arvis y cols., 2004), el *WDHMM* (Xu y J. Yang, 2005) y el *ML* (Aptoula y Lefevre, 2008). Se ha destacado en negrita el mejor resultado en cada BD.

Comparando únicamente los descriptores *PTD* y *JPTD*, el primero proporciona siempre unos resultados superiores al segundo, aunque en algunas BD (*VisTexP*, *CorelTex*, *CorelTex2*, *CorelV1Tex*, *CorelTexPatt*, *BDHomo*) los resultados son similares. Respecto a los dos modelos de clasificación utilizados en estos dos descriptores, en 5 BD (*CorelTex*, *CorelV1Tex*, *CorelSand*, *CorelCol*, *CorelShel*) la clasificación con modelo más cercano *PTD-mc* supera ligeramente al clasificador por vecino más cercano *PTD-vc*, pero en el conjunto de todas las BD la clasificación por vecino más cercano funciona mejor.

Comparando todos los descriptores con el clasificador más cercano, en 12 de las 16 BD el mejor resultado se ha obtenido con el descriptor *PTD* y en las 4 BD restantes el porcentaje de correcta clasificación obtenida por este descriptor es superada por poco por otros descriptores.

Considerando la eficiencia media (última fila de la tabla) el descriptor *PTD* supera en un 4% a la combinación de descriptores MPEG-7 *EHD+CSD*.

Los resultados demuestran de nuevo que el descriptor *PTD* tiene un buen comportamiento en la caracterización de las texturas.

Capítulo 4

Hacia una representación conceptual de texturas ¹

4.1. Introducción

Tal y como argumentan Tuytelaars y Mikolajczyk (2007) un hecho que es común en la mayoría de los descriptores basados en las representaciones *BoW* es la falta de interpretación semántica. Ésta es una de las ventajas que tiene el descriptor desarrollado en el capítulo anterior, presenta cierta semántica, aunque ésta es pobre y limitada puesto que las palabras visuales hacen referencia a términos muy básicos del lenguaje que describen como son pequeñas partes de la imagen (blobs).

Considerando esta ventaja en este capítulo se persigue un objetivo doble:

- Aumentar el grado de semántica del descriptor holístico proponiendo un descriptor de nivel intermedio. Este nuevo descriptor agrupa conjuntos de palabras visuales atendiendo a su significado perceptual. Para ello se han introducido mecanismos de organización perceptual que agrupan los blobs de la textura atendiendo a propiedades de similitud y proximidad. Estos grupos se identifican con el nombre de *Componentes Texturales*.
- Estudiar la versatilidad del descriptor de nivel intermedio en diferentes aplicaciones de diversa naturaleza: la recuperación de imágenes de textura utilizando como consulta una imagen o a partir de una sentencia que describe el contenido visual a buscar ² y la detección de texturas inmersas en imágenes.

4.2. Representación conceptual básica

Una característica que poseen gran parte de texturas es la repetitividad. Si bien esta característica es más común en el caso de texturas color artificiales, las texturas naturales también poseen

¹Contenido parcialmente publicado en Alvarez, Salvatella, Vanrell, y Otazu (2010a).

²En el resto del capítulo se utilizará el término en inglés *Browsing* para describir esta aplicación.

en mayor o menor grado esta característica. La repetitividad según Rao y Lohse (1996) es una de las dimensiones de la textura que está muy relacionada con otras características de la textura: la uniformidad y la regularidad. La primera hace referencia a la homogeneidad de los elementos que constituyen la textura y la segunda a la organización de estos. Por tanto, según el grado de repetitividad de una textura, así de similares serán las propiedades de los elementos que constituyan las texturas. Partiendo de la asunción que una textura ha de poseer cierto grado de repetitividad se ha desarrollado un modelo de representación de las texturas color basada en esta característica.

Siendo los blobs los elementos perceptuales de partida, se define el concepto semántico *componente textural* de la siguiente manera:

Componente Textural Es el conjunto de blobs que comparten atributos similares de forma y color.

Una textura con un cierto grado de repetitividad puede poseer diferentes componentes texturales. A modo de ejemplo y para ilustrar la idea de este concepto en la figura 4.1 se muestra la descomposición de una imagen en sus componentes texturales. La imagen de la figura 4.1.(a) está compuesta por dos componentes texturales, cada una de ellas compuesta a su vez por blobs alargados dispuestos verticalmente, en un caso con blobs de color gris y en otro caso con blobs de color azul. La imagen de la figura 4.1.(b) es más compleja ya que contiene 6 tipos de componentes texturales de formas y colores diferentes. Al pie de cada imagen se describen con palabras las propiedades de los blobs que forman las componentes texturales.

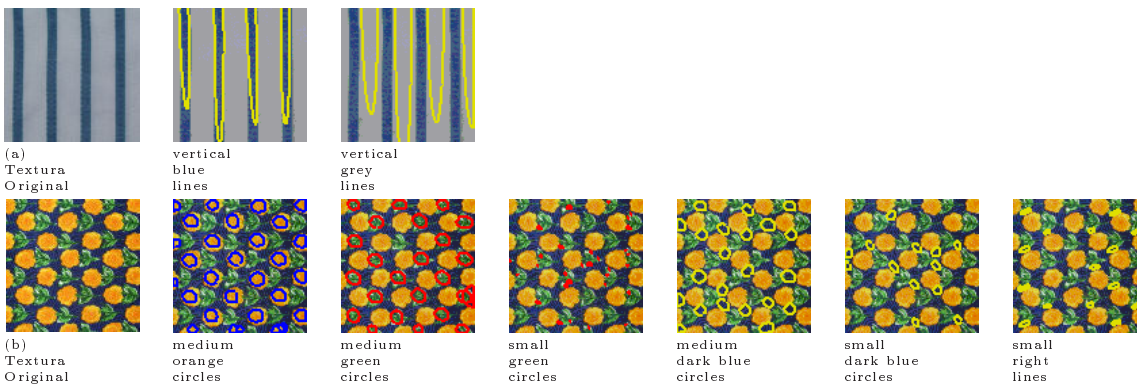


Figura 4.1: Componentes texturales y su descripción.

4.2.1. Agrupación por similitud

Dado que las componentes texturales están compuestas por grupos de blobs con atributos similares, el punto de partida para la obtención de las componentes texturales es la matriz dada en la ecuación 3.12, donde se ha obtenido la descomposición de la textura en sus textons. Esta descomposición presenta cierta redundancia dado que un mismo blob puede ser detectado a la vez en cualquiera de los tres canales del espacio color utilizado en la detección (espacio oponente). Esta

redundancia se ha eliminado filtrando los blobs redundantes utilizado para ello el procedimiento identificado como *filtrado perceptual* descrito en el apartado 3.4.1 para la detección de los blobs perceptuales, pero en este caso los blobs detectados en todos los canales del espacio color son el punto de partida del procedimiento. Como resultado del filtrado se obtienen los blobs perceptuales que se definen en la siguiente ecuación:

$$\mathbf{T}' = [\mathbf{T}'_{\text{loc}} \mathbf{T}'_{\text{sha}} \mathbf{T}'_{\text{col}}] \quad (4.1)$$

A partir de los blobs perceptuales y sus atributos el objetivo es encontrar los grupos de blobs cuyos atributos sean similares ya que ésta es la definición de la componente textural. Para poder agrupar los blobs según su similitud perceptual se han utilizado los espacios perceptuales de Textons definidos en el apartado 3.6.1 donde se representan los atributos de los blobs.

El espacio de textons donde se representan los atributos de forma presenta las siguientes características que relacionan la similitud entre textons:

- Los puntos localizados en un anillo alrededor del eje vertical corresponden a blobs con la misma área y forma con orientaciones diferentes.
- Los puntos localizados en un pequeño arco situado en un plano perpendicular al eje vertical corresponden a blobs que tienen la misma área y forma con orientaciones similares.
- Los puntos localizados a lo largo del eje vertical corresponden a blobs isotrópicos.
- Los puntos localizados a lo largo de un segmento de línea perpendicular al eje vertical corresponden a blobs con la misma área y orientación con diferente relación de aspecto.

Puesto que el espacio donde se representan los atributos de color de los blobs, HSI, es cilíndrico al igual que el espacio de representación de forma de los blobs, la distribución de los colores tiene las siguientes características:

- Los puntos situados en un pequeño arco alrededor del eje de la intensidad corresponden a colores con la misma intensidad y saturación con similar tono.
- Los puntos situados a lo largo de una línea paralela al eje de la intensidad corresponden a colores con la misma saturación y tono con diferente intensidad.
- Los puntos situados en una línea perpendicular al eje de la intensidad corresponden a colores con el mismo tono e intensidad y diferente saturación.
- Los puntos situados en el eje de la intensidad ($S = 0$) corresponden a grises.

Dadas las características de los espacios de textons, formas y colores similares corresponden a puntos del espacio localizados en diferentes variedades unidimensionales: líneas, arcos o anillos. Por tanto, para agrupar blobs con formas y colores similares es necesario utilizar un método de agrupamiento que sitúe los puntos con estas distribuciones dentro del mismo grupo. Además, este

método debe permitir combinar características representadas en espacios diferentes, como son los espacios de textons. El método de agrupamiento que tiene estas propiedades es el *Normalized Cut* (*N-cut*).

Normalized Cut

El algoritmo *N-cut* fue desarrollado por Shi y Malik (2000) para agrupar características perceptuales ya que basa la decisión de agrupar los elementos en su grado de similitud. Con este método se obtienen los grupos o clusters biparticionando un grafo de manera recursiva hasta que el parámetro llamado *n-cut* excede un límite determinado.

La decisión de dividir o no el grafo se basa en minimizar el coste de la propia división y ésta depende de los pesos de los arcos entre los nodos. Los nodos del grafo son los puntos del espacio de características y los arcos entre nodos tienen como peso la similitud entre nodos.

Dado un grafo $G = (V, E)$ donde V son los nodos del grafo y E los arcos entre nodos, la subdivisión de G en dos subconjuntos A y B , tal que $A \cup B = V$ y $A \cap B = \emptyset$, tiene el siguiente coste:

$$\text{N-cut}(A, B) = \frac{\text{cut}(A, B)}{\text{assoc}(A, V)} + \frac{\text{cut}(A, B)}{\text{assoc}(B, V)} \quad (4.2)$$

donde:

$$\begin{aligned} \text{cut}(A, B) &= \sum_{u \in A, v \in B} w(u, v) \\ \text{assoc}(A, V) &= \sum_{u \in A, t \in V} w(u, t) \end{aligned} \quad (4.3)$$

siendo $w(u, v)$ el peso asociado a uno de los arcos entre A y B , $\text{cut}(A, B)$ contabiliza los pesos de los arcos que unen A y B que han sido eliminados al hacer la partición, mientras que $\text{assoc}(A, V)$ representa la conexión total de nodos de A a todos los nodos del grafo, $\text{assoc}(B, V)$ se define de manera equivalente. La partición del grafo finaliza cuando el valor $\text{N-cut}(A, B)$ supera un determinado umbral, identificado como *n-cut*.

Para determinar la similitud entre nodos es necesario definir una medida de distancia. Dado que el espacio de atributos de forma ha sido diseñado para ser uniforme y el espacio de representación del color de los blobs es casi uniforme, es razonable utilizar la distancia Euclídea.

El algoritmo de agrupamiento *N-cut* aplicado sobre los espacios de textons genera k grupos tal y como se indica en la siguiente ecuación:

$$\text{NCUT}([U(\mathbf{T}'_{\text{sha}}), \text{HSI}(\mathbf{T}'_{\text{col}})], \mathbf{\Omega}) = \{\hat{\mathbf{T}}^1, \hat{\mathbf{T}}^2, \dots, \hat{\mathbf{T}}^k\} \quad (4.4)$$

donde $\mathbf{\Omega}$ es la matriz de pesos cuyos elementos definen la similitud entre nodos a través del cálculo de la distancia en cada uno de los espacios (forma y color) de manera independiente. Estos pesos

se definen,

$$\Omega_{pq} = \exp - \frac{\|U(\mathbf{T}'_{\text{sha}})_p - U(\mathbf{T}'_{\text{sha}})_q\|_2^2}{\sigma_{\text{sha}}^2} \cdot \exp - \frac{\|HSI(\mathbf{T}'_{\text{col}})_p - HSI(\mathbf{T}'_{\text{col}})_q\|_2^2}{\sigma_{\text{col}}^2} \quad (4.5)$$

El valor de Ω_{pq} representa la similitud entre el blob p y el blob q , que depende de la similitud entre sus formas y de la similitud entre sus colores. Los términos $U(\mathbf{T}'_{\text{sha}})_p$ y $HSI(\mathbf{T}'_{\text{col}})_p$ son las p -ésimas filas de las matrices $U(\mathbf{T}'_{\text{sha}})$ y $HSI(\mathbf{T}'_{\text{col}})$, respectivamente. Las funciones U y HSI son las transformaciones perceptuales indicadas en la ecuaciones 3.24 y 3.25 respectivamente.

Shi y Malik (2000) definieron σ como un porcentaje del rango total de la distancia en el espacio de características. En este caso, σ_{sha} está definida en el espacio de atributos de forma y σ_{col} en el espacio de atributos de color. Estos valores se han fijado empíricamente a 12% y 16% del rango de las distancias en el espacio de atributos de forma y color respectivamente.

Como resultado de aplicar el algoritmo de agrupamiento N-cut se obtienen las $\hat{\mathbf{T}}^i$, $\forall i = 1, \dots, k$ (donde k es el número total de grupos) y la i -ésima componente corresponde a la clase i que comprende a:

$$\hat{\mathbf{T}}^i = [\hat{\mathbf{T}}^i_{\text{loc}} \hat{\mathbf{T}}^i_{\text{sha}} \hat{\mathbf{T}}^i_{\text{col}}] \quad (4.6)$$

donde $\hat{\mathbf{T}}^i_{\text{loc}} = [X^i Y^i]$, $\hat{\mathbf{T}}^i_{\text{sha}} = [R^i Z^i \Phi^i]$ y $\hat{\mathbf{T}}^i_{\text{col}} = [H^i S^i I^i]$, siendo $X^i \subset \mathbf{X}$ un subconjunto de \mathbf{X} definido por los elementos pertenecientes al grupo i y el resto de los términos Y^i , R^i , Z^i , Φ^i , H^i , S^i y I^i se definen de manera equivalente.

Cada una de las clases o grupos obtenidos constituyen las *Componentes Texturales (TC)*. Para cada una de ellas se define su descriptor, el *Descriptor de la Componente Textural (TCD)* que está formado por los atributos *forma* y *color* del representante del grupo. Para el grupo i se define el descriptor *TCD*:

$$TCD^i = (r^i, z^i, \phi^i, h^i, s^i, i^i) \quad (4.7)$$

Para calcular el representante de cada grupo, dado que el resultado del clustering no es perfecto y dentro del mismo grupo pueden haber algunos blobs con formas y/o colores diferentes del que poseen la mayoría de los blobs del grupo, se ha utilizado un estimador estadístico robusto a este aspecto. En concreto, el representante del grupo i se calcula a partir de los atributos forma y color de los elementos pertenecientes a su grupo ($[\hat{\mathbf{T}}^i_{\text{sha}} \hat{\mathbf{T}}^i_{\text{col}}]$). Representando estos en sus respectivos espacios en cada eje se calcula la coordenada del representante de la siguiente manera:

$$\text{repre} = 0,25 \text{ quartil1} + 0,25 \text{ quartil2} + 0,5 \text{ mediana}$$

La representación conceptual de una imagen está formada por los descriptores de las compo-

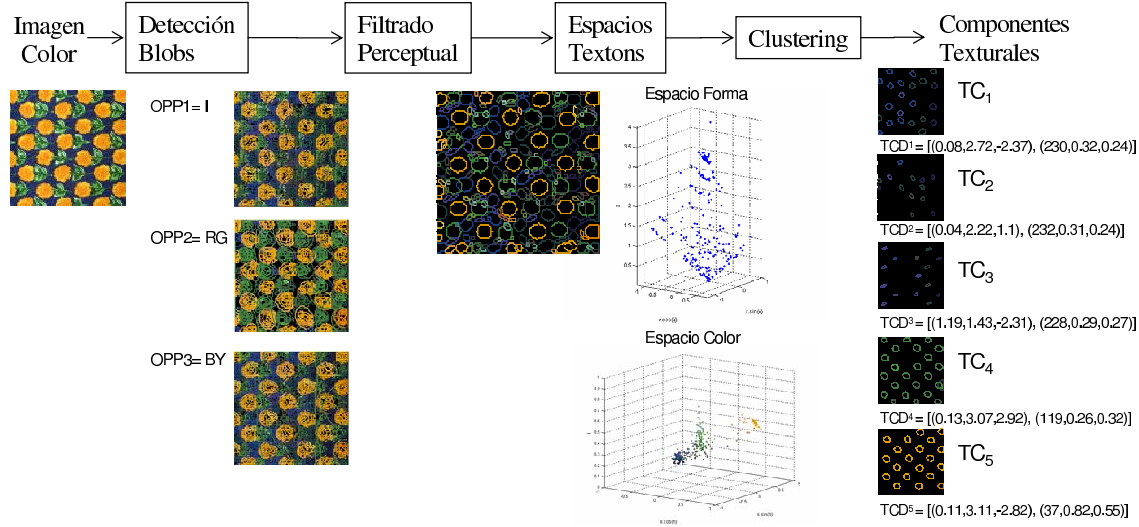


Figura 4.2: Esquema general de la construcción del descriptor TCD .

nentes texturales que contiene,

$$TCD = [TCD^1, \dots, TCD^k] \quad (4.8)$$

En la figura 4.2 se resumen las etapas que configuran la obtención del descriptor TCD que constituye la representación conceptual.

4.2.2. Similitud entre representaciones conceptuales

El número de componentes texturales que puede tener una textura depende de la complejidad de ésta y del número de características (formas y colores) diferentes que presente. Por tanto a la hora de comparar dos representaciones conceptuales hay que considerar que el número de descriptores que puede tener cada una de ellas puede ser diferente. Esta característica es determinante a la hora de decidir la métrica que ha de permitir calcular la similitud entre representaciones.

La métrica que permite calcular la disimilitud entre dos distribuciones multidimensionales de diferente dimensión es la *Earth Mover's Distance (EMD)* (Rubner y cols., 2000). El cálculo de la distancia EMD está basado en la solución del conocido problema del transporte de bienes entre productores y consumidores. En este problema se plantea minimizar el movimiento de bienes entre productores y consumidores de manera que se satisfaga la demanda de los consumidores y se cumplan una serie de restricciones. Siendo I los productores, J los consumidores y c_{ij} el coste del transporte de una unidad, se trata de encontrar el flujo del transporte, f_{ij} , que minimiza el coste del transporte total,

$$\sum_{i \in I} \sum_{j \in J} c_{ij} f_{ij} \quad (4.9)$$

Las restricciones son: el flujo de transporte tiene que ser entre productor y consumidor (no al

revés), el consumidor no puede recibir más bienes que los solicitados, el productor tiene un límite de productos y finalmente que la demanda total no puede exceder la producción total.

Una vez calculado el flujo f_{ij} la distancia EMD constituye el coste normalizado que se expresa de la siguiente manera:

$$EMD(x, y) = \frac{\sum_{i \in I} \sum_{j \in J} c_{ij} f_{ij}}{\sum_{i \in I} \sum_{j \in J} f_{ij}} \quad (4.10)$$

Este método aplicado al problema de calcular la distancia entre dos distribuciones de clusters P y Q , se traduce a calcular el coste de transformar una distribución en la otra. Definiendo la *signatura* de cada cluster como:

$$P = \{(p_1, w_{p_1}), \dots, (p_m, w_{p_m})\}$$

$$Q = \{(q_1, w_{q_1}), \dots, (q_n, w_{q_n})\}$$

siendo w_i el peso del cluster i , m y n el número de clusters de P y Q respectivamente, el valor $c_{ij} = dist(p_i, q_j)$ es la distancia entre dos clusters y las restricciones se expresan en términos de w :

$$\sum_{j=1}^n f_{ij} \leq w_{p_i}, \quad \sum_{i=1}^m f_{ij} \leq w_{q_j}$$

$$\sum_{i=1}^m \sum_{j=1}^n f_{ij} = \min\left(\sum_{i=1}^m w_{p_i}, \sum_{j=1}^n w_{q_j}\right)$$

Queda por definir la distancia entre dos clusters $dist(p_i, q_j)$ que es la distancia entre dos componentes texturales, ésta se define de la siguiente manera:

$$dist(TCD^i, TCD^j) = \alpha \cdot dist_{forma}(TCD^i, TCD^j) + \beta \cdot dist_{color}(TCD^i, TCD^j) \quad (4.11)$$

donde $dist_{forma}$ y $dist_{color}$ son las distancias Euclídeas en los espacios de textons, forma y color respectivamente. Los parámetros α y β son los pesos asignados a cada una de las distancias. Sobre el valor que deben de tener estos pesos Finlayson y Tian (1999) realizaron experimentos psicofísicos para determinar qué es más importante, la textura o el color a la hora de determinar la similitud entre dos imágenes, sus conclusiones apuntan que la similitud percibida es altamente dependiente del observador y del contexto, por tanto dependiendo de estos dos aspectos la importancia de ambos estímulos varia. Previamente al cálculo de $dist_{forma}$ y $dist_{color}$ es preciso normalizar los descriptores TCD de las componentes texturales, para ello se dividen las coordenadas en cada uno de los espacios por la cota máxima de cada coordenada, de esta manera las distancias $dist_{forma}$ y $dist_{color}$ tiene el mismo orden de magnitud.

El peso de una componente textural w_{TC_i} se define en términos del número de blobs de la componente textural respecto al número de blobs de la imagen de la siguiente manera:

$$w_{TC^i} = \frac{n^{\circ}blobs TC^i}{\sum_{i=1}^n n^{\circ}blobs TC^i} \quad (4.12)$$

siendo n el número de TC de una textura.

4.2.3. Vocabulario básico de texturas en color

Una característica importante del descriptor TCD es que puede ser traducido a un vocabulario de términos básicos del lenguaje natural. Se propone utilizar términos básicos del Inglés para describir la geometría y fotometría de los blobs de la imagen. Dado que una textura está formada por una lista de Componentes Texturales, cada una definida por los atributos de sus blobs, entonces se puede construir un sintagma para describir estos atributos. Por tanto, la descripción completa de la textura está formada por una lista de sintagmas que describen las componentes de la textura. Aún cuando esta descripción no permite cubrir toda la brecha semántica, constituye un paso importante al proporcionar propiedades semánticas que son nuevas en el contexto de la investigación de texturas.

A continuación, se detalla el procedimiento para traducir de manera automática el descriptor TCD (de la ecuación 4.8) a un sintagma. Para ello, primero se introducen los términos básicos que componen el vocabulario para describir los atributos de los blobs,

Color 22 términos, de estos 11 son los términos básicos definidos por Berlin y Kay (1969) y modelados por Benavente y cols. (2006). En el resto se incluyen los mismos 11 términos básicos pero con el modificador *-ish*.

Intensidad 2 términos diferentes (Dark o Light) para modificar los términos básicos del color. Estos se calculan de manera específica en cada uno de los colores utilizando únicamente la componente Intensidad del espacio de color.

Forma 2 términos básicos para describir la forma de los blobs, estos son “non-oriented”, para referirse a los blobs isotrópicos, y “oriented” para referirse a los blobs alargados.

Tamaño 3 términos básicos para describir el tamaño de los blobs: “small” (pequeño), “medium” (medio) y “large” (grande).

Orientación 4 modificadores diferentes para describir la orientación de los blobs alargados, se ha simplificado todo el rango de orientaciones a 4 usando los siguientes términos: horizontal ($\approx 0^\circ$), right ($\approx 45^\circ$), vertical ($\approx 90^\circ$) y left ($\approx 135^\circ$).

A continuación se proporciona la sintaxis para traducir sistemáticamente el TCD a un sintagma, éste se da en la gramática BNF (forma Backus Naur) usando el vocabulario previamente definido:

```
Texture_description := Texture_component [',' Texture_description]
Texture_component := Color_description Shape_description 'blobs'
Color_description := ['Dark' | 'Light'] Basic_term |Basic_termish-
```

```

Basic_term |Basic_term-Basic_term
Basic_term := 'red' | 'orange' | 'brown' | 'yellow' | 'green' | 'blue' | 'purple' | 'pink' |
'black' | 'grey' | 'white'
Basic_termish := 'reddish' | 'orangish' | 'brownish' | 'yellowish' | 'greenish' | 'bluish' |
'purplish' | 'pinkish' | 'blackish' | 'greyish' | 'whitish'
Shape_description := Orientation_description Size_description
Orientation_description := ['Non~Oriented' | Basic_orientation 'Oriented']
Basic_orientation := 'horizontal' | 'right' | 'vertical' | 'left'
Size_description := 'small' | 'medium' | 'large'

```

Para determinar los términos básicos a partir de los valores del TCD se han usado diferentes criterios. En la descripción del color se han asignado los nombres en base al sistema difuso definido por Benavente y cols. (2004). Según este sistema, los 11 términos básicos son parametrizados por funciones sigmoideas que asignan valores de pertenencia a cada término de color. A partir de este conjunto de pertenencias se determina el término usado para definir el color. Se asigna un único término de color si su pertenencia es alta, en este caso, se considera que es un color puro. Para el caso de colores no puros, solamente se usan los dos términos correspondientes a los valores con mayor pertenencia, por tanto se usa un término bilingüístico (ej: blue-green), son ejemplos de colores que están en la frontera entre dos términos de color. Si una de las pertenencias predomina respecto de la otra se usa el modificador *-ish* para el color que no predomina y para el que predomina se utiliza el término de color sin modificador (ej: bluish-green). La descripción de color además puede ser modificada por el término relacionado con la intensidad *dark* o *light*, en este caso el término hace referencia a la posición del color en el eje de la intensidad. El modificador *light* se asigna cuando la intensidad está por encima del 80 % de la intensidad del color, y el modificador *dark* se utiliza para intensidades por debajo del 20 %.

Para realizar la descripción de la forma se ha usado un vocabulario muy simplificado. La forma está limitada a dos formas simples de blobs, isótropa o no isótropa y alargada, la orientación ha sido limitada a cuatro términos y el tamaño se ha reducido a tres. Con respecto al tamaño, dado que el descriptor no es invariante a la escala (el área del blob se calcula en pixels), la asignación del término relacionado con el tamaño es dependiente del tamaño de la imagen. Ésta es una consideración a tener en cuenta a la hora de utilizar este vocabulario en aplicaciones de recuperación de imágenes. El tamaño especificado en la pregunta se debe ajustar al tamaño relativo de los blobs dentro de la imagen a buscar.

La gramática definida proporciona combinaciones de color imposibles, por ejemplo: *whitish-white* o *dark black*. Después de eliminar estas descripciones de color, se dispone de un diccionario semántico de aproximadamente 2085 sintagmas para explicar las componentes texturales.

Se pueden ver algunos ejemplos de estas descripciones en las figuras 4.3 y 4.10.

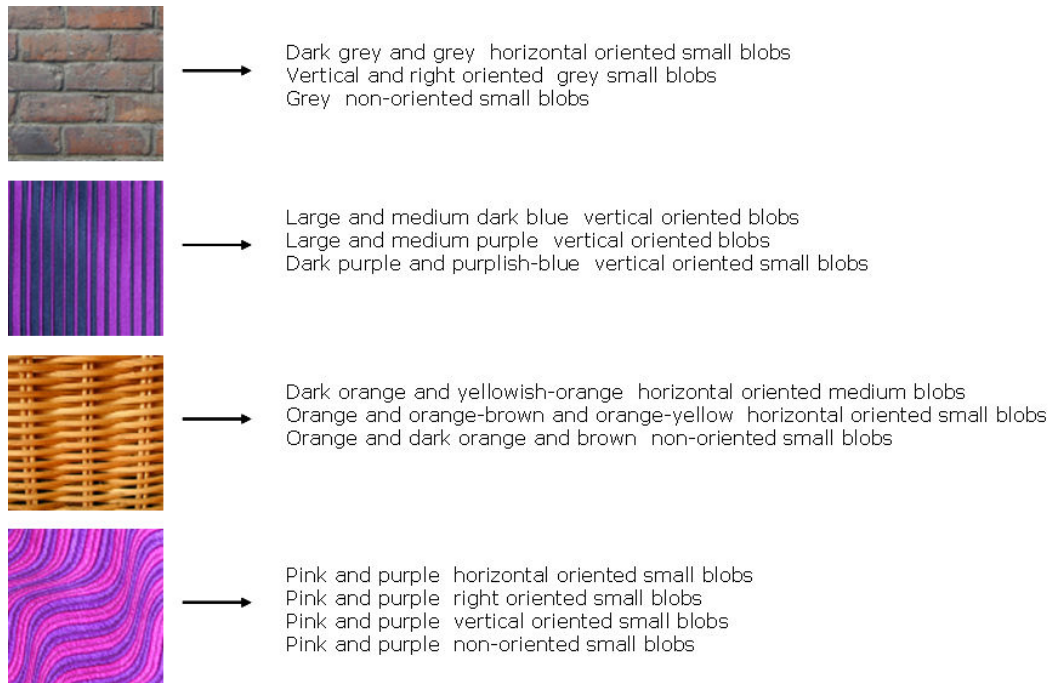


Figura 4.3: Diferentes texturas con la descripción de sus componentes texturales con términos del vocabulario.

4.3. Aplicación: Recuperación de imágenes de texturas

El objetivo de este experimento es comprobar la eficacia del descriptor *TCD* en la recuperación de imágenes en BD desde el punto de vista cuantitativo. Siguiendo con el procedimiento estándar de este tipo de aplicaciones, primero se calculan los descriptores de todas las imágenes de la BD y a continuación se efectúa la consulta sobre la BD utilizando para ello una imagen. Sobre la imagen que se utiliza como consulta se obtiene su descriptor *TCD* y se calcula la distancia de ésta a todas las imágenes de la BD de acuerdo con la función de similitud indicada en la ecuación 4.10. Repitiendo este procedimiento para todas las imágenes de la BD se evalúa la eficacia de esta aplicación con las medidas *Recall* y *Precision*, la primera definida en la ecuación 3.19 y la segunda se define de la siguiente manera:

$$precision(r) = \frac{\text{Retrieval}}{r} \quad (4.13)$$

siendo *Retrieval* el número de imágenes relevantes encontradas dentro de la r búsquedas.

La *precision* mide la habilidad del sistema de encontrar sólo las imágenes relevantes. Al igual que el *recall* esta medida también depende del número de imágenes a buscar (r).

Tabla 4.1: Recall promedio del descriptor TCD y otros descriptores.

BD	TCD	$LBP(RGB)_{8,1}$	$HTD+SCD$	$HTD+CSD$	PTD
<i>BDHomo</i>	96.39	91.11	97.85	97.78	98.30
<i>VisTexP</i>	77.26	70.38	83.44	80.24	84.80
<i>VisTexL</i>	81.10	73.20	87.41	83.45	88.70
<i>CorelTex</i>	73.11	61.89	67.33	72.36	84.89
<i>CorelTex2</i>	79.33	72.50	76.11	85.89	90.81
<i>CorelV1Tex</i>	86.25	77.53	85.94	91.89	95.00
<i>CorelV2Tex</i>	85.33	81.47	88.53	91.50	95.17
<i>CorelTexPat</i>	85.05	86.89	93.69	89.86	95.25
<i>CorelSand</i>	68.72	55.54	72.00	85.06	86.86

4.3.1. Resultados experimentales

Las BD de imágenes utilizadas en este experimento son las identificadas como *BDHomo*, *VisTexP*, *VisTexL*, *CorelTex*, *CorelTex2*, *CorelV1Tex*, *CorelV2Tex*, *CorelSand* y *CorelTexPat*, cuya construcción se ha definido en el capítulo anterior.

Se ha estudiado el efecto de variar los pesos α y β en el cálculo de la distancia entre representaciones (4.2.2), para ello se ha calculado el *Recall* promedio con 5 conjuntos de parámetros diferentes:

1. Igual peso de los atributos forma y color, $\alpha = 0,5$ y $\beta = 0,5$.
2. Más peso a los atributos forma, $\alpha = 0,7$ y $\beta = 0,3$.
3. Más peso a los atributos color, $\alpha = 0,3$ y $\beta = 0,7$.
4. Todo el peso a los atributos forma, $\alpha = 1$ y $\beta = 0$.
5. Todo el peso a los atributos color, $\alpha = 0$ y $\beta = 1$.

En las gráficas de las figuras 4.4 y 4.5 se encuentran los resultados obtenidos en todas las BD utilizadas con estos conjuntos de parámetros.

Para contrastar la eficacia del descriptor TCD con otros descriptores, en la la tabla 4.1 se muestran los resultados del \overline{recall} (definido en la ecuación 3.20) de cada una de las BD obtenidos con el descriptor TCD (con $\alpha = 0,5$ y $\beta = 0,5$), con los descriptores MPEG-7 adecuados para texturas homogéneas (las combinaciones $HTD+SCD$ y $HTD+CSD$) además del descriptor LBP ($LBP(RGB)_{8,1}$). En la última columna se han añadido los resultados obtenidos con el descriptor PTD . Se complementa la comparativa con las gráficas con las curvas *recall-precision* en las figuras 4.6 y 4.7.

4.3.2. Discusión sobre los resultados experimentales

En 6 de las 9 BD utilizadas (gráficas 4.4.(d), (e), (f) y gráficas 4.5) cuando se utiliza sólo la componente forma del descriptor TCD el *recall* promedio es bastante inferior al obtenido con las

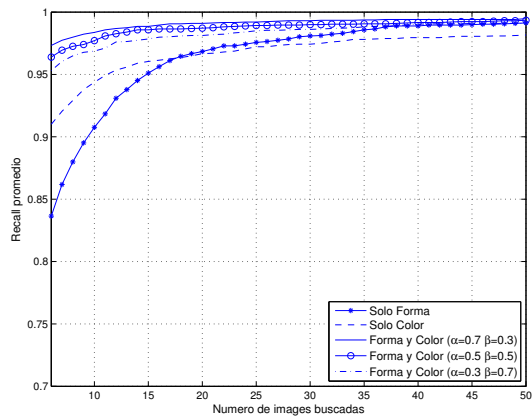
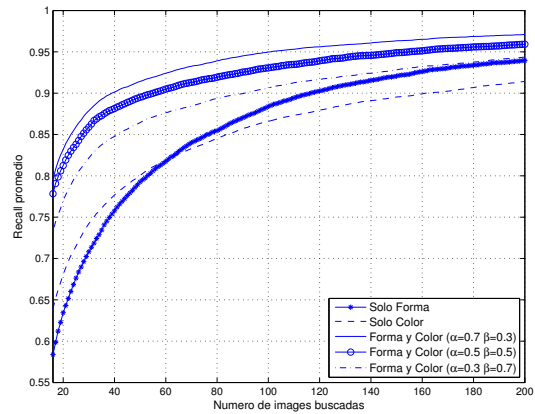
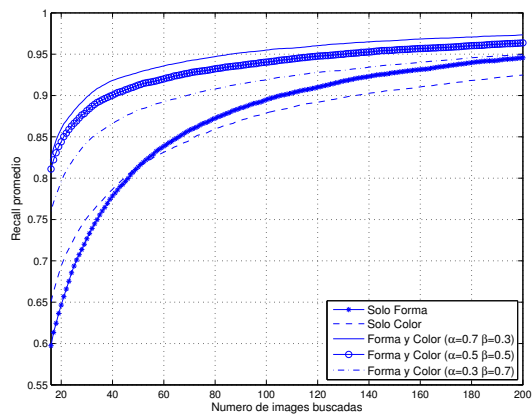
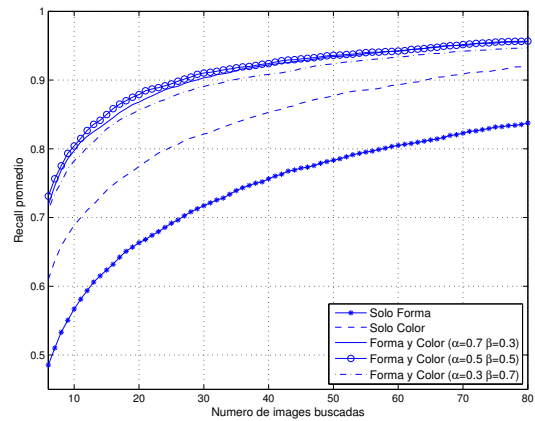
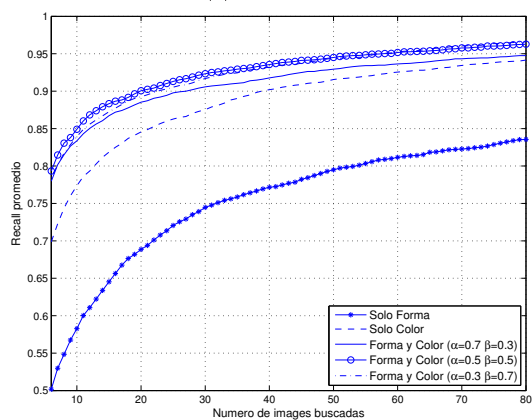
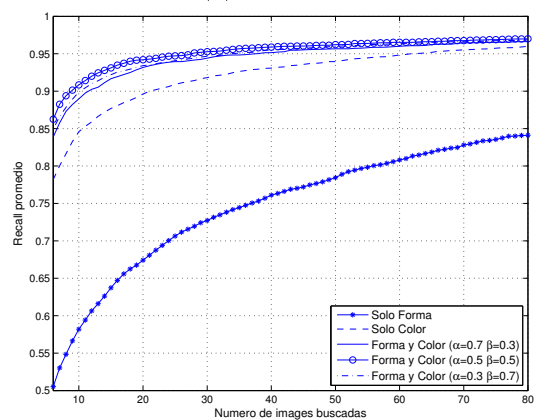
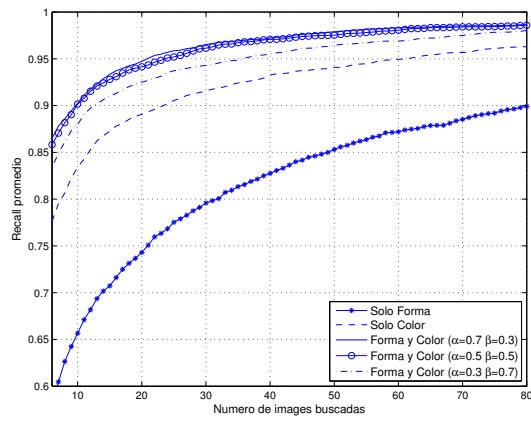
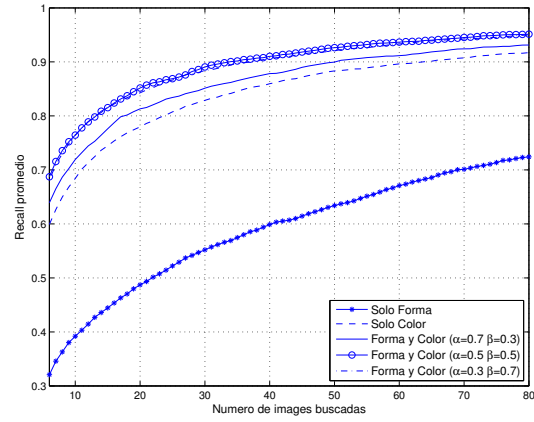
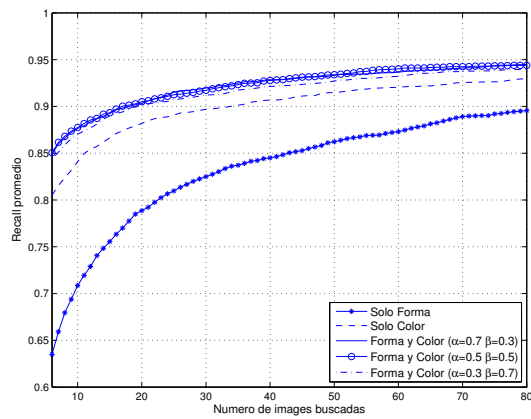
(a) *BDHomo*(b) *VisTexP*(c) *VisTexL*(d) *CorelTex*(e) *CorelTex2*(f) *CorelV1Tex*

Figura 4.4: *Recall promedio* del descriptor *TCD* utilizando diferentes pesos.

(a) *CorelV2Tex*(b) *CorelSand*(c) *CorelTexPat*Figura 4.5: *Recall promedio TCD* utilizando diferentes pesos.

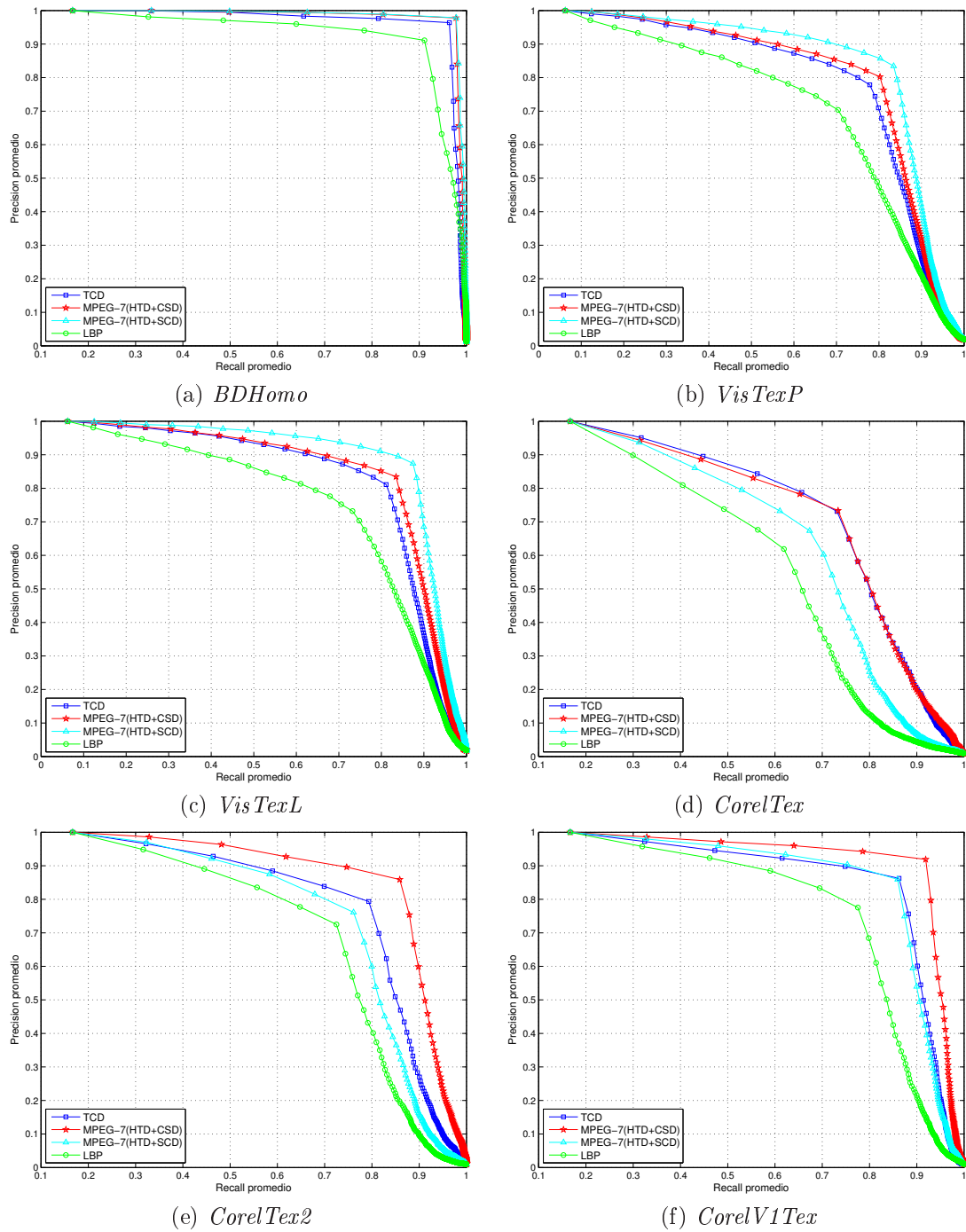


Figura 4.6: Curvas Recall-Precision.

componentes color del descriptor *TCD*. En todas las BD el uso combinado de ambas componentes producen un resultado superior al uso individual de éstas, existiendo poca diferencia en los resul-

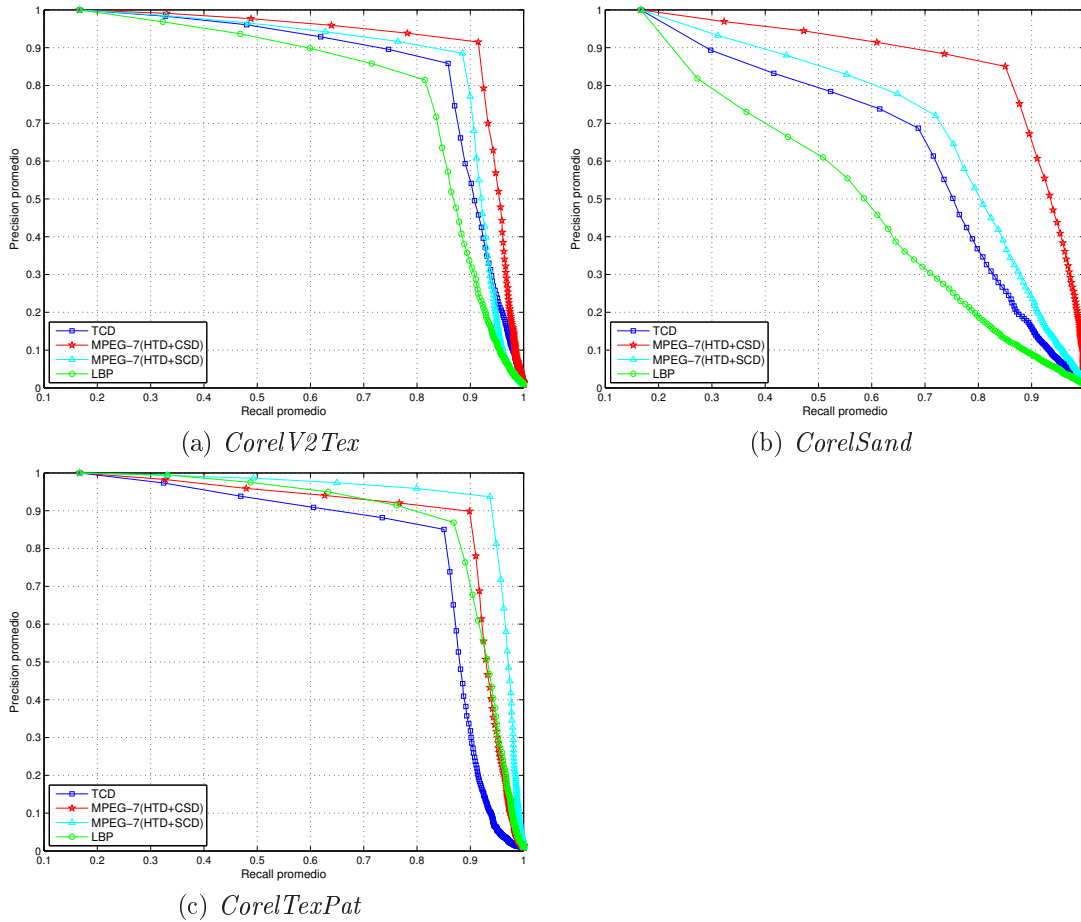


Figura 4.7: Curvas Recall-Precision.

tados con pequeñas variaciones de los pesos $\alpha = 0,5$ y $\beta = 0,5$. Por tanto se puede concluir que el valor de los pesos no es un factor crítico a la hora de calcular la similitud entre texturas.

La comparativa del descriptor *TCD* respecto a otros descriptores (tabla 4.1) muestra que el descriptor *TCD* sólo supera al resto de descriptores en la BD *CorelTex* (resultado en cursiva), sin embargo, en 8 de las 9 BD este descriptor mejora los resultados obtenidos con el descriptor *LBP*.

En la BD homogénea *BDHomo*, el descriptor *TCD* y los MPEG-7 ofrecen resultados prácticamente iguales, esta característica también se mantiene en la gráfica 4.6.(a) donde se evalúa la precisión en función del *recall*.

El descriptor *TCD* no mejora los resultados obtenidos con el descriptor *PTD*, éste continua siendo claramente superior. Lo que demuestra que la reducción de la información llevada a cabo en el descriptor *TCD* con respecto del *PTD* ha producido también un descenso en la eficiencia de la representación de las texturas en color. Sin embargo, se ha automatizado su descripción semántica.

4.3.3. Evaluación cualitativa

Desde el punto de vista cualitativo se ha analizado el comportamiento del descriptor *TCD* en la representación de texturas. Para ello se ha utilizado el método estadístico de *Escalado Multidimensional (MDS)* (Kruskal y Wish, 1978) que permite visualizar datos multidimensionales reduciendo su dimensión. A partir de una matriz de distancias entre los puntos de un espacio multidimensional, el algoritmo *MDS* genera una matriz de distancias en un nuevo espacio de menor dimensión y una medida de *stress*. Ésta última mide el error entre las distancias entre los puntos de salida y entre los puntos de entrada. El *stress* está acotado entre $[0, 1]$ y el valor de *stress*=0 indica un ajuste perfecto.

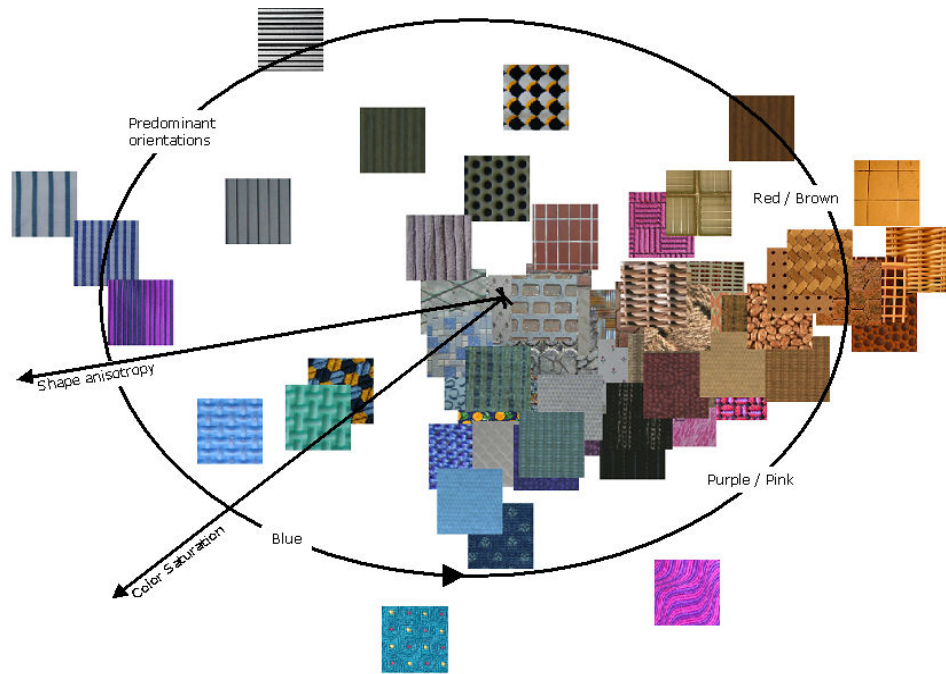
Utilizando el descriptor *TCD* se ha calculado la matriz de distancias entre 80 texturas diferentes de la BD *BDHomo*. A partir de ésta se ha aplicado la técnica *MDS* para obtener una representación 2D, produciendo un valor de *stress*=0.165. El resultado de la nueva representación se muestra en la figura 4.8a. En esta figura se puede observar cómo alrededor de la circunferencia exterior que se ha marcado, las imágenes tienen colores dominantes saturados y direcciones dominantes. Las imágenes con colores dominantes están a su vez agrupadas por tonos (azul, rosa-púrpura, marrón-rojo). En la figura 4.8b se muestra un zoom de la parte central donde predominan las imágenes con tonos grises y texturas con blobs isotrópicos.

4.4. Aplicación: *Browsing*

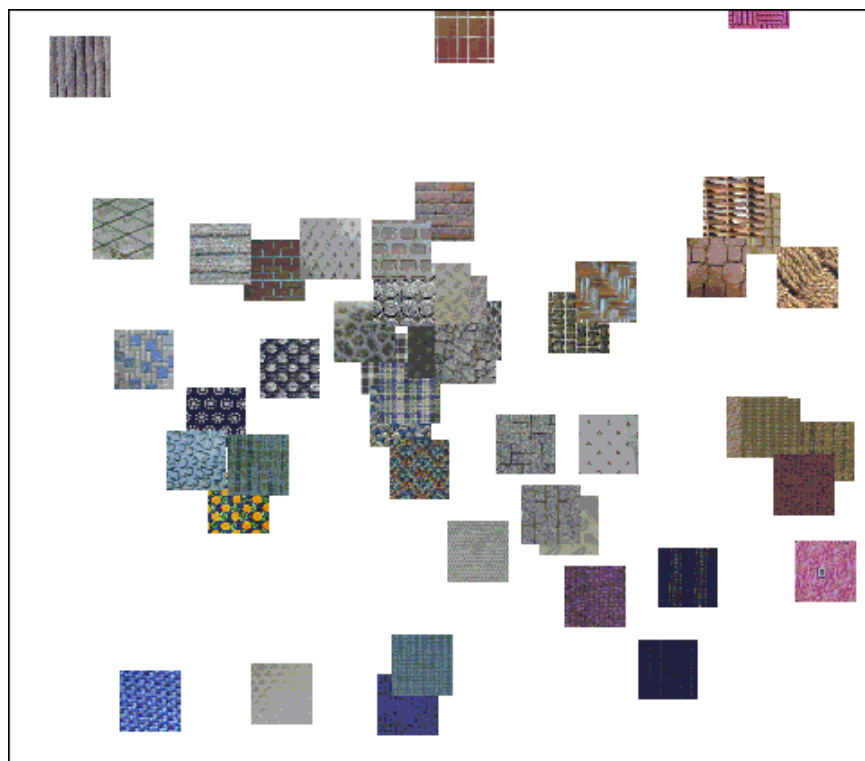
En lugar de utilizar imágenes como referencia para buscar imágenes similares en una BD, en este experimento se ha empleado un sintagma a partir del vocabulario definido en el apartado 4.2.3.

Dada una descripción en forma de sintagma se ha construido una función que proporciona el descriptor numérico equivalente, éste es el representante del área del espacio de textons que identifica la característica correspondiente. Con la descripción en forma de *TCD* se ha calculado la distancia de este descriptor a todas las imágenes de la BD (*BDHomo*), de entre éstas se han seleccionado las correspondientes a las 40 imágenes más cercanas. De nuevo, utilizando la matriz de distancias entre las 41 texturas como entrada al *MDS* se ha obtenido la nueva matriz de distancias de 2 dimensiones, ésta permite representar en un gráfico 2D la relación entre las texturas con características similares a las que se buscan.

En la figura 4.9 se muestran dos ejemplos de esta aplicación donde con la letra *Q* se ha indicado la posición donde se encuentra el sintagma con la descripción a buscar. En el primer ejemplo, como resultado a la consulta *Blue vertical oriented large blobs*, en la gráfica se observa la presencia de dos ejes esenciales, uno que indica la orientación-isotropía y otro la saturación del color, tal y como se ha expresado en la consulta. En el segundo ejemplo se muestra un consulta más compleja *Blue horizontal oriented small blobs AND Brown horizontal oriented small blobs*, que es representado por un *TCD* con dos componentes texturales. En este caso claramente se observa un eje en cuyos extremos se sitúan los dos colores de la consulta y en la parte central cercana a la localización de la consulta *Q* imágenes con ambos colores. En este ejemplo no hay una interpretación clara para



(a) Distribución de todas las imágenes.



(b) Zoom de la parte central de (a).

Figura 4.8: Distribución de las imágenes de acuerdo con la distancia obtenida por el MDS en 2D.

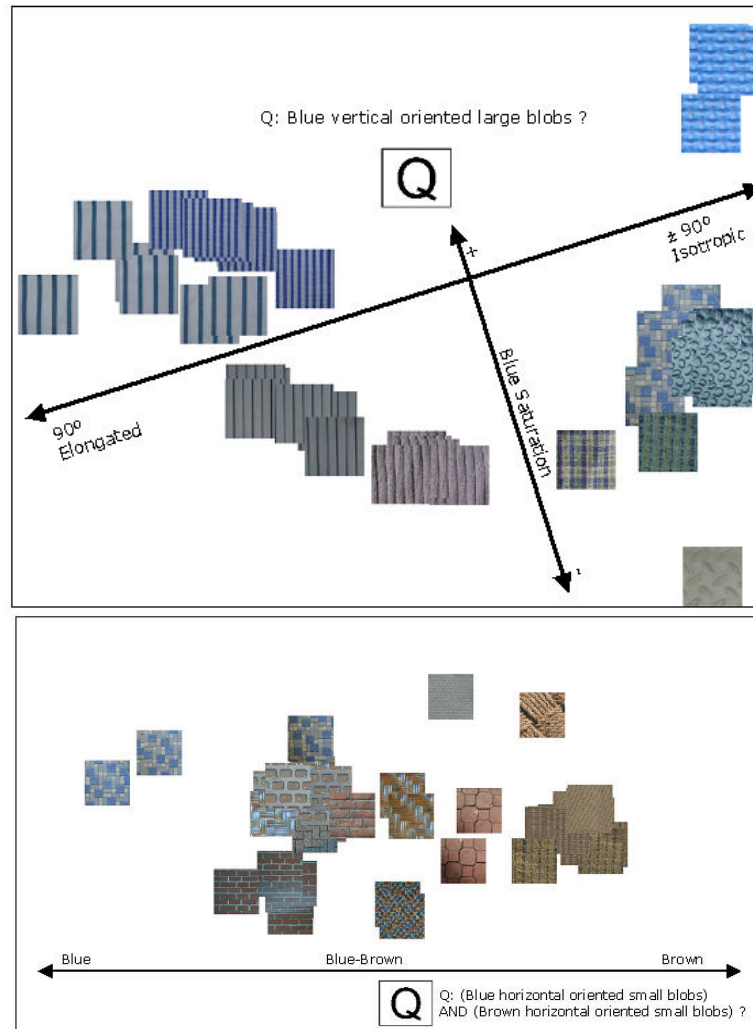


Figura 4.9: Resultados del MDS 2D utilizando las 40 imágenes más similares a la consulta realizada, esta se identifica con la letra Q.

la características relacionadas con la forma.

Este experimento cualitativo ilustra el comportamiento de la métrica propuesta para el descriptor *TCD* que es coherente con la descripción de la textura proporcionada en los términos lingüísticos básicos del vocabulario que se ha propuesto.

4.5. Aplicación: Detección de texturas en imágenes

En aplicaciones donde se quiere detectar y localizar una textura dentro de una escena es necesario efectuar primero una división de la imagen en regiones con características homogéneas, este procedimiento constituye la segmentación de la imagen. Para ello es necesario disponer de una

representación donde cada pixel de la imagen esté representado por un vector de características, a este tipo de representación se la denomina densa.

La representación conceptual, basada en los descriptores de las componentes texturales, se puede interpretar como una representación no densa ya que no todos los pixels de la imagen tienen representación. Por este motivo se ha desarrollado un método para extender la representación conceptual a todos los pixels de la imagen. La idea de este proceso es la de simular un agrupamiento por proximidad y similitud.

Antes de pasar a describir el proceso de agrupamiento es importante destacar que, el procedimiento desarrollado para obtener las componentes texturales parte de que las imágenes deben de estar previamente segmentadas. Cuando esto no es así, la presencia de fronteras que delimitan las texturas de una escena pueden producir componentes texturales erróneas. Éste problema se ha subsanado en parte al modificar el procedimiento para la obtención de los blobs que se describe a continuación.

El operador Laplaciano usado en la detección de los blobs (ecuación 3.3), con el parámetro adecuado (escala) produce una respuesta máxima en el centro de un blob cuyo tamaño esté cercano a la escala de detección, pero también produce una respuesta máxima ante la presencia de contornos entre texturas. Por lo tanto, es necesario filtrar los blobs producidos por los contornos entre texturas para caracterizar correctamente los elementos de la textura.

Para solucionar este problema se utiliza conjuntamente el operador gradiente ($\nabla L(\cdot; \sigma)$) y el operador Laplaciano en la detección de los blobs. El operador gradiente, al igual que el operador Laplaciano, produce respuesta máxima en los contornos pero no en el centro de un blob. Por tanto los puntos de la imagen donde los dos operadores tengan máxima respuesta no serán considerados blobs de la imagen.

4.5.1. Agrupación por proximidad y similitud

La representación densa de una imagen se puede obtener extendiendo espacialmente las propiedades de cada una de las componentes texturales (\hat{T}^i) a todos los pixels de su *área de influencia*. Este último concepto se define a continuación:

Área de influencia Región de la imagen que contiene el grupo de blobs que forman una componente textural.

La representación densa de una imagen con k componentes texturales tiene $k \times r \times c$ dimensiones, siendo $r \times c$ la dimensión de la imagen. De esta manera cada pixel está representado por un vector de k componentes binarios que representa la pertenencia a una componente textural específica, ésta a su vez identificada por su descriptor TCD. Para ilustrar la hipotética representación que se quiere obtener, en la figura 4.10 se muestra una imagen ejemplo con dos texturas y su correspondiente representación.

Se han diseñado dos maneras diferentes para extender las propiedades de una componente textural. En los siguientes apartados se detallan los dos procedimientos.

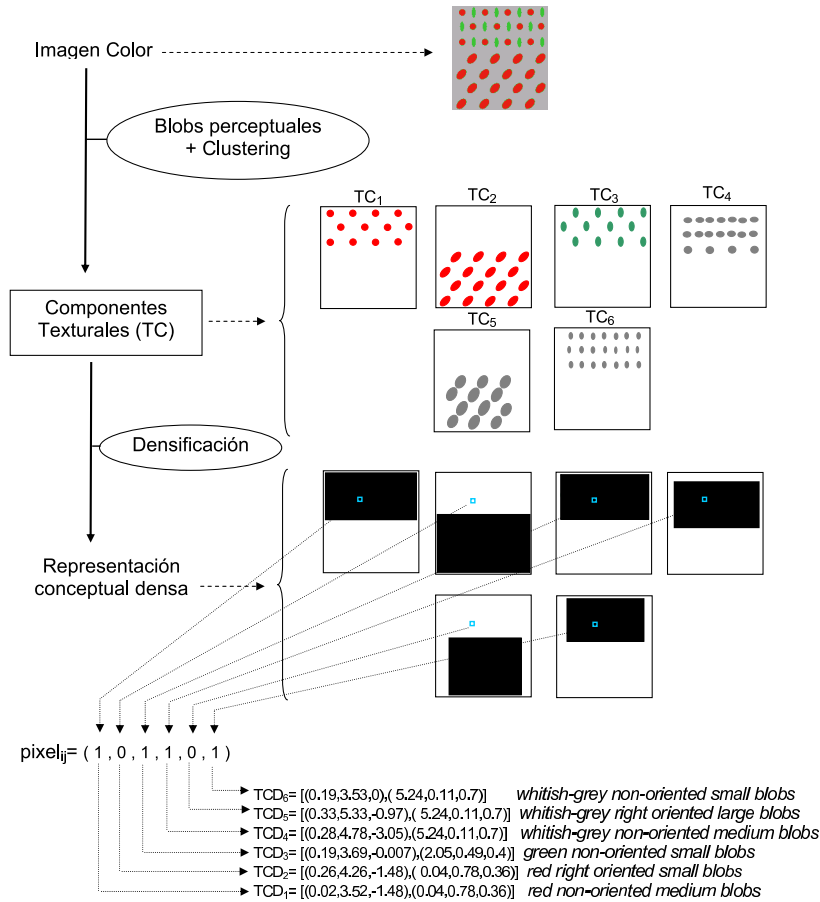


Figura 4.10: Etapas de la densificación de la representación conceptual.

Algoritmo del Potencial

Este procedimiento es adecuado para texturas cuya estructura es irregular, en la que no existe periodicidad. El objetivo de este algoritmo es englobar todos los blobs que pertenecen a una *TC* dentro de una misma área compacta. Este área debe cumplir una serie de requisitos, no debe tener agujeros y el área deber ser la menor posible. La resolución de este problema ha sido inspirada en el comportamiento del potencial gravitatorio provocado por varias masas. Según ésta, el potencial provocado por una masa respecto al resto es inversamente proporcional a la distancia que las separa y proporcional al producto de sus masas, esto es:

$$V = - \sum G \frac{Mm}{d}$$

A partir de esta idea se define P_1 para cada punto de la imagen como el potencial provocado por los blobs de una Componente Textural de la siguiente manera:

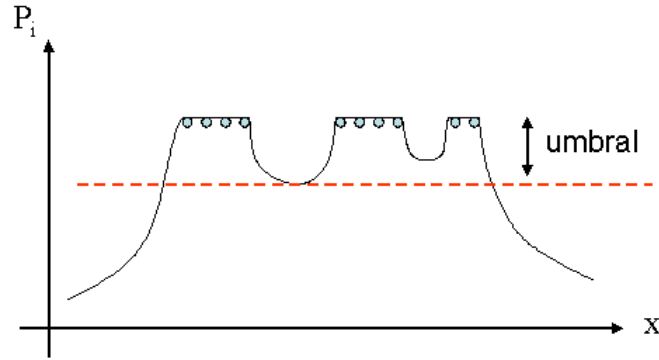


Figura 4.11: Esquema del umbral de agrupamiento sobre una función unidimensional.

$$P_1(p_j) = \begin{cases} \sum_{p_i \in B} \frac{1}{d(p_j, p_i)} & \text{si } p_j \in \overline{B} \\ P_{max} & \text{si } p_j \in B \end{cases} \quad (4.14)$$

siendo B el conjunto de puntos que pertenecen a los blobs de la componente textural y \overline{B} su complementario entonces $B \cap \overline{B} = \emptyset$, y P_{max} es el potencial máximo.

Además de la función P_1 también se han estudiado qué agrupaciones producen otras funciones derivadas de ésta, comprobando que en algunos casos se obtienen mejores resultados con la función P_2 definida de la siguiente manera:

$$P_2(p_j) = \begin{cases} \sum_{p_i \in B} \frac{1}{d^2(p_j, p_i)} & \text{si } p_j \in \overline{B} \\ P_{max} & \text{si } p_j \in B \end{cases} \quad (4.15)$$

La forma de cualquiera de las funciones P_1 o P_2 (P_i) en un espacio unidimensional se muestra en la figura 4.11.

Si sobre cualquiera de las funciones P_i se aplica un umbral correspondiente al mínimo local indicado en la figura 4.11 el resultado es la unión de todos los puntos. Extendiendo este modelo a un espacio bidimensional se consigue englobar los blobs en una única área. La forma que tendrá el área englobante vendrá determinada por el valor del umbral escogido.

Para determinar qué umbral debe de utilizarse es preciso definir qué características debe cumplir el área englobante. Éstas básicamente son tres:

- El área englobante debe de tener área mínima.
- El perímetro del área englobante debe de ser mínimo.
- El área englobante debe de ser convexa (sin agujeros).

A partir del conjunto de imágenes de test de la figura 4.12.(a) se han calculado las funciones que relacionan $perimetro = f(umbral)$ y $area = g(umbral)$ constatando que no existe un valor de umbral óptimo que produzcan una área y perímetro mínimo a la vez, por lo que es imposible cumplir

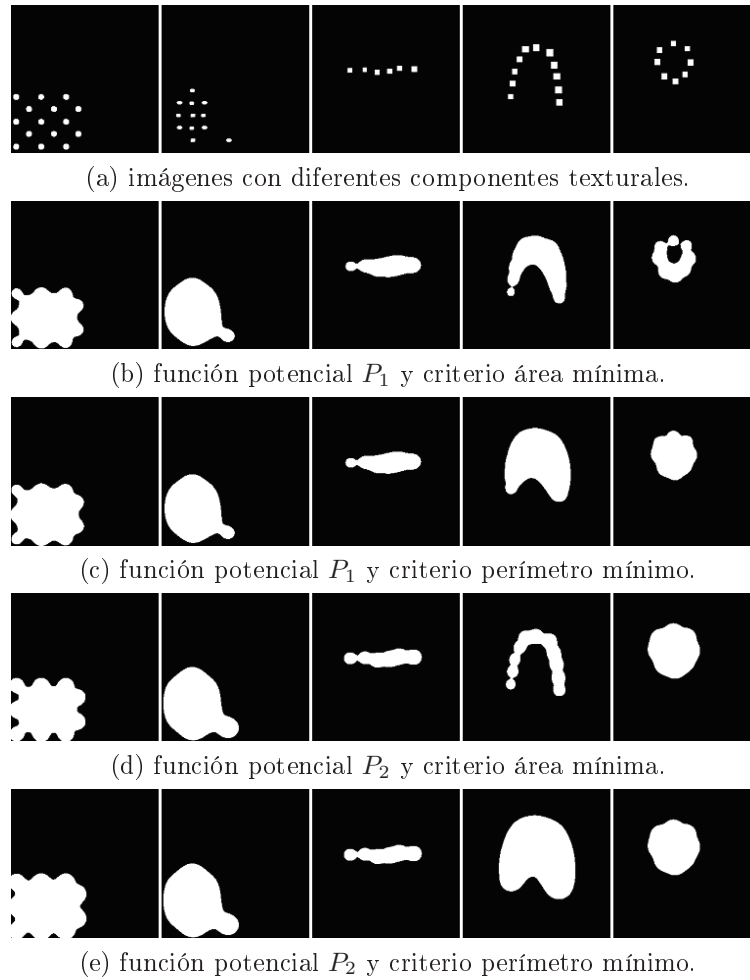


Figura 4.12: Resultados de las áreas obtenidas utilizando diferentes funciones potenciales y diferentes criterios.

todas las restricciones especificadas anteriormente. En la figura 4.12 se muestran los resultados de utilizar las funciones potenciales P_1 y P_2 con diferentes criterios.

Desde el punto de vista de conseguir una agrupación perceptual, las áreas englobantes obtenidas con la función P_1 con el criterio de perímetro mínimo son aceptables en todos los casos excepto cuando los blobs forman una figura cóncava, en este caso la función P_2 con el criterio de área mínima ofrece un mejor resultado. Por tanto el procedimiento final para calcular la agrupación es el siguiente:

1. Para cada componente textural ($\forall i = 1, 2, \dots, k$)
 - a. Reconstruir los blobs de la componente textural $\hat{\mathbf{T}}$ en la imagen $\mathbf{I}_{\hat{\mathbf{T}}}$, de las mismas dimensiones que la imagen original (I), poniendo a uno los pixels pertenecientes a los

blobs detectados:

$$\mathbf{I}_{\hat{\mathbf{T}}^i}(x, y) = \begin{cases} 1 & \text{si } \exists j \mid D_{w_j l_j \theta_j}(x, y) \leq 0 \\ 0 & \text{en otro caso} \end{cases} \quad (4.16)$$

donde j es la j -ésima fila de $\hat{\mathbf{T}}^i$, que corresponde al j -ésimo blob de la clase i , y $D_{w_j l_j \theta_j}(x, y)$ es la ecuación de la elipse rotada θ_j con semiejes w_j y l_j centrada en el punto (x_j, y_j) , que define la forma del j -ésimo blob.

- b. Determinar si el área envolvente puede ser cóncava. Para ello se calcula el centroide de la agrupación obtenida utilizando la función potencial P_2 y aplicando el umbral que produce un área englobante mínima. Si el centroide está incluido dentro de la agrupación es que el área envolvente no es cóncava.
- c. Si el área envolvente es cóncava utilizar la función potencial P_2 para englobar los blobs de la componente textural aplicando el umbral que produce un área englobante mínima sin agujeros.

$$\mathbf{I}_{\mathbf{C}^i} = \text{Binarizacion}(P_2(\mathbf{I}_{\hat{\mathbf{T}}^i}), \text{umbral}) \quad (4.17)$$

Si el área envolvente es convexa utilizar la función potencial P para englobar los blobs de la componente textural aplicando el umbral que produce una área englobante sin agujeros y cuyo perímetro es mínimo.

$$\mathbf{I}_{\mathbf{C}^i} = \text{Binarizacion}(P_1(\mathbf{I}_{\hat{\mathbf{T}}^i}), \text{umbral}) \quad (4.18)$$

Algoritmo *winner-take-all*

Este procedimiento es adecuado para texturas que presentan cierta regularidad y se basa en la estimación de la periodicidad de las componentes texturales de la textura. La expansión de las propiedades de las componentes texturales ha sido inspirada en el proceso de inhibición intracortical de Malik y Perona (1990). Ésta genera regiones compactas, que en este caso contiene blobs con formas, orientaciones y colores similares. A continuación se desglosan los diferentes pasos de este procedimiento:

1. Para cada componente textural ($\forall i = 1, 2, \dots, k$)
 - a. A partir de la localización de los centros de los blobs, $\hat{\mathbf{T}}_{\text{loc}}^i$ (ecuación 4.6), construir la matriz de distancias entre todos los blobs de la Componente Textural i . Ésta se identifica como \mathbf{DT}^i .
 - b. Calcular el histograma de la matriz de distancias, $\text{Hist}(\mathbf{DT}^i)$.
 - c. Estimar el periodo de la componente textural a partir de la localización del máximo del histograma,

$$p_i = \arg \max_d (\text{Hist}(\mathbf{DT}^i))$$

con $d \in \mathbf{DT}^i$

- d. Reconstruir los blobs de la componente textural en la imagen $\mathbf{I}_{\hat{\mathbf{T}}^i}$, de las mismas dimensiones que la imagen original, asignando el valor 1 los pixels pertenecientes a los blobs detectados:

$$\mathbf{I}_{\hat{\mathbf{T}}^i}(x, y) = \begin{cases} 1 & \text{si } \exists j \mid D_{w_j l_j \theta_j}(x, y) \leq 0 \\ 0 & \text{en otro caso} \end{cases} \quad (4.19)$$

donde j es la j -ésima fila de $\hat{\mathbf{T}}^i$ y $D_{w_j l_j \theta_j}(x, y)$ es la ecuación de la elipse rotada θ_j con semiejes w_j y l_j centrada en el punto (x_j, y_j) .

El resultado de la reconstrucción es una imagen binaria que contiene elipses con la misma localización, tamaño y orientación que los blobs originales de la componente textural i .

- e. Extender las propiedades de la TC a su área de influencia. Para ello realizar una operación morfológica de clausura (Serra, 1988) con un elemento estructural circular $EE_{p_i/2}$ de radio $p_i/2$:

$$\mathbf{I}_{\mathbf{C}^i} = ((\mathbf{I}_{\hat{\mathbf{T}}^i} \oplus EE_{p_i/2}) \ominus EE_{p_i/2}) \quad (4.20)$$

En la figura 4.13 se muestran los resultados de este procedimiento sobre unas componentes texturales de ejemplo.

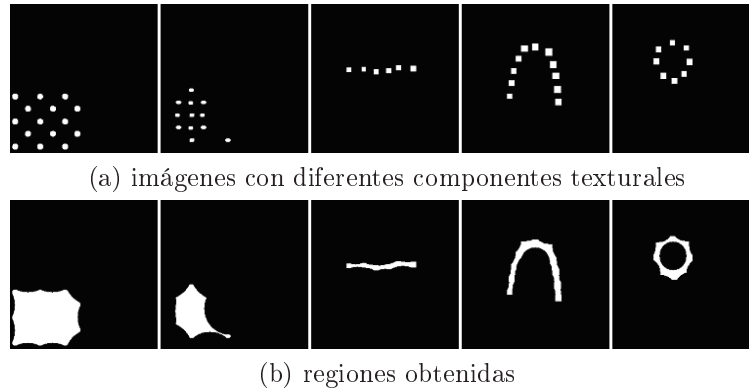


Figura 4.13: Expansion de características utilizando el procedimiento 1.

4.5.2. Integración de componentes texturales

Asumiendo que una textura está formada por varias componentes texturales localizadas en una misma región y que una imagen es un conjunto de texturas, a partir de la representación densa de una imagen y directamente aplicando cualquier método de clustering se pueden obtener las regiones de la imagen que comparten las propiedades de las componentes texturales. Consiguiendo de esta manera segmentar la imagen en las diferentes texturas que contenga.

Para integrar las componentes texturales y generar la segmentación de la imagen, en esta ocasión se ha utilizado una red neuronal SOM (Self Organising Map, Kohonen (1997)). Ésta es alimentada

por los k vectores binarios (de longitud $r \times c$) resultantes de la representación conceptual densa y obtiene como resultado un vector de dimensión $r \times c$ cuyos valores representan las etiquetas que identifican las diferentes regiones resultado de la clasificación/segmentación. El parámetro de entrada de este clasificador es el número de etiquetas (clases/regiones) que debe generar, N .

$$SOM(\mathbf{I}_{C^1}, \dots, \mathbf{I}_{C^k}) = \mathbf{IS} \quad (4.21)$$

siendo \mathbf{I}_{C^i} la representación densa de la imagen y $IS(p_i) \in [1 \dots N] \quad \forall p_i$ pixel de la imagen.

La segmentación obtenida con este método se puede clasificar dentro de la categoría de segmentaciones débiles, según la terminología utilizada por Smeulders y cols. (2000). Ésta se caracteriza por producir una segmentación poco fina sin delimitar correctamente la frontera entre las regiones.

4.5.3. Experimento 1: Recuperación de imágenes en BD de Mosaicos

El objetivo de este experimento es comprobar la eficacia de la representación densa del descriptor TCD para segmentar imágenes con texturas, cuando sobre éstas se quiere efectuar la búsqueda de una textura determinada. Para ello se ha utilizado una amplia BD de imágenes construida para este fin.

En los siguientes apartados se describe cómo se han construido las BD sobre las que se efectuará la búsqueda así como la descripción del procedimiento empleado y los resultados experimentales obtenidos.

Construcción de la BD

Según Manjunath y Ma (1996) una imagen puede considerarse como un mosaico de diferentes texturas, esta observación ha servido de inspiración a la hora de construir la base de datos de imágenes utilizada en este experimento. A partir de las imágenes de la BD $BDHomo$, cuya composición de ha descrito en el apartado 3.5.3, se ha construido una nueva BD formada por 1500 imágenes de mosaicos. Cada mosaico está compuesto por 9 imágenes seleccionadas aleatoriamente de la BD $BDHomo$, con la restricción de que un mosaico no puede contener ninguna textura repetida. Esta manera de construir los mosaicos ha determinado que el número de muestras pertenecientes a la misma textura contenidas en toda la BD varíe entre 137 y 194.

Procedimiento

Inicialmente, utilizando la representación densa del descriptor TCD se lleva a cabo la segmentación de todas las imágenes de la BD utilizando el método descrito en el apartado 4.5.2, en éste, dado que las texturas que contiene la BD son bastante regulares, se ha utilizado el algoritmo *winner-take-all* para extender las propiedades de las componentes texturales. Como las imágenes de mosaicos están formadas por 9 texturas el parámetro del clasificador SOM debe ser como mínimo 9, pero debido a la inexactitud de la segmentación, sobre todo en las zonas frontera entre texturas,

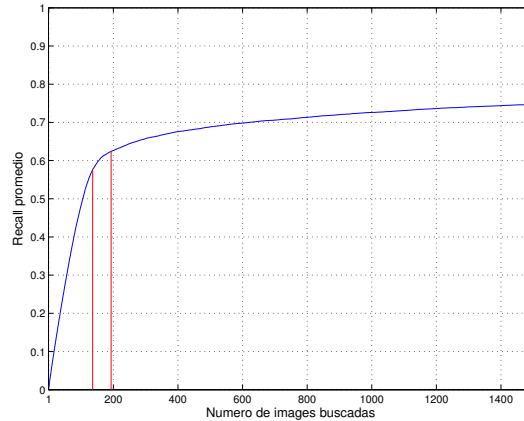


Figura 4.14: Eficiencia en la recuperación de imágenes en la BD Mosaicos.

se ha fijado este parámetro a 16. Es preferible que se produzca una sobre-segmentación a que se unan regiones con la misma textura.

A continuación en cada región obtenida por la segmentación se calcula el descriptor *TCD* dentro del mayor rectángulo inscrito en cada región. El cálculo del descriptor en una área rectangular es una restricción derivada del uso de los operadores utilizados en la detección de los blobs.

Una vez descritas las diferentes regiones que contienen las imágenes que componen la BD ya se puede realizar la búsqueda de texturas.

Resultados y discusión

En la BD de Mosaicos se efectúa la búsqueda de 80 imágenes, una de cada tipo de textura de la BD *BDHomo*, que son escogidas aleatoriamente de entre las 6 imágenes que hay de cada textura. Para valorar el resultado de la recuperación de imágenes al igual que en el capítulo anterior, se ha utilizado la medida \overline{recall} (J. Smith, 1998). En el caso ideal, las N primeras imágenes resultado de la búsqueda (entre 137 y 194) corresponderían a muestras de la misma textura. En la figura 4.14 se muestra la gráfica con el *recall* promedio, en ésta se ha marcado una línea roja entre los valores 137 y 194 donde la eficiencia de la búsqueda es del 58 % y 62 % respectivamente. Comparando esta eficiencia con la obtenida en el experimento previo para la BD *BDHomo*, podemos comprobar que la reducción de la eficiencia es debida en parte a las deficiencias del método de segmentación ya que si las texturas ya están segmentadas como en el experimento previo los resultados de la búsqueda de imágenes son mucho mejores.

Para ilustrar con ejemplos el resultado de la recuperación de imágenes, en la figura 4.15 se muestran los resultados obtenidos con dos imágenes. En las figuras 4.15a y 4.15c se muestran las imágenes que se ha buscado en la BD de Mosaicos y debajo de éstas, en las figuras 4.15b y 4.15d, se encuentran las imágenes encontradas. En la parte inferior de cada imagen se indica el lugar del ranking que éstas ocupan y se ha destacado (con un rectángulo blanco) la zona de detección. Así mismo, en la primera fila se muestran las 5 primeras imágenes encontradas y en la segunda fila las

primeras 5 imágenes donde la recuperación ha fallado.

Por otro lado, en la figura 4.15d se puede ver el comportamiento de la segmentación, en concreto en la imagen encontrada en 14^o lugar, el rectángulo blanco que indica el lugar donde se ha detectado la textura buscada engloba parte de dos texturas similares. En esta misma figura se puede observar como algunos de los errores de la recuperación son debidos a que la BD contiene texturas con alto grado de similitud, pero que sin embargo no están clasificadas dentro del mismo conjunto relevante.

4.5.4. Experimento 2: Recuperación de imágenes en BD con escenas naturales

Con el objetivo de comprobar la aplicabilidad de la representación densa del descriptor *TCD* y el propio *TCD* para localizar texturas en imágenes de escenas naturales, se ha realizado un experimento similar al anterior. En este caso, la BD está formada por 21 imágenes correspondientes a 7 tipos de animales, habiendo 3 imágenes diferentes de cada tipo animal. En la figura B.19 del apéndice A se muestran las imágenes que componen esta pequeña BD.

El procedimiento utilizado en este experimento de recuperación de imágenes es el mismo que el descrito en el apartado 4.5.3, con la única diferencia que el parámetro del clasificador *SOM* ha sido 4, delimitando así a este número las regiones que se obtendrán en la segmentación.

A continuación se ha utilizado como consulta a la BD de Animales una porción correspondiente a la textura de algunos de los animales de la BD. El resultado obtenido se muestra en la figura 4.16. Los resultados son correctos en algunos casos, en otros hay errores y aparentemente ciertos equívocos. Los errores debidos a la incorrecta segmentación de las imágenes se ilustran al efectuar la búsqueda de las plumas de un pelícano. En este caso en 5^o lugar se ha localizado la textura en un lugar donde sólo está en parte. Otro caso similar se produce como resultado de la sobre-segmentación de la imagen, éste produce que se localice la textura en dos partes de la misma imagen (sobre la misma búsqueda anterior, las imágenes en 3^{er} y 4^o lugar). Finalmente el aparente equívoco se ilustra al efectuar la primer consulta, al buscar la piel del guepardo se encuentran en 3^{er} y 4^o lugar jirafas, cuando en los tres primeros lugares del ranking deberían aparecer los 3 guepardos. Este caso y desde el punto de vista perceptual no se debe considerar como un error ya que la piel de ambos animales es muy similar.



(a) consulta



n° 1



n° 2



n° 3



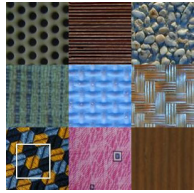
n° 4



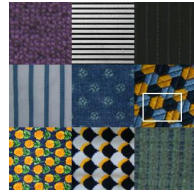
n° 5



n° 27



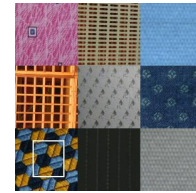
n° 41



n° 45



n° 52

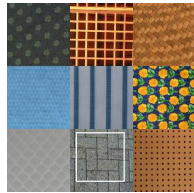


n° 60

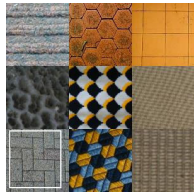
(b) Resultado de la consulta



(c) consulta



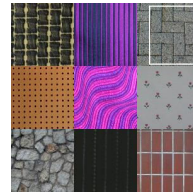
n° 1



n° 2



n° 3



n° 4



n° 5



n° 14



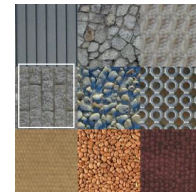
n° 46



n° 54



n° 56



n° 57

(d) Resultado de la consulta

Figura 4.15: En la parte superior (a) y (c) imagen a buscar. En (b) y (d) 1^a fila: 5 primeras imágenes encontradas. 2^a fila: primeros 5 errores.

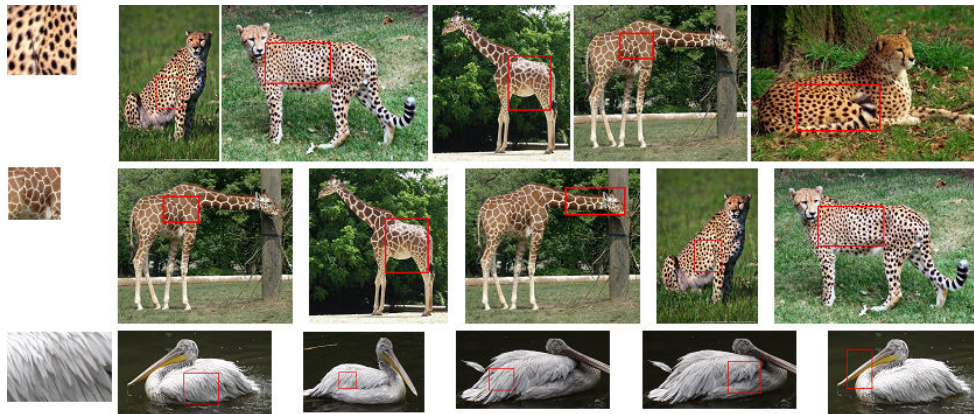


Figura 4.16: A la izquierda imágenes utilizadas en la consulta y a la derecha los 5 primeros resultados de cada una de las consultas.

Capítulo 5

Conclusiones

Al igual que otras representaciones computacionales, la desarrollada en este trabajo se ha inspirado en los resultados sobre percepción. La *representación holística* propuesta va más allá de los estudios psicofísicos ya que implementa la definición original de los *textons* de Julesz y Bergen (1983) pero añadiendo el color para representar las texturas.

La representación de la textura presentada en este trabajo se basa en el concepto de *blob perceptual*, elemento donde se integran las características *textura y color*, constituyendo así una manera de combinar estos estímulos visuales que no se ajusta a ninguno de los métodos utilizados hasta ahora para combinar estas características. El modelo de construcción de esta representación utiliza el enfoque holístico *bag-of-words* que ya había sido empleado para representar texturas monocromas (Renninger y Malik, 2004; Varma y Zisserman, 2005).

La adecuación de la teoría de los *textons* a este enfoque holístico ha permitido obtener diferentes descriptores basados en las densidades de los *textons* que se han mostrado muy efectivos y han proporcionado un buen marco de trabajo para la experimentación.

Los descriptores holísticos (*PTD* y *JPTD*) presentan las cualidades perseguidas en los objetivos de este trabajo:

- Integran textura y color bajo un único concepto, el *blob perceptual*, que por definición es perceptual y le proporciona una cierta semántica muy básica.
- Combinan los atributos forma y color de los *blobs perceptuales* sin necesidad de realizar ninguna normalización de estos atributos.
- Son simples, ya que se derivan de la cuantificación de los espacios de *textons* lo que permite obtener un vocabulario único y general, sin necesidad de aplicar ninguna técnica de aprendizaje.
- Cada término del vocabulario tiene una semántica asociada. Permite una traducción directa a palabras del vocabulario humano y por tanto es más comprensible.

- Ha demostrado ser un modelo de representación de texturas general, ya que ha proporcionado buenos resultados en aplicaciones de recuperación y clasificación de texturas sobre BD con alta diversidad.
- Tiene un coste computacional relativamente bajo.

5.1. Aportaciones

A continuación se resumen las principales aportaciones de este trabajo:

- Se han propuesto una serie de descriptores de texturas (*TD*, *JPTD*, *PTD*) basados en la cuantificación directa de los espacios de *textons*, de baja dimensión, consiguiendo con ello un alto grado de efectividad. Los diferentes vocabularios utilizados han permitido extraer conclusiones interesantes:
 - La introducción de transformaciones uniformes, o perceptuales, en los espacios de representación de los *textons* (tanto de color como de forma) ha permitido incrementos del orden del 5 % en experimentos de recuperación de imágenes.
 - La conclusión anterior se reafirma si se analiza el tamaño de los vocabularios construidos. En cuantificaciones menos perceptuales el tamaño del vocabulario puede hacer variar el rendimiento del orden del 6 %, mientras que en espacios más perceptuales el rendimiento sólo varía del orden del 2 %. Por tanto, utilizando espacios perceptuales se pueden emplear vocabularios más pequeños.
 - Los descriptores derivados (*PTD* y *JPTD*) a partir del descriptor *TD* son significativamente mejores debido a la introducción de la co-ocurrencia de las propiedades de forma y color de los blobs.
 - Los descriptores propuestos han permitido una mejora del estado del arte respecto a los descriptores más utilizados como es el caso de los *MPEG-7* y *LBP*. En aplicaciones de recuperación de imágenes y clasificación se han obtenido incrementos del orden del 4 % respecto al mejor de los descriptores.
 - Los vocabularios propuestos presentan la propiedad de tener *palabras visuales* con una semántica muy básica, siendo ésta es una de las propiedades de la que carecen los vocabularios definidos en los enfoques *bag-of-words*.
- Las diferentes representaciones de la imagen a partir de los vocabularios propuestos ha hecho emerger una conclusión interesante sobre cómo tiene que ser la combinación de la textura y el color, de acuerdo con las teorías de la percepción hay caminos visuales diferentes para la percepción de la forma y el color.
- Se ha estudiado la viabilidad de la definición de una representación conceptual de nivel intermedio. El estudio se ha llevado a cabo con la realización de una serie de experimentos, en estos:

- Se ha propuesto un procedimiento basado en propiedades de organización perceptual para conseguir un descriptor denso que ha permitido llevar a cabo la localización de texturas en escenas.
- Se ha explorado el espacio de representación conjunta de la textura y el color de imágenes de texturas orientado a aplicaciones de *browsing* de texturas.
- Se ha propuesto una gramática que genera un vocabulario en lenguaje natural de nivel muy básico. Esto ha permitido generar expresiones muy primitivas pero que son una novedad en este ámbito.

5.2. Líneas futuras

A pesar de las cualidades que presentan los descriptores desarrollados en esta tesis, admiten como todo mejoras. Así mismo también abren diversas vías de continuación en diferentes direcciones, algunas de las cuales se presentan a continuación.

- Aplicar el descriptor desarrollado para el reconocimiento de objetos y escenas. Para ello es necesario:
 - Invarianza a la rotación y a otras transformaciones geométricas. La representación holística, construida sobre dos espacios perceptuales, permite hacer transformaciones sobre estos para adaptarlos a diferentes requerimientos. En este sentido es factible conseguir que la representación sea invariante a la rotación y a la escala ya que la orientación de los blobs y el área de estos son las componentes del espacio perceptual donde se representa la forma de los blobs.
 - Completar y validar las técnicas desarrolladas en la construcción de representaciones densas a partir de los descriptores definidos. La completa validación del descriptor *TCD* se debe de realizar con mayor número de experimentos en escenas naturales. Así mismo, el modelo de construcción de la representación densa es también aplicable a los descriptores holísticos.
- Estudiar la aplicación de los diferentes descriptores desarrollados para la definición de un vocabulario de materiales de textura: ladrillos, hierba, paja, mármol, arena, piedra, etc.
- Añadir más semántica a los descriptores introduciendo nuevos atributos que capturen la organización espacial de los blobs perceptuales dentro de la imagen (ej: modelos de regularidad, aleatoriedad). Para ello será necesario utilizar la información sobre la localización de los blobs de que ya se dispone. De esta manera, se añade poder discriminatorio en las aplicaciones que requieran un nivel de descripción más detallado, como es el caso del *browsing*.
- Completar la gramática desarrollada para la descripción conjunta de la textura y el color introduciendo la traducción de los atributos de color a nombres de colores.

Apéndice A

Tablas con resultados experimentales

A.0.1. Resultados experimentales con el descriptor *JPTD*

Los resultados de la recuperación de imágenes utilizando el descriptor *JPTD* sobre todas las BD se muestran en las tablas: A.1-A.16. Se han usado diferentes vocabularios (variando la función de cuantificación Q_s y Q_c) según los modelos descritos en el apartado 3.6.2 y se ha calculado el *recall promedio*(r) en cada experimento. En las tablas se muestra este índice siendo r el número de imágenes relevantes de cada BD, idealmente para este valor el *recall promedio* debería de ser 100. Los espacios de color utilizados en este caso son el espacio HSI-Carron y el espacio HSV-Smith. La similitud entre las imágenes se ha calculado con la función χ^2 .

En cada una de las tablas se ha destacado el negrita el mejor resultado obtenido.

Tabla A.1: Recall promedio en la BD Outex.

Modelos de cuantización	Tamaño del vocabulario						
	$(N_1, N_2, N_3), N, N \times M$	$(5, 5, 5), 125, 15625$	$(5, 5, 5), 125, 60750$	$(6, 6, 6), 216, 55296$	$(7, 7, 7), 216, 55296$	$(7, 7, 7), 343, 166698$	$(5, 5, 7), 175, 105000$
$Q_{\#}^S, Q_{\#}^C$	$(M_1, M_2, M_3), M$	$(5, 5, 5), 125$	$(9, 9, 6), 486$	$(8, 8, 4), 256$	$(8, 8, 4), 256$	$(9, 9, 6), 486$	$(10, 10, 6), 600$
	<i>HSI-Carron</i>	52.18	58.14	61.13	57.17	59.42	58.20
	<i>HSV-Smith</i>	58.11	60.73	63.93	60.40	61.40	64.68
$Q_{\#}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	$(5, 5, 5), 125, 32000$	$(7, 7, 7), 343, 87808$	$(5, 5, 7), 175, 100800$			
	$(M_1, M_2, M_3), M$	$(4, 16, 4), 256$	$(4, 16, 4), 256$	$(6, 16, 6), 576$			
	<i>HSI-Carron</i>	60.54	61.19	64.47			
<i>HSV-Smith</i>	62.55	63.32	65.70				
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	$(4, 8, 5), 160, 20480$	$(4, 8, 5), 160, 40860$	$(4, 8, 7), 224, 57344$	$(3, 8, 7), 168, 96768$		
	$(M_1, M_2, M_3), M$	$(4, 8, 4), 128$	$(4, 16, 4), 256$	$(4, 16, 4), 256$	$(6, 16, 6), 576$		
	<i>HSI-Carron</i>	66.88	66.32	66.45	67.79		
<i>HSV-Smith</i>	67.25	66.76	66.66	67.17			
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	$(3, 8, 7), 168, 43008$	$(3, 8, 7), 168, 67200$	$(3, 8, 7), 168, 96768$			
	$(M_1, M_2, M_3), M$	$(4, 16, 4), 256$	$(5, 16, 5), 400$	$(6, 16, 6), 576$			
	<i>HSI-Carron</i>	66.79	65.68	68.02			
<i>HSV-Smith</i>	67.03	66.83	68.25				

Tabla A.2: Recall promedio en la BD VisTexL.

Modelos de cuantización	Tamaño del vocabulario						
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N \times M$	(5,5,5),125, 16925	(5,5,5),125, 60750	(6,6,6),216, 55296	(7,7,7),216, 55296	(7,7,7),343, 166698	(5,5,7),175, 105000
	$(M_1, M_2, M_3), M$	(5,5,5),125	(9,9,9),486	(8,8,4),256	(8,8,4),256	(9,9,9),486	(10,10,6),600
	<i>HSI-Carron</i>	76.06	81.94	80.83	78.71	82.40	80.53
<i>HSV-Smith</i>	80.85	81.75	84.04	82.00	83.38	83.24	
$Q_{\#}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	(5,5,5),125, 32000	(7,7,7),343, 87808	(5,5,7),175, 100800			
	$(M_1, M_2, M_3), M$	(4,1,6,4),256	(4,1,6,4),256	(6,1,6,6),576			
	<i>HSI-Carron</i>	82.31	82.78	84.45			
<i>HSV-Smith</i>	82.91	83.44	83.79				
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	(4,8,5),160, 20480	(4,8,5),160, 40960	(4,8,7),224, 57344	(3,8,7),168, 96768		
	$(M_1, M_2, M_3), M$	(4,8,4),128	(4,1,6,4),256	(4,1,6,4),256	(6,1,6,6),576		
	<i>HSI-Carron</i>	82.28	84.21	84.19	85.42		
<i>HSV-Smith</i>	82.46	84.33	84.31	84.15			
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	(3,8,7),168, 43008	(3,8,7),168, 67200	(3,8,7),168, 96768			
	$(M_1, M_2, M_3), M$	(4,1,6,4),256	(5,1,6,5),400	(6,1,6,6),576			
	<i>HSI-Carron</i>	83.97	84.50	85.08			
<i>HSV-Smith</i>	83.88	84.91	83.98				

Tabla A.3: Recall promedio en la BD VisTexP.

Modelos de cuantización	Tamaño del vocabulario						
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N \times M$	(5,5,5),125, 16925	(5,5,5),125, 60750	(6,6,6),216, 55296	(7,7,7),216, 55296	(7,7,7),343, 166698	(5,5,7),175, 105000
	$(M_1, M_2, M_3), M$	(5,5,5),125	(9,9,9),486	(8,8,4),256	(8,8,4),256	(9,9,9),486	(10,10,6),600
	<i>HSI-Carron</i>	73.80	78.58	76.80	74.26	78.96	77.32
<i>HSV-Smith</i>	78.49	78.88	80.33	77.44	79.44	79.28	
$Q_{\#}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	(5,5,5),125, 32000	(7,7,7),343, 87808	(5,5,7),175, 100800			
	$(M_1, M_2, M_3), M$	(4,1,6,4),256	(4,1,6,4),256	(6,1,6,6),576			
	<i>HSI-Carron</i>	77.21	77.93	80.43			
<i>HSV-Smith</i>	78.84	79.29	80.02				
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	(4,8,5),160, 20480	(4,8,5),160, 40960	(4,8,7),224, 57344	(3,8,7),168, 96768		
	$(M_1, M_2, M_3), M$	(4,8,4),128	(4,1,6,4),256	(4,1,6,4),256	(6,1,6,6),576		
	<i>HSI-Carron</i>	78.33	79.99	80.09	81.79		
<i>HSV-Smith</i>	79.50	80.48	80.53	80.82			
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	(3,8,7),168, 43008	(3,8,7),168, 67200	(3,8,7),168, 96768			
	$(M_1, M_2, M_3), M$	(4,1,6,4),256	(5,1,6,5),400	(6,1,6,6),576			
	<i>HSI-Carron</i>	80.16	81.26	81.79			
<i>HSV-Smith</i>	80.45	81.94	80.82				

Tabla A.4: Recall promedio en la BD CorelTex.

Modelos de cuantización	Tamaño del vocabulario						
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N \times M$	(5,5,5),125, 16925	(5,5,5),125, 60750	(6,6,6),216, 55296	(7,7,7),216, 55296	(7,7,7),343, 166698	(5,5,7),175, 105000
	$(M_1, M_2, M_3), M$	(5,5,5),125	(9,9,9),486	(8,8,4),256	(8,8,4),256	(9,9,9),486	(10,10,6),600
	<i>HSI-Carron</i>	64.06	70.44	75.67	67.67	70.81	77.86
<i>HSV-Smith</i>	67.86	74.72	79.33	72.64	74.83	74.92	
$Q_{\#}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	(5,5,5),125, 32000	(7,7,7),343, 87808	(5,5,7),175, 100800			
	$(M_1, M_2, M_3), M$	(4,1,6,4),256	(4,1,6,4),256	(6,1,6,6),576			
	<i>HSI-Carron</i>	73.31	73.58	75.64			
<i>HSV-Smith</i>	74.36	74.42	77.33				
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	(4,8,5),160, 20480	(4,8,5),160, 40960	(4,8,7),224, 57344	(3,8,7),168, 96768		
	$(M_1, M_2, M_3), M$	(4,8,4),128	(4,1,6,4),256	(4,1,6,4),256	(6,1,6,6),576		
	<i>HSI-Carron</i>	77.17	80.56	80.75	81.64		
<i>HSV-Smith</i>	78.5	80.92	80.03	81.89			
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	(3,8,7),168, 43008	(3,8,7),168, 67200	(3,8,7),168, 96768			
	$(M_1, M_2, M_3), M$	(4,1,6,4),256	(5,1,6,5),400	(6,1,6,6),576			
	<i>HSI-Carron</i>	81.06	81.47	81.67			
<i>HSV-Smith</i>	81.11	81.67	82.03				

Tabla A.5: Recall promedio en la BD *CorelTex2*.

Modelos de cuantización	Tamaño del vocabulario						
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3).N, N \times M$	(5,5,5),125, 15625	(5,5,5),125, 60750	(6,6,6),216, 55296	(7,7,7),216, 55296	(7,7,7),343, 166698	(5,5,7),175, 105000
	$(M_1, M_2, M_3).M$	(5,5,5),125	(9,9,6),486	(8,8,4),256	(8,8,4),256	(9,9,6),486	(10,10,6),600
	<i>HSl-Carron</i>	71.31	81.08	80.78	78.19	81.69	81.44
	<i>HSV-Smith</i>	75.92	93.11	84.86	82.94	83.27	85.17
$Q_{\#}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3).N, N \times M$	(5,5,5),125, 32000	(7,7,7),343, 87808	(5,5,7),175, 100800			
	$(M_1, M_2, M_3).M$	(4,16,4),256	(4,16,4),256	(6,16,6),576			
	<i>HSl-Carron</i>	82.69	83.14	85.97			
	<i>HSV-Smith</i>	84.94	85.11	87.44			
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3).N, N \times M$	(4,8,5),100, 20480	(4,8,5),100, 40960	(4,8,7),224, 57344	(3,8,7),168, 96768		
	$(M_1, M_2, M_3).M$	(4,8,4),128	(4,16,4),256	(4,16,4),256	(6,16,6),576		
	<i>HSl-Carron</i>	81.28	84.31	84.33	87.36		
	<i>HSV-Smith</i>	84.36	86.61	86.53	88.61		
$Q_{\otimes}^S, Q_{\#}^C$	$(N_1, N_2, N_3).N, N \times M$	(3,8,7),168, 43008	(3,8,7),168, 67200	(3,8,7),168, 96768			
	$(M_1, M_2, M_3).M$	(4,16,4),256	(5,16,5),400	(6,16,6),576			
	<i>HSl-Carron</i>	84.36	85.94	87.31			
	<i>HSV-Smith</i>	84.64	87.50	88.28			

Tabla A.6: Recall promedio en la BD *CorelV1Tex*.

Modelos de cuantización	Tamaño del vocabulario						
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3).N, N \times M$	(5,5,5),125, 15625	(5,5,5),125, 60750	(6,6,6),216, 55296	(7,7,7),216, 55296	(7,7,7),343, 166698	(5,5,7),175, 105000
	$(M_1, M_2, M_3).M$	(5,5,5),125	(9,9,6),486	(8,8,4),256	(8,8,4),256	(9,9,6),486	(10,10,6),600
	<i>HSl-Carron</i>	83.67	88.86	89.50	88.75	89.31	90.19
	<i>HSV-Smith</i>	88.67	91.06	90.31	90.19	91.41	91.94
$Q_{\#}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3).N, N \times M$	(5,5,5),125, 32000	(7,7,7),343, 87808	(5,5,7),175, 100800			
	$(M_1, M_2, M_3).M$	(4,16,4),256	(4,16,4),256	(6,16,6),576			
	<i>HSl-Carron</i>	90.67	90.69	92.44			
	<i>HSV-Smith</i>	91.42	91.50	92.86			
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3).N, N \times M$	(4,8,5),100, 20480	(4,8,5),100, 40960	(4,8,7),224, 57344	(3,8,7),168, 96768		
	$(M_1, M_2, M_3).M$	(4,8,4),128	(4,16,4),256	(4,16,4),256	(6,16,6),576		
	<i>HSl-Carron</i>	90.14	90.67	90.56	91.94		
	<i>HSV-Smith</i>	89.97	91.44	91.14	92.50		
$Q_{\otimes}^S, Q_{\#}^C$	$(N_1, N_2, N_3).N, N \times M$	(3,8,7),168, 43008	(3,8,7),168, 67200	(3,8,7),168, 96768			
	$(M_1, M_2, M_3).M$	(4,16,4),256	(5,16,5),400	(6,16,6),576			
	<i>HSl-Carron</i>	90.67	90.83	91.97			
	<i>HSV-Smith</i>	91.25	92.36	92.25			

Tabla A.7: Recall promedio en la BD *CorelV2Tex*.

Modelos de cuantización	Tamaño del vocabulario						
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3).N, N \times M$	(5,5,5),125, 15625	(5,5,5),125, 60750	(6,6,6),216, 55296	(7,7,7),216, 55296	(7,7,7),343, 166698	(5,5,7),175, 105000
	$(M_1, M_2, M_3).M$	(5,5,5),125	(9,9,6),486	(8,8,4),256	(8,8,4),256	(9,9,6),486	(10,10,6),600
	<i>HSl-Carron</i>	87.42	91.36	90.89	90.31	91.64	91.28
	<i>HSV-Smith</i>	89.53	92.17	91.39	91.08	92.05	92.28
$Q_{\#}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3).N, N \times M$	(5,5,5),125, 32000	(7,7,7),343, 87808	(5,5,7),175, 100800			
	$(M_1, M_2, M_3).M$	(4,16,4),256	(4,16,4),256	(6,16,6),576			
	<i>HSl-Carron</i>	91.17	91.14	91.83			
	<i>HSV-Smith</i>	91.25	91.33	92.33			
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3).N, N \times M$	(4,8,5),100, 20480	(4,8,5),100, 40960	(4,8,7),224, 57344	(3,8,7),168, 96768		
	$(M_1, M_2, M_3).M$	(4,8,4),128	(4,16,4),256	(4,16,4),256	(6,16,6),576		
	<i>HSl-Carron</i>	90.75	91.44	91.44	91.28		
	<i>HSV-Smith</i>	89.91	91.08	91.14	91.22		
$Q_{\otimes}^S, Q_{\#}^C$	$(N_1, N_2, N_3).N, N \times M$	(3,8,7),168, 43008	(3,8,7),168, 67200	(3,8,7),168, 96768			
	$(M_1, M_2, M_3).M$	(4,16,4),256	(5,16,5),400	(6,16,6),576			
	<i>HSl-Carron</i>	91.39	91.22	91.03			
	<i>HSV-Smith</i>	90.97	91.00	91.00			

Tabla A.8: Recall promedio en la BD *CorelTexPat*.

Modelos de cuantización	Tamaño del vocabulario						
	$(N_1, N_2, N_3), N, N \times M$	$(5,5,5),125, 16625$	$(5,5,5),125, 60750$	$(6,6,6),216, 55296$	$(7,7,7),216, 55296$	$(7,7,7),343, 166698$	$(5,5,7),175, 105000$
$Q_{\#}^S, Q_{\#}^C$	$(M_1, M_2, M_3), M$	$(5,5,5),125$	$(9,9,9),486$	$(8,8,4),256$	$(8,8,4),256$	$(9,9,9),486$	$(10,10,6),600$
	<i>HSI-Carron</i>	87.05	90.66	91.30	89.75	90.83	91.22
	<i>HSV-Smith</i>	89.11	92.22	92.61	90.58	93.33	92.22
$Q_{\#}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	$(5,5,5),125, 32000$	$(7,7,7),343, 87808$	$(5,5,7),175, 100800$			
	$(M_1, M_2, M_3), M$	$(4,1,6,4),256$	$(4,1,6,4),256$	$(6,1,6,6),576$			
	<i>HSI-Carron</i>	90.75	91.03	91.94			
<i>HSV-Smith</i>	91.06	91.33	92.44				
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	$(4,8,5),160, 20480$	$(4,8,5),160, 40960$	$(4,8,7),224, 57344$	$(3,8,7),168, 96768$		
	$(M_1, M_2, M_3), M$	$(4,8,4),128$	$(4,1,6,4),256$	$(4,1,6,4),256$	$(6,1,6,6),576$		
	<i>HSI-Carron</i>	91.38	92.55	92.50	93.08		
<i>HSV-Smith</i>	91.33	92.44	92.36	93.61			
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	$(3,8,7),168, 43008$	$(3,8,7),168, 67200$	$(3,8,7),168, 96768$			
	$(M_1, M_2, M_3), M$	$(4,1,6,4),256$	$(5,1,6,5),400$	$(6,1,6,6),576$			
	<i>HSI-Carron</i>	92.64	93.06	93.19			
<i>HSV-Smith</i>	92.58	93.72	93.75				

Tabla A.9: Recall promedio en la BD *CorelSand*.

Modelos de cuantización	Tamaño del vocabulario						
	$(N_1, N_2, N_3), N, N \times M$	$(5,5,5),125, 16625$	$(5,5,5),125, 60750$	$(6,6,6),216, 55296$	$(7,7,7),216, 55296$	$(7,7,7),343, 166698$	$(5,5,7),175, 105000$
$Q_{\#}^S, Q_{\#}^C$	$(M_1, M_2, M_3), M$	$(5,5,5),125$	$(9,9,9),486$	$(8,8,4),256$	$(8,8,4),256$	$(9,9,9),486$	$(10,10,6),600$
	<i>HSI-Carron</i>	66.81	77.69	78.03	78.31	77.64	80.69
	<i>HSV-Smith</i>	74.33	82.28	79.44	80.50	82.08	83.28
$Q_{\#}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	$(5,5,5),125, 32000$	$(7,7,7),343, 87808$	$(5,5,7),175, 100800$			
	$(M_1, M_2, M_3), M$	$(4,1,6,4),256$	$(4,1,6,4),256$	$(6,1,6,6),576$			
	<i>HSI-Carron</i>	83.33	82.97	84.64			
<i>HSV-Smith</i>	84.25	83.33	85.92				
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	$(4,8,5),160, 20480$	$(4,8,5),160, 40960$	$(4,8,7),224, 57344$	$(3,8,7),168, 96768$		
	$(M_1, M_2, M_3), M$	$(4,8,4),128$	$(4,1,6,4),256$	$(4,1,6,4),256$	$(6,1,6,6),576$		
	<i>HSI-Carron</i>	78.67	80.08	78.94	79.81		
<i>HSV-Smith</i>	78.64	80.19	79.17	80.33			
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	$(3,8,7),168, 43008$	$(3,8,7),168, 67200$	$(3,8,7),168, 96768$			
	$(M_1, M_2, M_3), M$	$(4,1,6,4),256$	$(5,1,6,5),400$	$(6,1,6,6),576$			
	<i>HSI-Carron</i>	77.39	78.08	78.36			
<i>HSV-Smith</i>	77.42	77.53	78.33				

Tabla A.10: Recall promedio en la BD *CorelBark*.

Modelos de cuantización	Tamaño del vocabulario						
	$(N_1, N_2, N_3), N, N \times M$	$(5,5,5),125, 16625$	$(5,5,5),125, 60750$	$(6,6,6),216, 55296$	$(7,7,7),216, 55296$	$(7,7,7),343, 166698$	$(5,5,7),175, 105000$
$Q_{\#}^S, Q_{\#}^C$	$(M_1, M_2, M_3), M$	$(5,5,5),125$	$(9,9,9),486$	$(8,8,4),256$	$(8,8,4),256$	$(9,9,9),486$	$(10,10,6),600$
	<i>HSI-Carron</i>	54.03	62.86	66.75	61.50	62.97	65.17
	<i>HSV-Smith</i>	59.92	68.64	71.36	66.97	68.97	70.56
$Q_{\#}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	$(5,5,5),125, 32000$	$(7,7,7),343, 87808$	$(5,5,7),175, 100800$			
	$(M_1, M_2, M_3), M$	$(4,1,6,4),256$	$(4,1,6,4),256$	$(6,1,6,6),576$			
	<i>HSI-Carron</i>	66.69	66.94	70.33			
<i>HSV-Smith</i>	69.50	69.83	71.67				
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	$(4,8,5),160, 20480$	$(4,8,5),160, 40960$	$(4,8,7),224, 57344$	$(3,8,7),168, 96768$		
	$(M_1, M_2, M_3), M$	$(4,8,4),128$	$(4,1,6,4),256$	$(4,1,6,4),256$	$(6,1,6,6),576$		
	<i>HSI-Carron</i>	67.28	70.94	70.81	72.97		
<i>HSV-Smith</i>	70.69	73.17	73.22	74.00			
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	$(3,8,7),168, 43008$	$(3,8,7),168, 67200$	$(3,8,7),168, 96768$			
	$(M_1, M_2, M_3), M$	$(4,1,6,4),256$	$(5,1,6,5),400$	$(6,1,6,6),576$			
	<i>HSI-Carron</i>	70.83	72.06	72.67			
<i>HSV-Smith</i>	72.83	72.89	73.47				

Tabla A.11: Recall promedio en la BD *CorelCol*.

Modelos de cuantización	Tamaño del vocabulario						
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3).N, N \times M$	(5,5,5),125, 15625	(5,5,5),125, 60750	(6,6,6),216, 55296	(7,7,7),216, 55296	(7,7,7),343, 166698	(5,5,7),175, 105000
	$(M_1, M_2, M_3).M$	(5,5,5),125	(9,9,6),486	(8,8,4),256	(8,8,4),256	(9,9,6),486	(10,10,6),600
	<i>HSI-Carron</i>	71.86	74.90	75.16	74.12	75.36	75.52
	<i>HSV-Smith</i>	73.30	76.63	77.03	75.78	76.56	77.09
$Q_{\#}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3).N, N \times M$	(5,5,5),125, 32000	(7,7,7),343, 87808	(5,5,7),175, 100800			
	$(M_1, M_2, M_3).M$	(4,16,4),256	(4,16,4),256	(6,16,6),576			
	<i>HSI-Carron</i>	73.95	74.12	76.76			
	<i>HSV-Smith</i>	76.14	76.34	77.32			
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3).N, N \times M$	(4,8,5),100, 20480	(4,8,5),100, 40960	(4,8,7),224, 57344	(3,8,7),168, 96768		
	$(M_1, M_2, M_3).M$	(4,8,4),128	(4,16,4),256	(4,16,4),256	(6,16,6),576		
	<i>HSI-Carron</i>	74.38	75.36	75.20	77.48		
	<i>HSV-Smith</i>	75.78	77.19	77.06	78.01		
$Q_{\otimes}^S, Q_{\#}^C$	$(N_1, N_2, N_3).N, N \times M$	(3,8,7),168, 43008	(3,8,7),168, 67200	(3,8,7),168, 96768			
	$(M_1, M_2, M_3).M$	(4,16,4),256	(5,16,5),400	(6,16,6),576			
	<i>HSI-Carron</i>	75.33	76.24	76.9			
	<i>HSV-Smith</i>	77.19	77.39	77.55			

Tabla A.12: Recall promedio en la BD *CorelMarb*.

Modelos de cuantización	Tamaño del vocabulario						
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3).N, N \times M$	(5,5,5),125, 15625	(5,5,5),125, 60750	(6,6,6),216, 55296	(7,7,7),216, 55296	(7,7,7),343, 166698	(5,5,7),175, 105000
	$(M_1, M_2, M_3).M$	(5,5,5),125	(9,9,6),486	(8,8,4),256	(8,8,4),256	(9,9,6),486	(10,10,6),600
	<i>HSI-Carron</i>	57.39	69.06	69.69	66.19	69.42	70.00
	<i>HSV-Smith</i>	66.22	75.00	74.25	71.03	75.00	75.25
$Q_{\#}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3).N, N \times M$	(5,5,5),125, 32000	(7,7,7),343, 87808	(5,5,7),175, 100800			
	$(M_1, M_2, M_3).M$	(4,16,4),256	(4,16,4),256	(6,16,6),576			
	<i>HSI-Carron</i>	73.39	72.86	78.86			
	<i>HSV-Smith</i>	77.33	77.03	81.28			
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3).N, N \times M$	(4,8,5),100, 20480	(4,8,5),100, 40960	(4,8,7),224, 57344	(3,8,7),168, 96768		
	$(M_1, M_2, M_3).M$	(4,8,4),128	(4,16,4),256	(4,16,4),256	(6,16,6),576		
	<i>HSI-Carron</i>	71.92	72.28	74.42	80.08		
	<i>HSV-Smith</i>	76.14	78.42	77.97	81.42		
$Q_{\otimes}^S, Q_{\#}^C$	$(N_1, N_2, N_3).N, N \times M$	(3,8,7),168, 43008	(3,8,7),168, 67200	(3,8,7),168, 96768			
	$(M_1, M_2, M_3).M$	(4,16,4),256	(5,16,5),400	(6,16,6),576			
	<i>HSI-Carron</i>	73.94	76.72	79.44			
	<i>HSV-Smith</i>	77.67	79.39	80.97			

Tabla A.13: Recall promedio en la BD *CorelPain*.

Modelos de cuantización	Tamaño del vocabulario						
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3).N, N \times M$	(5,5,5),125, 15625	(5,5,5),125, 60750	(6,6,6),216, 55296	(7,7,7),216, 55296	(7,7,7),343, 166698	(5,5,7),175, 105000
	$(M_1, M_2, M_3).M$	(5,5,5),125	(9,9,6),486	(8,8,4),256	(8,8,4),256	(9,9,6),486	(10,10,6),600
	<i>HSI-Carron</i>	63.92	71.81	67.53	66.39	71.94	71.50
	<i>HSV-Smith</i>	68.00	73.28	72.47	71.14	73.22	75.22
$Q_{\#}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3).N, N \times M$	(5,5,5),125, 32000	(7,7,7),343, 87808	(5,5,7),175, 100800			
	$(M_1, M_2, M_3).M$	(4,16,4),256	(4,16,4),256	(6,16,6),576			
	<i>HSI-Carron</i>	71.75	71.83	76.36			
	<i>HSV-Smith</i>	74.03	74.19	76.42			
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3).N, N \times M$	(4,8,5),100, 20480	(4,8,5),100, 40960	(4,8,7),224, 57344	(3,8,7),168, 96768		
	$(M_1, M_2, M_3).M$	(4,8,4),128	(4,16,4),256	(4,16,4),256	(6,16,6),576		
	<i>HSI-Carron</i>	67.83	71.78	71.58	75.06		
	<i>HSV-Smith</i>	69.33	73.97	73.72	75.72		
$Q_{\otimes}^S, Q_{\#}^C$	$(N_1, N_2, N_3).N, N \times M$	(3,8,7),168, 43008	(3,8,7),168, 67200	(3,8,7),168, 96768			
	$(M_1, M_2, M_3).M$	(4,16,4),256	(5,16,5),400	(6,16,6),576			
	<i>HSI-Carron</i>	71.28	73.08	74.36			
	<i>HSV-Smith</i>	73.36	74.33	75.36			

Tabla A.14: Recall promedio en la BD *CorelShel*.

Modelos de cuantización	Tamaño del vocabulario						
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N \times M$	(5,5,5),125, 16625	(5,5,5),125, 60750	(6,6,6),216, 55296	(7,7,7),216, 55296	(7,7,7),343, 166698	(5,5,7),175, 105000
	$(M_1, M_2, M_3), M$	(5,5,5),125	(9,9,9),486	(8,8,4),256	(8,8,4),256	(9,9,9),486	(10,10,6),600
	<i>HSI-Carron</i>	43.25	50.53	43.39	46.25	50.52	49.89
<i>HSV-Smith</i>	47.25	52.75	49.39	50.22	52.69	53.78	
$Q_{\#}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	(5,5,5),125, 32000	(7,7,7),343, 87808	(5,5,7),175, 100800			
	$(M_1, M_2, M_3), M$	(4,1,6,4),256	(4,1,6,4),256	(6,1,6,6),576			
	<i>HSI-Carron</i>	50.56	50.69	54.11			
<i>HSV-Smith</i>	52.44	52.47	53.67				
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	(4,8,5),160, 20480	(4,8,5),160, 40960	(4,8,7),224, 57944	(3,8,7),168, 96768		
	$(M_1, M_2, M_3), M$	(4,8,4),128	(4,1,6,4),256	(4,1,6,4),256	(6,1,6,6),576		
	<i>HSI-Carron</i>	44.03	48.03	48.06	51.22		
<i>HSV-Smith</i>	45.92	49.86	50.11	50.53			
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	(3,8,7),168, 43008	(3,8,7),168, 67200	(3,8,7),168, 96768			
	$(M_1, M_2, M_3), M$	(4,1,6,4),256	(5,1,6,5),400	(6,1,6,6),576			
	<i>HSI-Carron</i>	48.75	49.89	51.36			
<i>HSV-Smith</i>	50.75	50.39	50.72				

Tabla A.15: Recall promedio en la BD *Arenas*.

Modelos de cuantización	Tamaño del vocabulario			
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N \times M$	(5,5,5),125, 16625	(5,5,5),125, 24500	(6,6,6),216, 55296
	$(M_1, M_2, M_3), M$	(5,5,5),125	(7,7,4),196	(8,8,4),256
	<i>HSI-Carron</i>	75.85	75.84	71.03
<i>HSV-Smith</i>	79.31	78.24	74.86	
$Q_{\#}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	(5,5,5),125, 16000	(5,5,5),125, 32000	
	$(M_1, M_2, M_3), M$	(4,8,4),128	(4,1,6,4),256	
	<i>HSI-Carron</i>	73.01	75.56	
<i>HSV-Smith</i>	78.50	76.38		
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	(4,8,5),160, 20480	(4,8,5),160, 40960	
	$(M_1, M_2, M_3), M$	(4,8,4),128	(4,1,6,4),256	
	<i>HSI-Carron</i>	67.76	68.55	
<i>HSV-Smith</i>	71.02	72.89		
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	(3,8,5),120, 24000	(3,8,7),168, 43008	
	$(M_1, M_2, M_3), M$	(5,8,5),200	(4,1,6,4),256	
	<i>HSI-Carron</i>	68.33	66.07	
<i>HSV-Smith</i>	71.33	71.74		

Tabla A.16: Recall promedio en la BD *BDHomo*.

Modelos de cuantización	Tamaño del vocabulario						
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N \times M$	(5,5,5),125, 16625	(5,5,5),125, 60750	(6,6,6),216, 55296	(7,7,7),216, 55296	(7,7,7),343, 166698	(5,5,7),175, 105000
	$(M_1, M_2, M_3), M$	(5,5,5),125	(9,9,9),486	(8,8,4),256	(8,8,4),256	(9,9,9),486	(10,10,6),600
	<i>HSI-Carron</i>	88.26	94.55	95.10	94.06	94.97	95.28
<i>HSV-Smith</i>	93.69	95.59	95.75	94.93	95.86	96.04	
$Q_{\#}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	(5,5,5),125, 32000	(7,7,7),343, 87808	(5,5,7),175, 100800			
	$(M_1, M_2, M_3), M$	(4,1,6,4),256	(4,1,6,4),256	(6,1,6,6),576			
	<i>HSI-Carron</i>	95.49	95.66	97.01			
<i>HSV-Smith</i>	95.56	96.63	97.19				
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	(4,8,5),160, 20480	(4,8,5),160, 40960	(4,8,7),224, 57944	(3,8,7),168, 96768		
	$(M_1, M_2, M_3), M$	(4,8,4),128	(4,1,6,4),256	(4,1,6,4),256	(6,1,6,6),576		
	<i>HSI-Carron</i>	96.63	96.70	96.63	97.22		
<i>HSV-Smith</i>	96.88	96.64	96.81	97.22			
$Q_{\otimes}^S, Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N \times M$	(3,8,7),168, 43008	(3,8,7),168, 67200	(3,8,7),168, 96768			
	$(M_1, M_2, M_3), M$	(4,1,6,4),256	(5,1,6,5),400	(6,1,6,6),576			
	<i>HSI-Carron</i>	96.88	96.98	97.43			
<i>HSV-Smith</i>	96.94	96.91	97.22				

A.0.2. Resultados experimentales con el descriptor *PTD*

Los resultados experimentales de la recuperación de imágenes utilizando el descriptor *PTD* se muestran en las tablas: A.17-A.32. Se ha calculado el *recall promedio*(r), en todas las BD de texturas. En las tablas se muestra este índice siendo r el número de imágenes relevantes de cada BD, idealmente para este valor el *recall promedio* debería de ser 100.

Se ha realizado la experimentación utilizando diferentes vocabularios (según los modelos descritos en el apartado 3.6.2) variando las funciones cuantificación Q_s y Q_c , además de utilizar diferentes espacios para representar el color. La función χ^2 ha sido la utilizada para calcular la similitud entre representaciones.

En cada una de las tablas que se muestran a continuación se ha destacado el negrita el mejor resultado obtenido.

Tabla A.17: Recall promedio en la BD Outex.

Modelos de cuantización	Tamaño del vocabulario						
$Q_{\#}^S Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(5,5,5),125, 611	(7,7,7),343, 829	(9,9,9),729, 1458	(6,6,6),216, 1512		
	$(M_1, M_2, M_3), M$	(9,9,6),486	(9,9,6),486	(9,9,9),729	(9,16,9),1296		
	<i>HSI-Carron</i>	57.85	58.93	61.21	65.64		
	<i>HSV-Yagi</i>	50.24	51.76	54.78	62.71		
<i>HSV-Smith</i>	60.67	61.31	62.21	69.43			
$Q_{\#}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	(5,5,5),125, 681	(7,7,7),343, 1639	(8,8,8),512, 1808			
	$(M_1, M_2, M_3), M$	(4,16,4),256	(9,16,9),1296	(9,16,9),1296			
	<i>HSI-Carron</i>	60.87	68.06	68.67			
	<i>HSV-Yagi</i>	58.06	64.39	65.18			
<i>HSV-Smith</i>	62.99	66.17	66.96				
$Q_{\otimes}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	(4,8,5),160, 288	(4,8,5),160, 416	(4,8,7),224, 480	(4,8,7),224, 1520		
	$(M_1, M_2, M_3), M$	(4,8,4),128	(4,16,4),256	(4,16,4),256	(9,16,9),1296		
	<i>HSI-Carron</i>	67.77	67.50	67.74	71.61		
	<i>HSV-Yagi</i>	65.85	65.42	65.75	69.12		
<i>HSV-Smith</i>	68.60	68.11	68.53	70.73			
$Q_{\otimes}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	(8,8,7),168, 424	(8,8,7),168, 744	(8,8,7),168, 1192	(8,8,7),168, 1464	(4,8,7),224, 1520	(8,8,7),168, 1768
	$(M_1, M_2, M_3), M$	(4,16,4),256	(6,16,6),576	(8,16,8),1024	(9,16,9),1296	(9,16,9),1296	(10,16,10),1600
	<i>HSI-Carron</i>	68.25	69.76	70.88	71.99	72.42	71.36
	<i>HSV-Yagi</i>	66.35	67.08	68.37	69.54	69.94	69.26
<i>HSV-Smith</i>	68.96	70.30	71.02	71.05	71.69	72.14	

Tabla A.18: Recall promedio en la BD VisTexL.

Modelos de cuantización	Tamaño del vocabulario						
$Q_{\#}^S Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(5,5,5),125, 611	(7,7,7),343, 829	(9,9,9),729, 1458	(6,6,6),216, 1512		
	$(M_1, M_2, M_3), M$	(9,9,6),486	(9,9,6),486	(9,9,9),729	(12,12,9),1296		
	<i>HSI-Carron</i>	82.81	82.88	83.52	85.87		
	<i>HSV-Yagi</i>	77.77	78.52	80.67	83.41		
<i>HSV-Smith</i>	82.31	82.98	83.99	86.73			
$Q_{\#}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	(5,5,5),125, 681	(7,7,7),343, 1639	(8,8,8),512, 1808			
	$(M_1, M_2, M_3), M$	(4,16,4),256	(9,16,9),1296	(9,16,9),1296			
	<i>HSI-Carron</i>	82.65	85.64	85.80			
	<i>HSV-Yagi</i>	80.67	85.31	85.55			
<i>HSV-Smith</i>	83.47	85.17	85.40				
$Q_{\otimes}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	(4,8,5),160, 288	(4,8,5),160, 416	(4,8,7),224, 480	(4,8,7),224, 1520		
	$(M_1, M_2, M_3), M$	(4,8,4),128	(4,16,4),256	(4,16,4),256	(9,16,9),1296		
	<i>HSI-Carron</i>	83.84	86.02	86.24	88.00		
	<i>HSV-Yagi</i>	81.76	84.00	84.20	87.82		
<i>HSV-Smith</i>	84.45	86.19	86.42	87.86			
$Q_{\otimes}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	(8,8,7),168, 424	(8,8,7),168, 744	(8,8,7),168, 1192	(8,8,7),168, 1464	(4,8,7),224, 1520	(8,8,7),168, 1768
	$(M_1, M_2, M_3), M$	(4,16,4),256	(6,16,6),576	(8,16,8),1024	(9,16,9),1296	(9,16,9),1296	(10,16,10),1600
	<i>HSI-Carron</i>	86.74	88.22	88.47	88.46	88.79	88.15
	<i>HSV-Yagi</i>	84.93	87.23	87.98	88.37	88.66	88.07
<i>HSV-Smith</i>	87.00	87.21	88.37	88.35	88.70	88.49	

Tabla A.19: Recall promedio en la BD VisTexP.

Modelos de cuantización	Tamaño del vocabulario						
		(5,5,5),125, 611	(7,7,7),343, 829	(9,9,9),729, 1458	(6,6,6),216, 1512		
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(5,5,5),125, 611	(7,7,7),343, 829	(9,9,9),729, 1458	(6,6,6),216, 1512		
	$(M_1, M_2, M_3), M$	(9,9,6),486	(9,9,6),486	(9,9,9),729	(12,12,9),1296		
	<i>HSI-Carron</i>	78.80	79.34	80.22	82.69		
	<i>HSV-Yagi</i>	74.52	74.92	76.78	80.10		
<i>HSV-Smith</i>	79.28	79.83	80.61	82.93			
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(5,5,5),125, 681	(7,7,7),343, 1639	(8,8,8),512, 1808			
	$(M_1, M_2, M_3), M$	(4,16,4),256	(9,16,9),1296	(9,16,9),1296			
	<i>HSI-Carron</i>	77.39	81.47	81.71			
	<i>HSV-Yagi</i>	76.00	81.02	81.20			
<i>HSV-Smith</i>	79.24	81.18	81.35				
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(4,8,5),160, 288	(4,8,5),160, 416	(4,8,7),224, 480	(4,8,7),224, 1520		
	$(M_1, M_2, M_3), M$	(4,8,4),128	(4,16,4),256	(4,16,4),256	(9,16,9),1296		
	<i>HSI-Carron</i>	79.50	81.32	81.71	84.08		
	<i>HSV-Yagi</i>	78.28	80.17	80.33	83.95		
<i>HSV-Smith</i>	80.69	82.07	82.18	83.92			
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(3,8,7),168, 424	(3,8,7),168, 744	(3,8,7),168, 1192	(3,8,7),168, 1464	(4,8,7),224, 1520	(3,8,7),168, 1768
	$(M_1, M_2, M_3), M$	(4,16,4),256	(6,16,6),576	(8,16,8),1024	(9,16,9),1296	(9,16,9),1296	(10,16,10),1600
	<i>HSI-Carron</i>	82.33	84.06	84.38	84.64	85.05	84.44
	<i>HSV-Yagi</i>	80.99	82.99	83.93	84.65	85.05	84.20
<i>HSV-Smith</i>	82.87	83.38	84.66	84.33	84.80	84.78	

Tabla A.20: Recall promedio en la BD CorelTex.

Modelos de cuantización	Tamaño del vocabulario						
		(5,5,5),125, 611	(7,7,7),343, 829	(9,9,9),729, 1458	(6,6,6),216, 1512		
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(5,5,5),125, 611	(7,7,7),343, 829	(9,9,9),729, 1458	(6,6,6),216, 1512		
	$(M_1, M_2, M_3), M$	(9,9,6),486	(9,9,6),486	(9,9,9),729	(12,12,9),1296		
	<i>HSI-Carron</i>	70.33	71.00	73.50	80.47		
	<i>HSV-Yagi</i>	62.69	63.47	66.19	74.64		
<i>HSV-Smith</i>	74.72	75.28	76.08	83.28			
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(5,5,5),125, 681	(7,7,7),343, 1639	(8,8,8),512, 1808			
	$(M_1, M_2, M_3), M$	(4,16,4),256	(9,16,9),1296	(9,16,9),1296			
	<i>HSI-Carron</i>	73.44	78.11	78.89			
	<i>HSV-Yagi</i>	69.56	76.33	76.64			
<i>HSV-Smith</i>	74.47	78.92	79.39				
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(4,8,5),160, 288	(4,8,5),160, 416	(4,8,7),224, 480	(4,8,7),224, 1520		
	$(M_1, M_2, M_3), M$	(4,8,4),128	(4,16,4),256	(4,16,4),256	(9,16,9),1296		
	<i>HSI-Carron</i>	77.28	81.00	81.14	83.97		
	<i>HSV-Yagi</i>	74.83	77.97	78.28	81.81		
<i>HSV-Smith</i>	78.33	81.19	81.33	84.00			
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(3,8,7),168, 424	(3,8,7),168, 744	(3,8,7),168, 1192	(3,8,7),168, 1464	(4,8,7),224, 1520	(3,8,7),168, 1768
	$(M_1, M_2, M_3), M$	(4,16,4),256	(6,16,6),576	(8,16,8),1024	(9,16,9),1296	(9,16,9),1296	(10,16,10),1600
	<i>HSI-Carron</i>	81.86	83.19	84.42	84.64	84.89	84.56
	<i>HSV-Yagi</i>	79.36	80.81	82.03	82.61	83.19	82.89
<i>HSV-Smith</i>	82.56	84.00	84.47	84.69	84.89	84.97	

Tabla A.21: Recall promedio en la BD CorelTex2.

Modelos de cuantización	Tamaño del vocabulario						
		(5,5,5),125, 611	(7,7,7),343, 829	(9,9,9),729, 1458	(6,6,6),216, 1512		
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(5,5,5),125, 611	(7,7,7),343, 829	(9,9,9),729, 1458	(6,6,6),216, 1512		
	$(M_1, M_2, M_3), M$	(9,9,6),486	(9,9,6),486	(9,9,9),729	(12,12,9),1296		
	<i>HSI-Carron</i>	81.06	81.75	82.42	85.46		
	<i>HSV-Yagi</i>	72.44	73.39	75.67	71.03		
<i>HSV-Smith</i>	83.17	83.53	84.47	88.64			
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(5,5,5),125, 681	(7,7,7),343, 1639	(8,8,8),512, 1808			
	$(M_1, M_2, M_3), M$	(4,16,4),256	(9,16,9),1296	(9,16,9),1296			
	<i>HSI-Carron</i>	82.67	88.19	88.67			
	<i>HSV-Yagi</i>	79.61	86.78	87.11			
<i>HSV-Smith</i>	85.19	88.86	89.19				
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(4,8,5),160, 288	(4,8,5),160, 416	(4,8,7),224, 480	(4,8,7),224, 1520		
	$(M_1, M_2, M_3), M$	(4,8,4),128	(4,16,4),256	(4,16,4),256	(9,16,9),1296		
	<i>HSI-Carron</i>	81.89	85.00	85.22	89.72		
	<i>HSV-Yagi</i>	79.58	83.11	83.50	88.50		
<i>HSV-Smith</i>	84.64	87.72	87.83	89.94			
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(3,8,7),168, 424	(3,8,7),168, 744	(3,8,7),168, 1192	(3,8,7),168, 1464	(4,8,7),224, 1520	(3,8,7),168, 1768
	$(M_1, M_2, M_3), M$	(4,16,4),256	(6,16,6),576	(8,16,8),1024	(9,16,9),1296	(9,16,9),1296	(10,16,10),1600
	<i>HSI-Carron</i>	85.92	88.83	89.33	90.11	90.22	90.36
	<i>HSV-Yagi</i>	84.47	87.42	88.61	89.08	89.36	89.56
<i>HSV-Smith</i>	88.31	89.75	90.31	90.50	90.81	90.64	

Tabla A.22: Recall promedio en la BD *CorelV1Tex*.

Modelos de cuantización	Tamaño del vocabulario						
	$(N_1, N_2, N_3), N, N+M$	$(5, 5, 5), 125, 611$	$(7, 7, 7), 343, 829$	$(9, 9, 9), 729, 1458$	$(6, 6, 6), 216, 1512$		
$Q_{\#}^S Q_{\#}^C$	$(M_1, M_2, M_3), M$	$(9, 9, 6), 486$	$(9, 9, 6), 486$	$(9, 9, 9), 729$	$(12, 12, 9), 1296$		
	<i>HSI-Carron</i>	88.94	89.44	90.08	91.94		
	<i>HSV-Yagi</i>	84.28	84.89	86.31	89.50		
	<i>HSV-Smith</i>	91.25	91.64	92.28	93.69		
$Q_{\#}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	$(5, 5, 5), 125, 611$	$(7, 7, 7), 343, 1639$	$(8, 8, 8), 512, 1808$			
	$(M_1, M_2, M_3), M$	$(4, 16, 4), 256$	$(9, 16, 9), 1296$	$(9, 16, 9), 1296$			
	<i>HSI-Carron</i>	91.25	93.19	93.47			
	<i>HSV-Yagi</i>	89.22	92.56	92.69			
$Q_{\otimes}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	$(4, 8, 5), 160, 288$	$(4, 8, 5), 160, 416$	$(4, 8, 7), 224, 480$	$(4, 8, 7), 224, 1520$		
	$(M_1, M_2, M_3), M$	$(4, 8, 4), 128$	$(4, 16, 4), 256$	$(4, 16, 4), 256$	$(9, 16, 9), 1296$		
	<i>HSI-Carron</i>	91.14	92.39	92.58	93.94		
	<i>HSV-Yagi</i>	88.47	90.53	90.72	93.33		
$Q_{\otimes}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	$(3, 8, 7), 168, 424$	$(3, 8, 7), 168, 744$	$(3, 8, 7), 168, 1192$	$(3, 8, 7), 168, 1464$	$(4, 8, 7), 224, 1520$	$(3, 8, 7), 168, 1768$
	$(M_1, M_2, M_3), M$	$(4, 16, 4), 256$	$(6, 16, 6), 576$	$(8, 16, 8), 1024$	$(9, 16, 9), 1296$	$(9, 16, 9), 1296$	$(10, 16, 10), 1600$
	<i>HSI-Carron</i>	93.22	94.22	94.67	94.53	94.61	94.53
	<i>HSV-Yagi</i>	91.81	92.94	93.69	93.92	94.06	94.08
<i>HSV-Smith</i>	93.58	94.25	94.86	94.83	95.00	94.72	

Tabla A.23: Recall promedio en la BD *CorelV2Tex*.

Modelos de cuantización	Tamaño del vocabulario						
	$(N_1, N_2, N_3), N, N+M$	$(5, 5, 5), 125, 611$	$(7, 7, 7), 343, 829$	$(9, 9, 9), 729, 1458$	$(6, 6, 6), 216, 1512$		
$Q_{\#}^S Q_{\#}^C$	$(M_1, M_2, M_3), M$	$(9, 9, 6), 486$	$(9, 9, 6), 486$	$(9, 9, 9), 729$	$(12, 12, 9), 1296$		
	<i>HSI-Carron</i>	91.28	91.53	91.83	93.06		
	<i>HSV-Yagi</i>	87.11	87.58	88.42	91.33		
	<i>HSV-Smith</i>	92.00	92.36	93.03	93.69		
$Q_{\#}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	$(5, 5, 5), 125, 611$	$(7, 7, 7), 343, 1639$	$(8, 8, 8), 512, 1808$			
	$(M_1, M_2, M_3), M$	$(4, 16, 4), 256$	$(9, 16, 9), 1296$	$(9, 16, 9), 1296$			
	<i>HSI-Carron</i>	91.22	92.89	93.31			
	<i>HSV-Yagi</i>	89.14	92.31	92.69			
$Q_{\otimes}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	$(4, 8, 5), 160, 288$	$(4, 8, 5), 160, 416$	$(4, 8, 7), 224, 480$	$(4, 8, 7), 224, 1520$		
	$(M_1, M_2, M_3), M$	$(4, 8, 4), 128$	$(4, 16, 4), 256$	$(4, 16, 4), 256$	$(9, 16, 9), 1296$		
	<i>HSI-Carron</i>	91.14	92.50	92.53	93.75		
	<i>HSV-Yagi</i>	88.56	90.36	90.33	93.11		
$Q_{\otimes}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	$(3, 8, 7), 168, 424$	$(3, 8, 7), 168, 744$	$(3, 8, 7), 168, 1192$	$(3, 8, 7), 168, 1464$	$(4, 8, 7), 224, 1520$	$(3, 8, 7), 168, 1768$
	$(M_1, M_2, M_3), M$	$(4, 16, 4), 256$	$(6, 16, 6), 576$	$(8, 16, 8), 1024$	$(9, 16, 9), 1296$	$(9, 16, 9), 1296$	$(10, 16, 10), 1600$
	<i>HSI-Carron</i>	93.61	93.94	94.28	94.58	94.75	94.67
	<i>HSV-Yagi</i>	91.47	93.19	94.00	94.22	94.39	94.25
<i>HSV-Smith</i>	93.47	94.50	95.14	95.03	95.17	94.89	

Tabla A.24: Recall promedio en la BD *CorelTexPat*.

Modelos de cuantización	Tamaño del vocabulario						
	$(N_1, N_2, N_3), N, N+M$	$(5, 5, 5), 125, 611$	$(7, 7, 7), 343, 829$	$(9, 9, 9), 729, 1458$	$(6, 6, 6), 216, 1512$		
$Q_{\#}^S Q_{\#}^C$	$(M_1, M_2, M_3), M$	$(9, 9, 6), 486$	$(9, 9, 6), 486$	$(9, 9, 9), 729$	$(12, 12, 9), 1296$		
	<i>HSI-Carron</i>	91.03	91.19	91.67	93.53		
	<i>HSV-Yagi</i>	86.28	86.61	88.75	91.61		
	<i>HSV-Smith</i>	92.44	92.47	93.19	94.06		
$Q_{\#}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	$(5, 5, 5), 125, 611$	$(7, 7, 7), 343, 1639$	$(8, 8, 8), 512, 1808$			
	$(M_1, M_2, M_3), M$	$(4, 16, 4), 256$	$(9, 16, 9), 1296$	$(9, 16, 9), 1296$			
	<i>HSI-Carron</i>	90.97	93.28	93.61			
	<i>HSV-Yagi</i>	90.56	92.89	93.19			
$Q_{\otimes}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	$(4, 8, 5), 160, 288$	$(4, 8, 5), 160, 416$	$(4, 8, 7), 224, 480$	$(4, 8, 7), 224, 1520$		
	$(M_1, M_2, M_3), M$	$(4, 8, 4), 128$	$(4, 16, 4), 256$	$(4, 16, 4), 256$	$(9, 16, 9), 1296$		
	<i>HSI-Carron</i>	92.03	93.19	93.50	94.50		
	<i>HSV-Yagi</i>	90.39	92.14	92.33	94.19		
$Q_{\otimes}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	$(3, 8, 7), 168, 424$	$(3, 8, 7), 168, 744$	$(3, 8, 7), 168, 1192$	$(3, 8, 7), 168, 1464$	$(4, 8, 7), 224, 1520$	$(3, 8, 7), 168, 1768$
	$(M_1, M_2, M_3), M$	$(4, 16, 4), 256$	$(6, 16, 6), 576$	$(8, 16, 8), 1024$	$(9, 16, 9), 1296$	$(9, 16, 9), 1296$	$(10, 16, 10), 1600$
	<i>HSI-Carron</i>	93.67	94.19	94.56	94.72	94.81	94.97
	<i>HSV-Yagi</i>	92.69	93.53	94.69	94.58	94.72	94.72
<i>HSV-Smith</i>	93.44	94.61	94.75	95.19	95.25	95.11	

Tabla A.25: Recall promedio en la BD *CorelSand*.

Modelos de cuantización	Tamaño del vocabulario						
		(5,5),125, 611	(7,7),343, 829	(9,9),729, 1458	(6,6),216, 1512		
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(5,5),125, 611	(7,7),343, 829	(9,9),729, 1458	(6,6),216, 1512		
	$(M_1, M_2, M_3), M$	(9,9),486	(9,9),486	(9,9),729	(12,12),1296		
	<i>HSL-Carron</i>	76.39	76.61	77.78	82.58		
	<i>HSV-Yagi</i>	62.56	63.42	64.81	78.39		
<i>HSV-Smith</i>	81.75	81.50	82.53	84.82			
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(5,5),125, 681	(7,7),343, 1639	(8,8),512, 1808			
	$(M_1, M_2, M_3), M$	(4,16,4),256	(9,16,9),1296	(9,16,9),1296			
	<i>HSL-Carron</i>	83.14	85.89	86.08			
	<i>HSV-Yagi</i>	80.42	84.64	84.97			
<i>HSV-Smith</i>	84.28	86.67	86.81				
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(4,8,5),160, 288	(4,8,5),160, 416	(4,8,7),224, 480	(4,8,7),224, 1520		
	$(M_1, M_2, M_3), M$	(4,8,4),128	(4,16,4),256	(4,16,4),256	(9,16,9),1296		
	<i>HSL-Carron</i>	80.94	82.75	82.89	95.92		
	<i>HSV-Yagi</i>	78.17	80.81	80.94	84.89		
<i>HSV-Smith</i>	82.39	84.14	84.22	86.11			
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(3,8,7),168, 424	(3,8,7),168, 744	(3,8,7),168, 1192	(3,8,7),168, 1464	(4,8,7),224, 1520	(3,8,7),168, 1768
	$(M_1, M_2, M_3), M$	(4,16,4),256	(6,16,6),576	(8,16,8),1024	(9,16,9),1296	(9,16,9),1296	(10,16,10),1600
	<i>HSL-Carron</i>	83.56	85.19	86.31	86.31	86.44	86.67
	<i>HSV-Yagi</i>	81.67	83.67	85.00	85.06	85.03	85.47
<i>HSV-Smith</i>	84.78	86.14	86.83	86.64	86.86	87.06	

Tabla A.26: Recall promedio en la BD *CorelBark*.

Modelos de cuantización	Tamaño del vocabulario						
		(5,5),125, 611	(7,7),343, 829	(9,9),729, 1458	(6,6),216, 1512		
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(5,5),125, 611	(7,7),343, 829	(9,9),729, 1458	(6,6),216, 1512		
	$(M_1, M_2, M_3), M$	(9,9),486	(9,9),486	(9,9),729	(12,12),1296		
	<i>HSL-Carron</i>	62.64	63.17	64.69	71.94		
	<i>HSV-Yagi</i>	55.89	57.08	59.03	67.03		
<i>HSV-Smith</i>	68.69	69.17	70.61	75.42			
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(5,5),125, 681	(7,7),343, 1639	(8,8),512, 1808			
	$(M_1, M_2, M_3), M$	(4,16,4),256	(9,16,9),1296	(9,16,9),1296			
	<i>HSL-Carron</i>	66.72	72.89	73.47			
	<i>HSV-Yagi</i>	62.03	71.61	72.22			
<i>HSV-Smith</i>	69.58	73.72	74.08				
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(4,8,5),160, 288	(4,8,5),160, 416	(4,8,7),224, 480	(4,8,7),224, 1520		
	$(M_1, M_2, M_3), M$	(4,8,4),128	(4,16,4),256	(4,16,4),256	(9,16,9),1296		
	<i>HSL-Carron</i>	67.58	71.14	71.58	76.56		
	<i>HSV-Yagi</i>	64.56	68.83	69.03	76.08		
<i>HSV-Smith</i>	71.28	73.81	74.11	77.89			
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(3,8,7),168, 424	(3,8,7),168, 744	(3,8,7),168, 1192	(3,8,7),168, 1464	(4,8,7),224, 1520	(3,8,7),168, 1768
	$(M_1, M_2, M_3), M$	(4,16,4),256	(6,16,6),576	(8,16,8),1024	(9,16,9),1296	(9,16,9),1296	(10,16,10),1600
	<i>HSL-Carron</i>	72.39	75.44	76.61	77.31	78.33	77.97
	<i>HSV-Yagi</i>	70.17	73.97	76.33	76.61	76.92	77.31
<i>HSV-Smith</i>	75.00	76.78	78.17	78.22	79.36	78.83	

Tabla A.27: Recall promedio en la BD *CorelCol*.

Modelos de cuantización	Tamaño del vocabulario						
		(5,5),125, 611	(7,7),343, 829	(9,9),729, 1458	(6,6),216, 1512		
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(5,5),125, 611	(7,7),343, 829	(9,9),729, 1458	(6,6),216, 1512		
	$(M_1, M_2, M_3), M$	(9,9),486	(9,9),486	(9,9),729	(12,12),1296		
	<i>HSL-Carron</i>	75.95	76.31	77.16	78.79		
	<i>HSV-Yagi</i>	71.27	71.80	72.94	77.25		
<i>HSV-Smith</i>	77.29	77.91	78.92	79.90			
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(5,5),125, 681	(7,7),343, 1639	(8,8),512, 1808			
	$(M_1, M_2, M_3), M$	(4,16,4),256	(9,16,9),1296	(9,16,9),1296			
	<i>HSL-Carron</i>	74.71	78.66	79.35			
	<i>HSV-Yagi</i>	73.27	77.84	78.65			
<i>HSV-Smith</i>	76.83	79.31	79.71				
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(4,8,5),160, 288	(4,8,5),160, 416	(4,8,7),224, 480	(4,8,7),224, 1520		
	$(M_1, M_2, M_3), M$	(4,8,4),128	(4,16,4),256	(4,16,4),256	(9,16,9),1296		
	<i>HSL-Carron</i>	75.78	77.58	77.75	80.49		
	<i>HSV-Yagi</i>	74.87	76.34	76.57	79.67		
<i>HSV-Smith</i>	77.78	79.22	79.35	80.88			
$Q_{\#}^S, Q_{\#}^C$	$(N_1, N_2, N_3), N, N+M$	(3,8,7),168, 424	(3,8,7),168, 744	(3,8,7),168, 1192	(3,8,7),168, 1464	(4,8,7),224, 1520	(3,8,7),168, 1768
	$(M_1, M_2, M_3), M$	(4,16,4),256	(6,16,6),576	(8,16,8),1024	(9,16,9),1296	(9,16,9),1296	(10,16,10),1600
	<i>HSL-Carron</i>	78.17	80.46	81.31	80.92	81.34	81.57
	<i>HSV-Yagi</i>	77.39	79.15	79.67	80.33	80.75	80.69
<i>HSV-Smith</i>	80.20	80.59	81.21	81.41	81.70	81.60	

Tabla A.28: Recall promedio en la BD *CorelMarb*.

Modelos de cuantización	Tamaño del vocabulario						
	$(N_1, N_2, N_3), N, N+M$	$(5, 5, 5), 125, 611$	$(7, 7, 7), 343, 829$	$(9, 9, 9), 729, 1458$	$(6, 6, 6), 216, 1512$		
$Q_{\#}^S Q_{\#}^C$	$(M_1, M_2, M_3), M$	$(9, 9, 6), 486$	$(9, 9, 6), 486$	$(9, 9, 9), 729$	$(12, 12, 9), 1296$		
	<i>HSI-Carron</i>	68.75	69.19	71.06	77.69		
	<i>HSV-Yagi</i>	58.86	59.17	62.33	72.47		
	<i>HSV-Smith</i>	75.25	75.61	76.56	81.03		
$Q_{\#}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	$(5, 5, 5), 125, 611$	$(7, 7, 7), 343, 1639$	$(8, 8, 8), 512, 1808$			
	$(M_1, M_2, M_3), M$	$(4, 16, 4), 256$	$(9, 16, 9), 1296$	$(9, 16, 9), 1296$			
	<i>HSI-Carron</i>	73.19	82.17	82.36			
	<i>HSV-Yagi</i>	70.89	79.97	80.31			
$Q_{\otimes}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	$(4, 8, 5), 160, 288$	$(4, 8, 5), 160, 416$	$(4, 8, 7), 224, 480$	$(4, 8, 7), 224, 1520$		
	$(M_1, M_2, M_3), M$	$(4, 8, 4), 128$	$(4, 16, 4), 256$	$(4, 16, 4), 256$	$(9, 16, 9), 1296$		
	<i>HSI-Carron</i>	73.97	77.33	77.47	84.97		
	<i>HSV-Yagi</i>	71.58	75.33	75.33	82.39		
$Q_{\otimes}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	$(3, 8, 7), 168, 424$	$(3, 8, 7), 168, 744$	$(3, 8, 7), 168, 1192$	$(3, 8, 7), 168, 1464$	$(4, 8, 7), 224, 1520$	$(3, 8, 7), 168, 1768$
	$(M_1, M_2, M_3), M$	$(4, 16, 4), 256$	$(6, 16, 6), 576$	$(8, 16, 8), 1024$	$(9, 16, 9), 1296$	$(9, 16, 9), 1296$	$(10, 16, 10), 1600$
	<i>HSI-Carron</i>	77.94	82.67	83.61	84.86	85.06	84.72
	<i>HSV-Yagi</i>	75.94	80.56	81.50	82.67	82.89	82.72
<i>HSV-Smith</i>	80.78	84.39	85.78	86.58	86.56	86.56	

Tabla A.29: Recall promedio en la BD *CorelPain*.

Modelos de cuantización	Tamaño del vocabulario						
	$(N_1, N_2, N_3), N, N+M$	$(5, 5, 5), 125, 611$	$(7, 7, 7), 343, 829$	$(9, 9, 9), 729, 1458$	$(6, 6, 6), 216, 1512$		
$Q_{\#}^S Q_{\#}^C$	$(M_1, M_2, M_3), M$	$(9, 9, 6), 486$	$(9, 9, 6), 486$	$(9, 9, 9), 729$	$(12, 12, 9), 1296$		
	<i>HSI-Carron</i>	71.58	72.00	73.03	75.28		
	<i>HSV-Yagi</i>	66.81	67.08	68.61	73.03		
	<i>HSV-Smith</i>	73.58	73.61	75.17	78.19		
$Q_{\#}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	$(5, 5, 5), 125, 611$	$(7, 7, 7), 343, 1639$	$(8, 8, 8), 512, 1808$			
	$(M_1, M_2, M_3), M$	$(4, 16, 4), 256$	$(9, 16, 9), 1296$	$(9, 16, 9), 1296$			
	<i>HSI-Carron</i>	71.83	77.75	77.89			
	<i>HSV-Yagi</i>	69.33	75.69	75.97			
$Q_{\otimes}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	$(4, 8, 5), 160, 288$	$(4, 8, 5), 160, 416$	$(4, 8, 7), 224, 480$	$(4, 8, 7), 224, 1520$		
	$(M_1, M_2, M_3), M$	$(4, 8, 4), 128$	$(4, 16, 4), 256$	$(4, 16, 4), 256$	$(9, 16, 9), 1296$		
	<i>HSI-Carron</i>	68.86	73.42	73.61	78.92		
	<i>HSV-Yagi</i>	67.28	71.69	71.94	78.03		
$Q_{\otimes}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	$(3, 8, 7), 168, 424$	$(3, 8, 7), 168, 744$	$(3, 8, 7), 168, 1192$	$(3, 8, 7), 168, 1464$	$(4, 8, 7), 224, 1520$	$(3, 8, 7), 168, 1768$
	$(M_1, M_2, M_3), M$	$(4, 16, 4), 256$	$(6, 16, 6), 576$	$(8, 16, 8), 1024$	$(9, 16, 9), 1296$	$(9, 16, 9), 1296$	$(10, 16, 10), 1600$
	<i>HSI-Carron</i>	73.97	77.28	78.19	79.06	79.11	79.58
	<i>HSV-Yagi</i>	72.39	76.08	77.97	78.25	78.42	78.92
<i>HSV-Smith</i>	76.25	78.56	79.42	80.00	80.14	79.92	

Tabla A.30: Recall promedio en la BD *CorelShel*.

Modelos de cuantización	Tamaño del vocabulario						
	$(N_1, N_2, N_3), N, N+M$	$(5, 5, 5), 125, 611$	$(7, 7, 7), 343, 829$	$(9, 9, 9), 729, 1458$	$(6, 6, 6), 216, 1512$		
$Q_{\#}^S Q_{\#}^C$	$(M_1, M_2, M_3), M$	$(9, 9, 6), 486$	$(9, 9, 6), 486$	$(9, 9, 9), 729$	$(12, 12, 9), 1296$		
	<i>HSI-Carron</i>	49.33	50.17	52.53	51.53		
	<i>HSV-Yagi</i>	44.58	45.72	47.78	47.25		
	<i>HSV-Smith</i>	51.72	52.53	53.92	55.69		
$Q_{\#}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	$(5, 5, 5), 125, 611$	$(7, 7, 7), 343, 1639$	$(8, 8, 8), 512, 1808$			
	$(M_1, M_2, M_3), M$	$(4, 16, 4), 256$	$(9, 16, 9), 1296$	$(9, 16, 9), 1296$			
	<i>HSI-Carron</i>	49.72	55.53	56.25			
	<i>HSV-Yagi</i>	46.97	54.36	55.08			
$Q_{\otimes}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	$(4, 8, 5), 160, 288$	$(4, 8, 5), 160, 416$	$(4, 8, 7), 224, 480$	$(4, 8, 7), 224, 1520$		
	$(M_1, M_2, M_3), M$	$(4, 8, 4), 128$	$(4, 16, 4), 256$	$(4, 16, 4), 256$	$(9, 16, 9), 1296$		
	<i>HSI-Carron</i>	44.94	50.14	50.36	55.86		
	<i>HSV-Yagi</i>	42.22	47.53	48.00	53.75		
$Q_{\otimes}^S Q_{\otimes}^C$	$(N_1, N_2, N_3), N, N+M$	$(3, 8, 7), 168, 424$	$(3, 8, 7), 168, 744$	$(3, 8, 7), 168, 1192$	$(3, 8, 7), 168, 1464$	$(4, 8, 7), 224, 1520$	$(3, 8, 7), 168, 1768$
	$(M_1, M_2, M_3), M$	$(4, 16, 4), 256$	$(6, 16, 6), 576$	$(8, 16, 8), 1024$	$(9, 16, 9), 1296$	$(9, 16, 9), 1296$	$(10, 16, 10), 1600$
	<i>HSI-Carron</i>	52.33	55.86	56.78	57.25	57.56	57.44
	<i>HSV-Yagi</i>	49.78	52.86	54.92	55.89	56.42	56.19
<i>HSV-Smith</i>	53.97	55.22	57.44	57.67	58.03	57.78	

Tabla A.31: Recall promedio en la BD Arenas.

Modelos de cuantización	Tamaño del vocabulario						
		(5,5,5),125, 611	(7,7,7),343, 829	(9,9,9),729, 1458	(6,6,6),216, 1512		
$Q_{\#}^{\otimes} Q_{\#}^{\otimes}$	$(N_1, N_2, N_3), N, N+M$	(5,5,5),125, 611	(7,7,7),343, 829	(9,9,9),729, 1458	(6,6,6),216, 1512		
	$(M_1, M_2, M_3), M$	(9,9,9),486	(9,9,9),486	(9,9,9),729	(12,12,9),1296		
	<i>HSI-Carron</i>	80.17	80.27	80.76	77.91		
	<i>HSV-Yagi</i>	77.11	77.31	78.12	79.82		
<i>HSV-Smith</i>	81.82	81.90	81.86	81.15			
$Q_{\#}^{\otimes} Q_{\otimes}^{\otimes}$	$(N_1, N_2, N_3), N, N+M$	(5,5,5),125, 681	(7,7,7),343, 1639	(8,8,8),512, 1808			
	$(M_1, M_2, M_3), M$	(4,16,4),256	(9,16,9),1296	(9,16,9),1296			
	<i>HSI-Carron</i>	75.54	80.22	80.59			
	<i>HSV-Yagi</i>	77.37	80.25	80.63			
<i>HSV-Smith</i>	78.53	80.12	80.45				
$Q_{\otimes}^{\otimes} Q_{\otimes}^{\otimes}$	$(N_1, N_2, N_3), N, N+M$	(4,8,5),160, 288	(4,8,5),160, 416	(4,8,7),224, 480	(4,8,7),224, 1520		
	$(M_1, M_2, M_3), M$	(4,8,4),128	(4,16,4),256	(4,16,4),256	(9,16,9),1296		
	<i>HSI-Carron</i>	70.93	73.82	73.96	78.99		
	<i>HSV-Yagi</i>	73.28	75.98	75.66	79.21		
<i>HSV-Smith</i>	75.06	77.34	77.38	79.13			
$Q_{\otimes}^{\otimes} Q_{\#}^{\otimes}$	$(N_1, N_2, N_3), N, N+M$	(3,8,7),168, 424	(3,8,7),168, 744	(3,8,7),168, 1192	(3,8,7),168, 1464	(4,8,7),224, 1520	(3,8,7),168, 1768
	$(M_1, M_2, M_3), M$	(4,16,4),256	(6,16,6),576	(8,16,8),1024	(9,16,9),1296	(9,16,9),1296	(10,16,10),1600
	<i>HSI-Carron</i>	76.37	79.28	80.56	80.48	81.28	80.57
	<i>HSV-Yagi</i>	78.41	79.83	80.29	80.61	81.25	80.55
<i>HSV-Smith</i>	79.12	80.69	80.57	80.47	81.12	80.47	

Tabla A.32: Recall promedio en la BD *BDHomo*.

Modelos de cuantización	Tamaño del vocabulario						
		(5,5,5),125, 611	(7,7,7),343, 829	(9,9,9),729, 1458	(6,6,6),216, 1512		
$Q_{\#}^{\otimes} Q_{\#}^{\otimes}$	$(N_1, N_2, N_3), N, N+M$	(5,5,5),125, 611	(7,7,7),343, 829	(9,9,9),729, 1458	(6,6,6),216, 1512		
	$(M_1, M_2, M_3), M$	(9,9,9),486	(9,9,9),486	(9,9,9),729	(12,12,9),1296		
	<i>HSI-Carron</i>	94.50	94.65	95.10	93.28		
	<i>HSV-Yagi</i>	91.94	92.64	93.37	95.73		
<i>HSV-Smith</i>	95.76	95.94	95.38	96.46			
$Q_{\#}^{\otimes} Q_{\otimes}^{\otimes}$	$(N_1, N_2, N_3), N, N+M$	(5,5,5),125, 681	(7,7,7),343, 1639	(8,8,8),512, 1808			
	$(M_1, M_2, M_3), M$	(4,16,4),256	(9,16,9),1296	(9,16,9),1296			
	<i>HSI-Carron</i>	95.87	97.19	97.33			
	<i>HSV-Yagi</i>	95.31	96.88	96.94			
<i>HSV-Smith</i>	96.63	97.43	97.47				
$Q_{\otimes}^{\otimes} Q_{\otimes}^{\otimes}$	$(N_1, N_2, N_3), N, N+M$	(4,8,5),160, 288	(4,8,5),160, 416	(4,8,7),224, 480	(4,8,7),224, 1520		
	$(M_1, M_2, M_3), M$	(4,8,4),128	(4,16,4),256	(4,16,4),256	(9,16,9),1296		
	<i>HSI-Carron</i>	96.49	96.79	96.08	97.60		
	<i>HSV-Yagi</i>	94.35	96.88	96.91	97.57		
<i>HSV-Smith</i>	96.84	97.26	97.26	97.67			
$Q_{\otimes}^{\otimes} Q_{\#}^{\otimes}$	$(N_1, N_2, N_3), N, N+M$	(3,8,7),168, 424	(3,8,7),168, 744	(3,8,7),168, 1192	(3,8,7),168, 1464	(4,8,7),224, 1520	(3,8,7),168, 1768
	$(M_1, M_2, M_3), M$	(4,16,4),256	(6,16,6),576	(8,16,8),1024	(9,16,9),1296	(9,16,9),1296	(10,16,10),1600
	<i>HSI-Carron</i>	97.47	98.26	98.19	98.19	98.37	98.13
	<i>HSV-Yagi</i>	97.29	98.19	98.19	98.09	98.37	98.37
<i>HSV-Smith</i>	97.67	98.16	98.30	98.19	98.30	98.26	

Apéndice B

BD de imágenes

A continuación se muestra el contenido en las diferentes Bases de datos de imágenes utilizadas en la experimentación.



Figura B.1: BD Arenas.



Figura B.2: BD VisTexP.

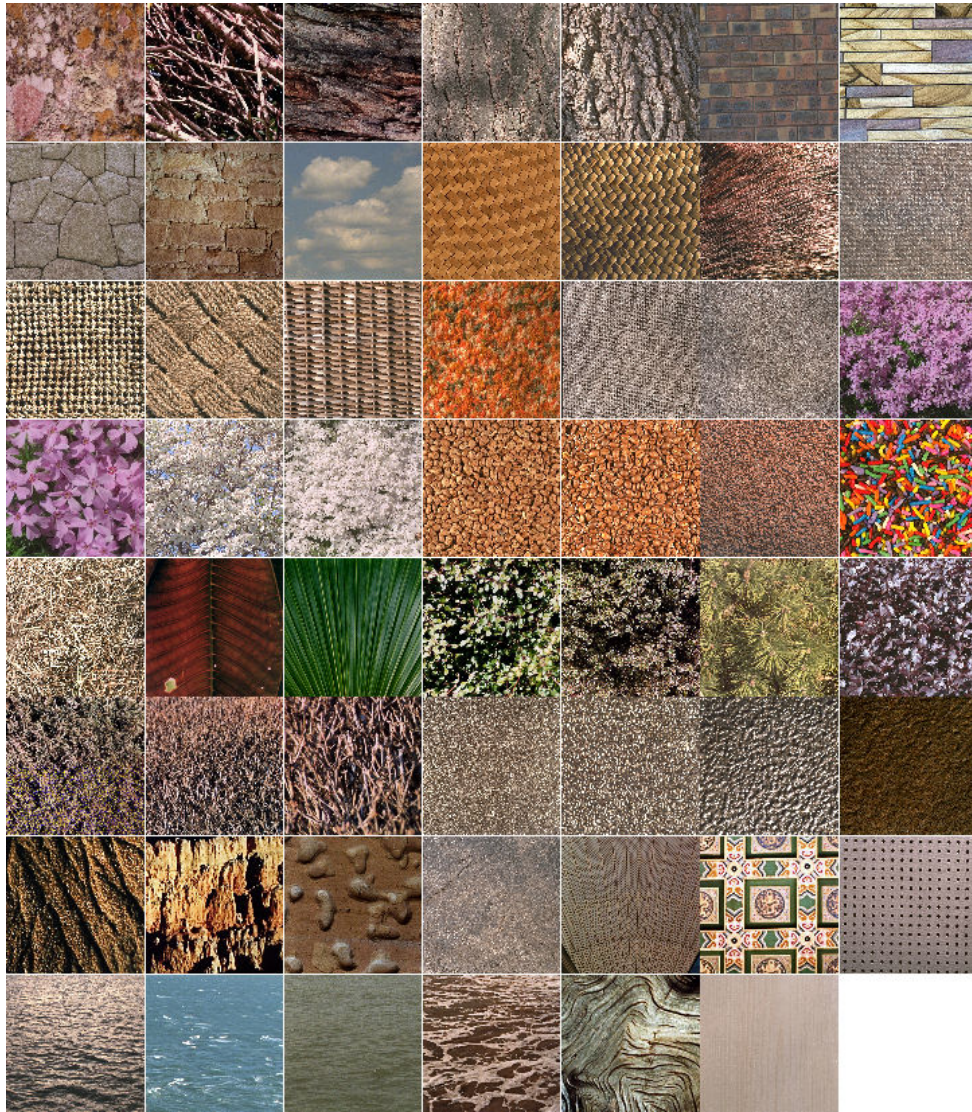


Figura B.3: BD VisTexL.

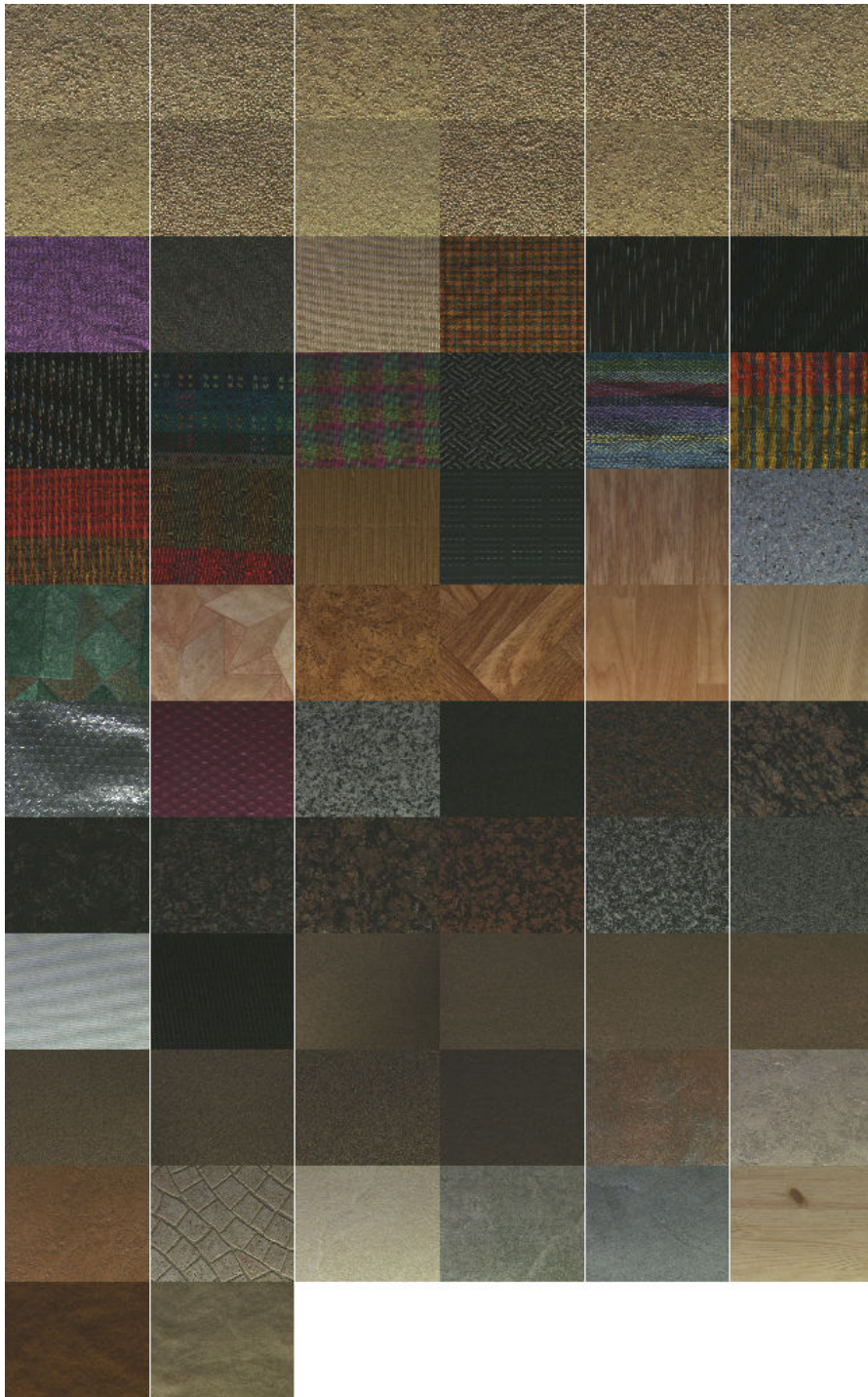


Figura B.4: BD Outex.



Figura B.5: Imágenes de la BD Mayang incluidas en *BDHomo*.



Figura B.6: Imágenes de la BD Outex incluidas en *BDHomo*.



Figura B.7: Imágenes de la BD VisTex incluidas en *BDHomo*.

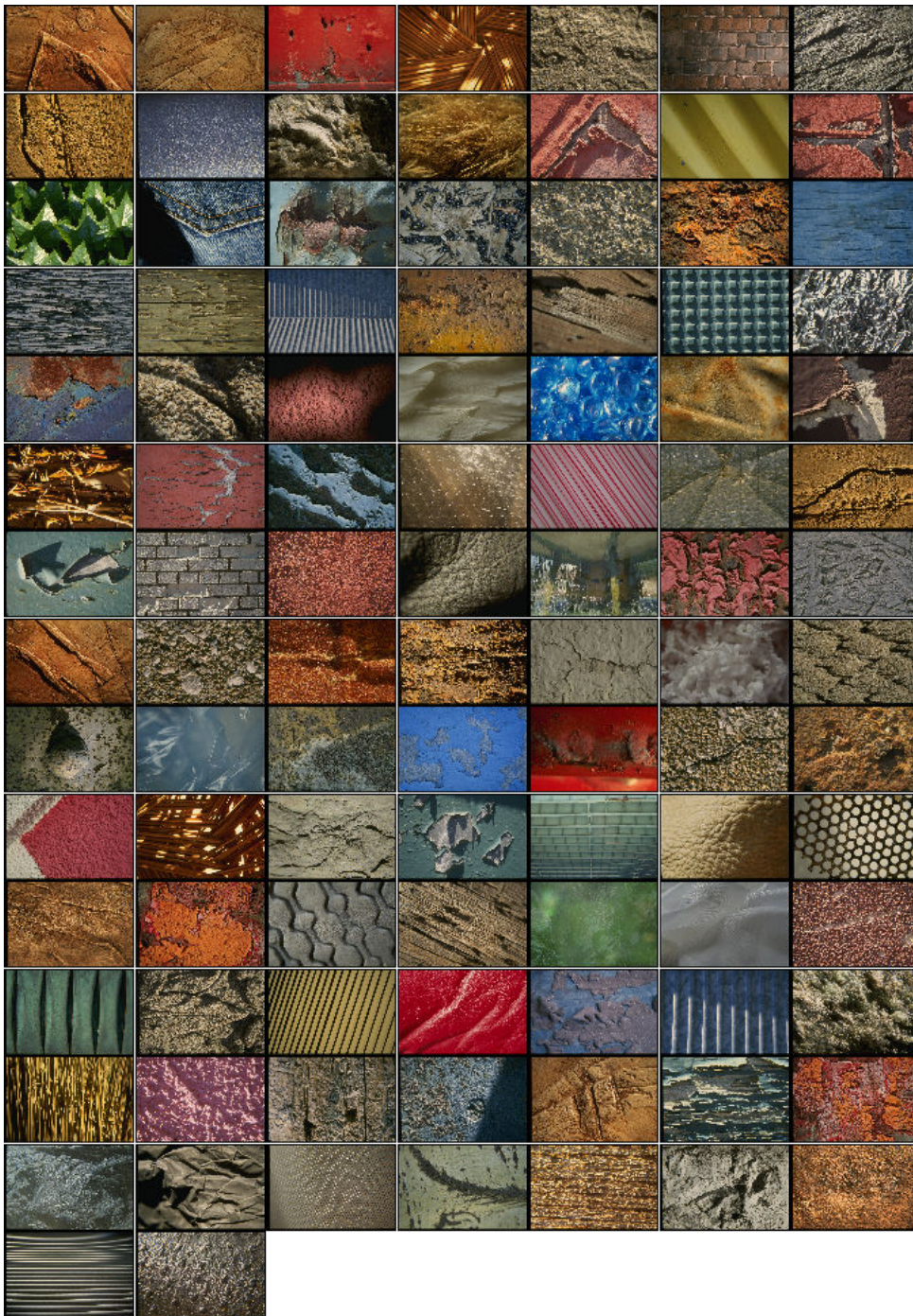


Figura B.8: BD CorelTex.

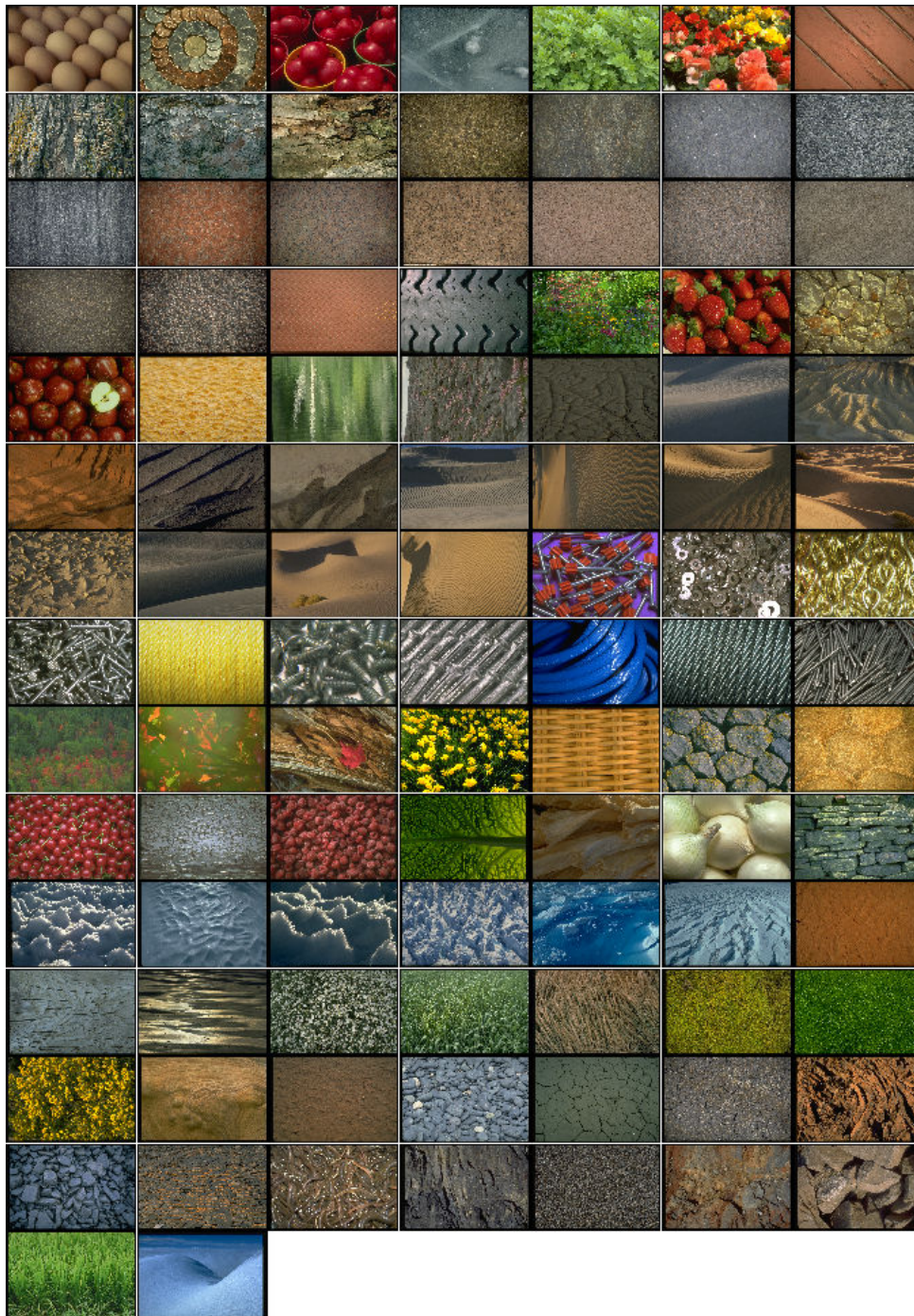


Figura B.9: BD CorelTex2.



Figura B.10: BD CoreV1Tex.

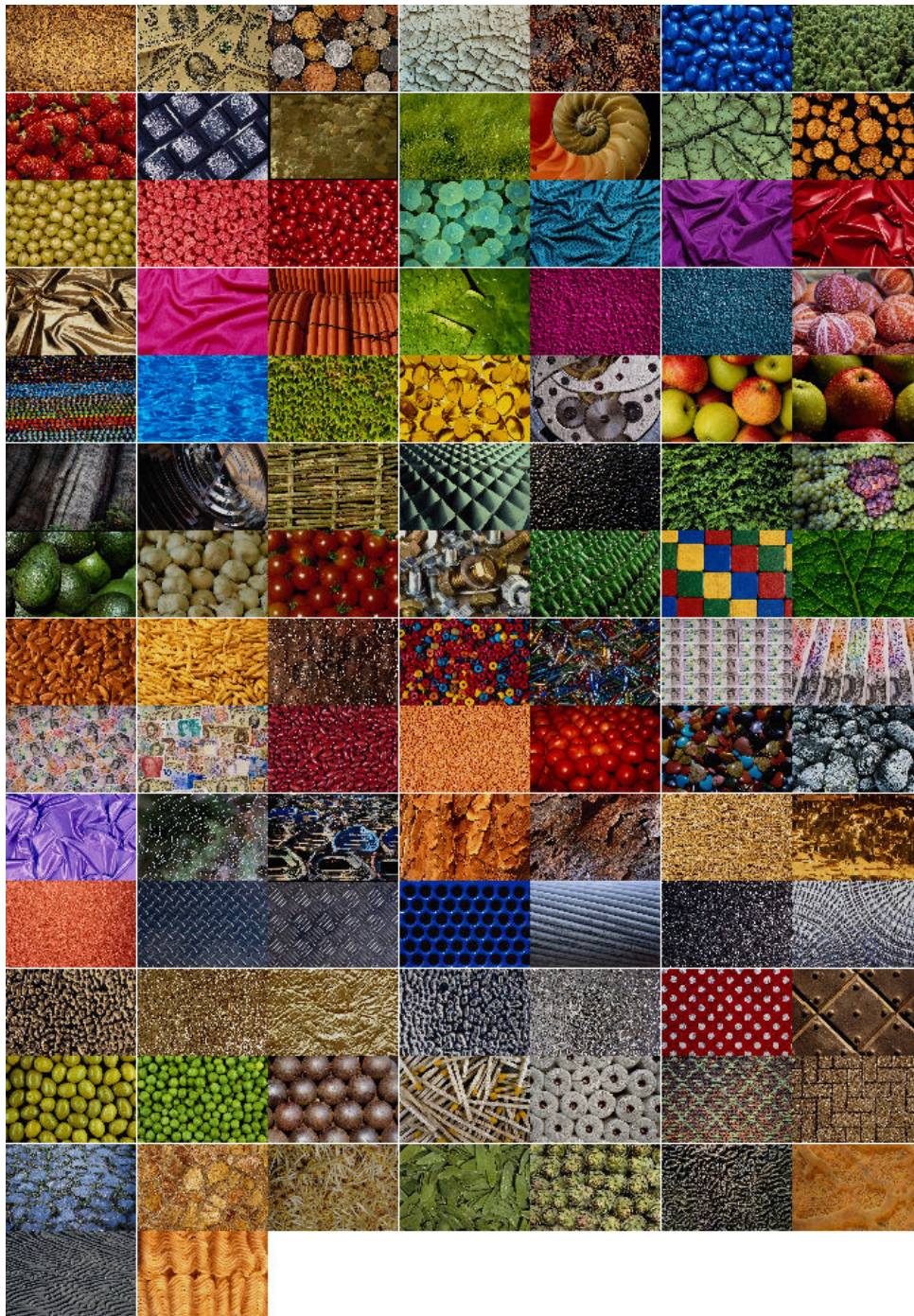


Figura B.11: BD CorelV2Tex.



Figura B.12: BD CorelTexPattern.

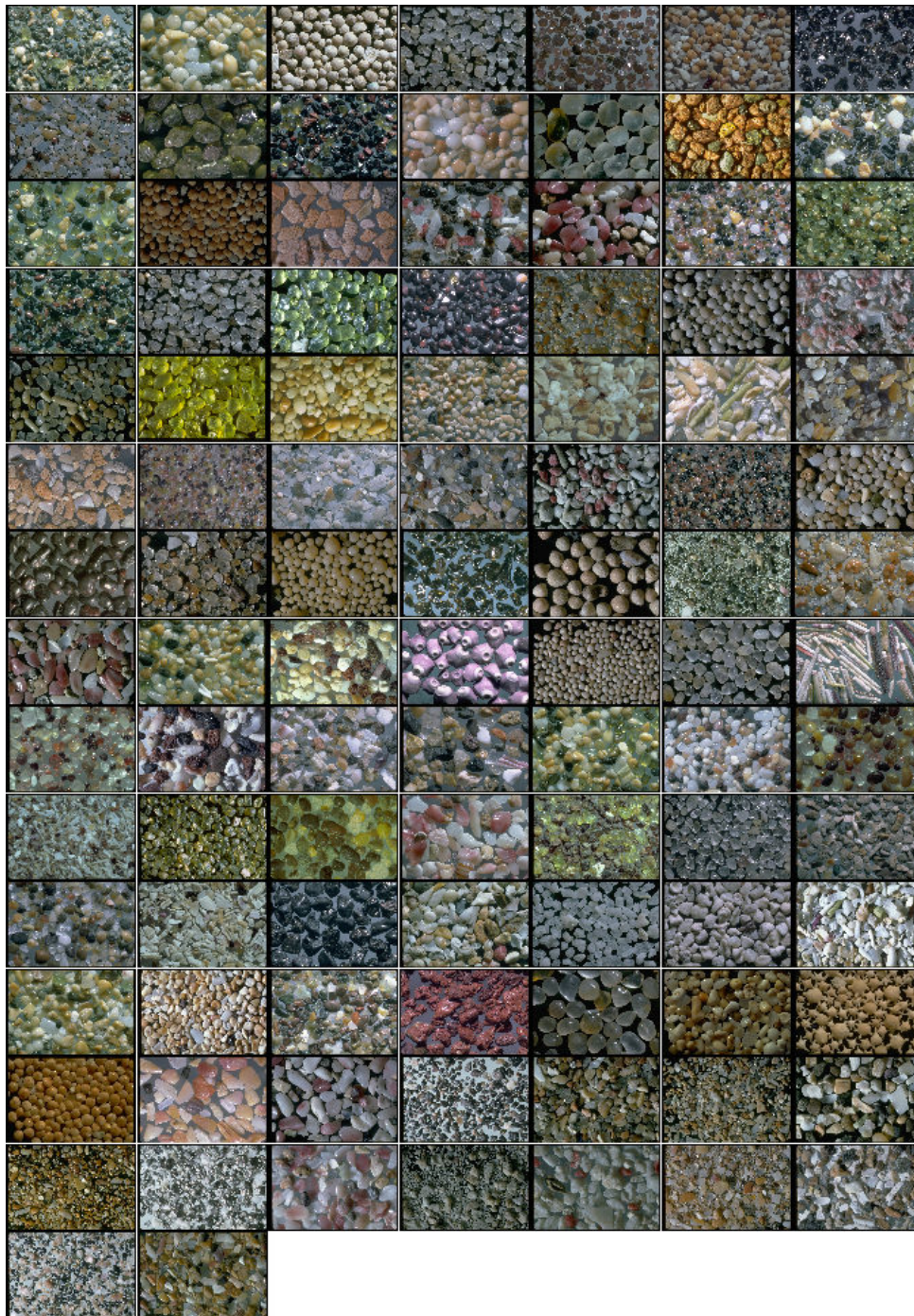


Figura B.13: BD CorelSandPeb.

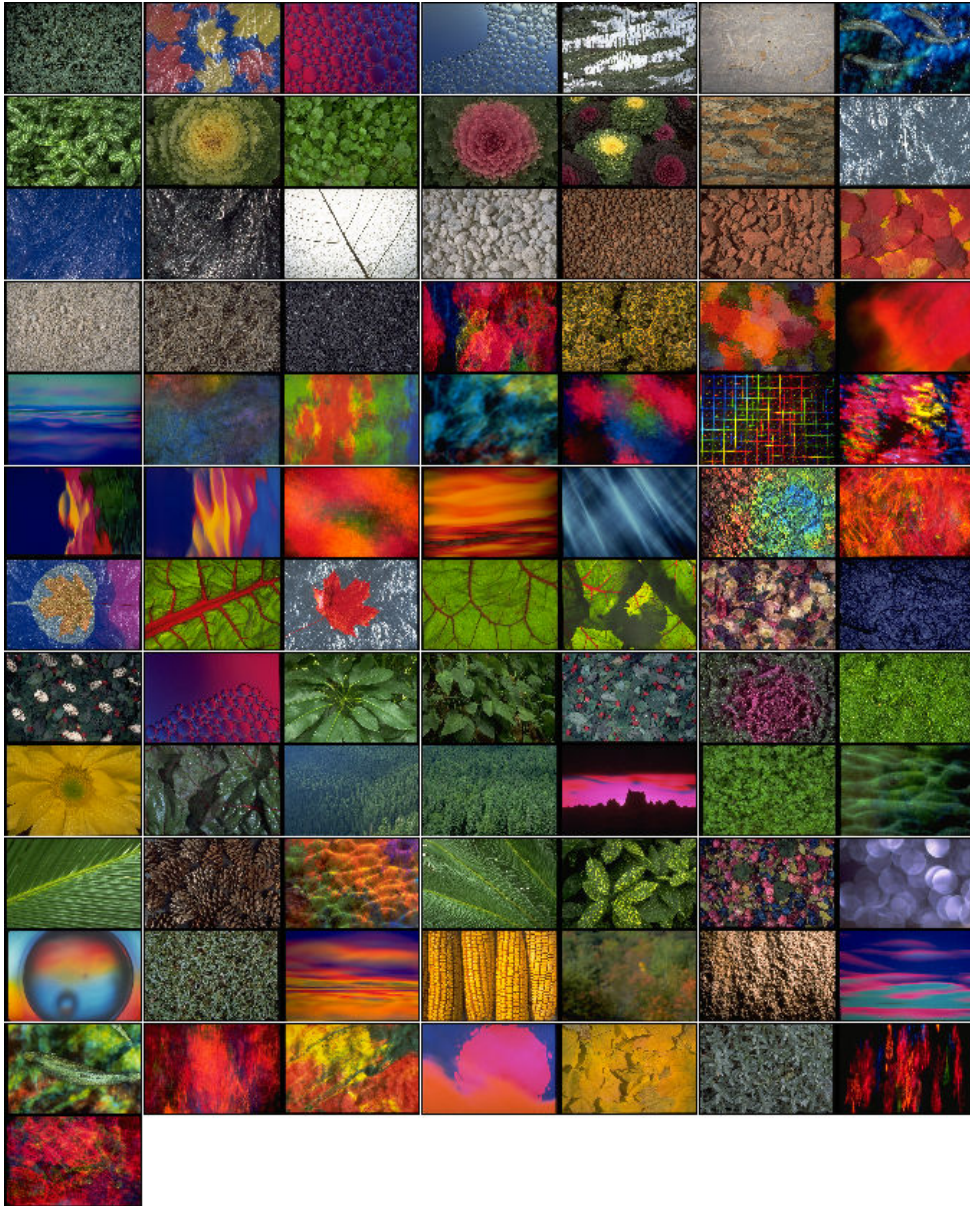


Figura B.14: BD CorelCol.



Figura B.15: BD CorelBark.

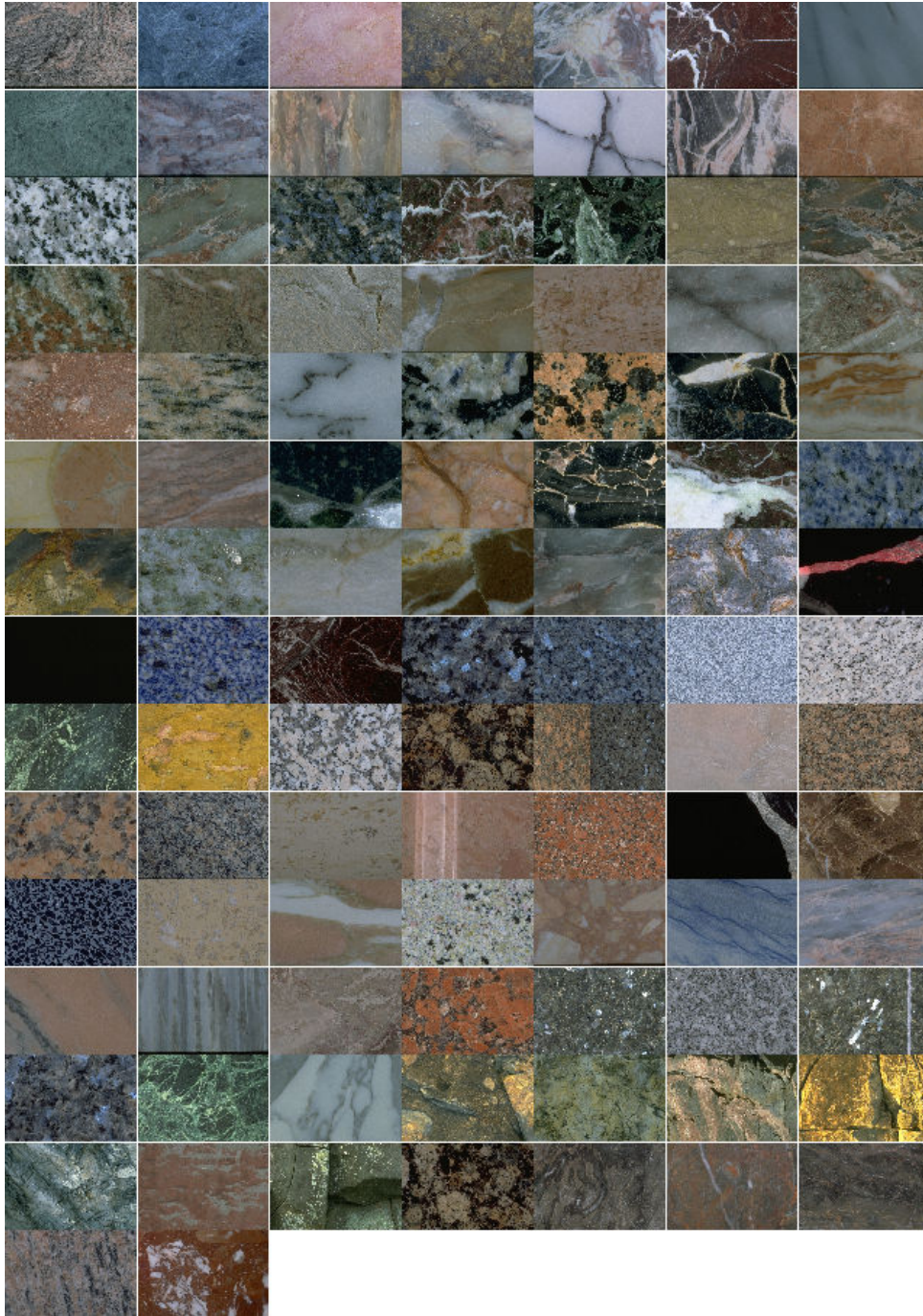


Figura B.16: BD CorelMarb.

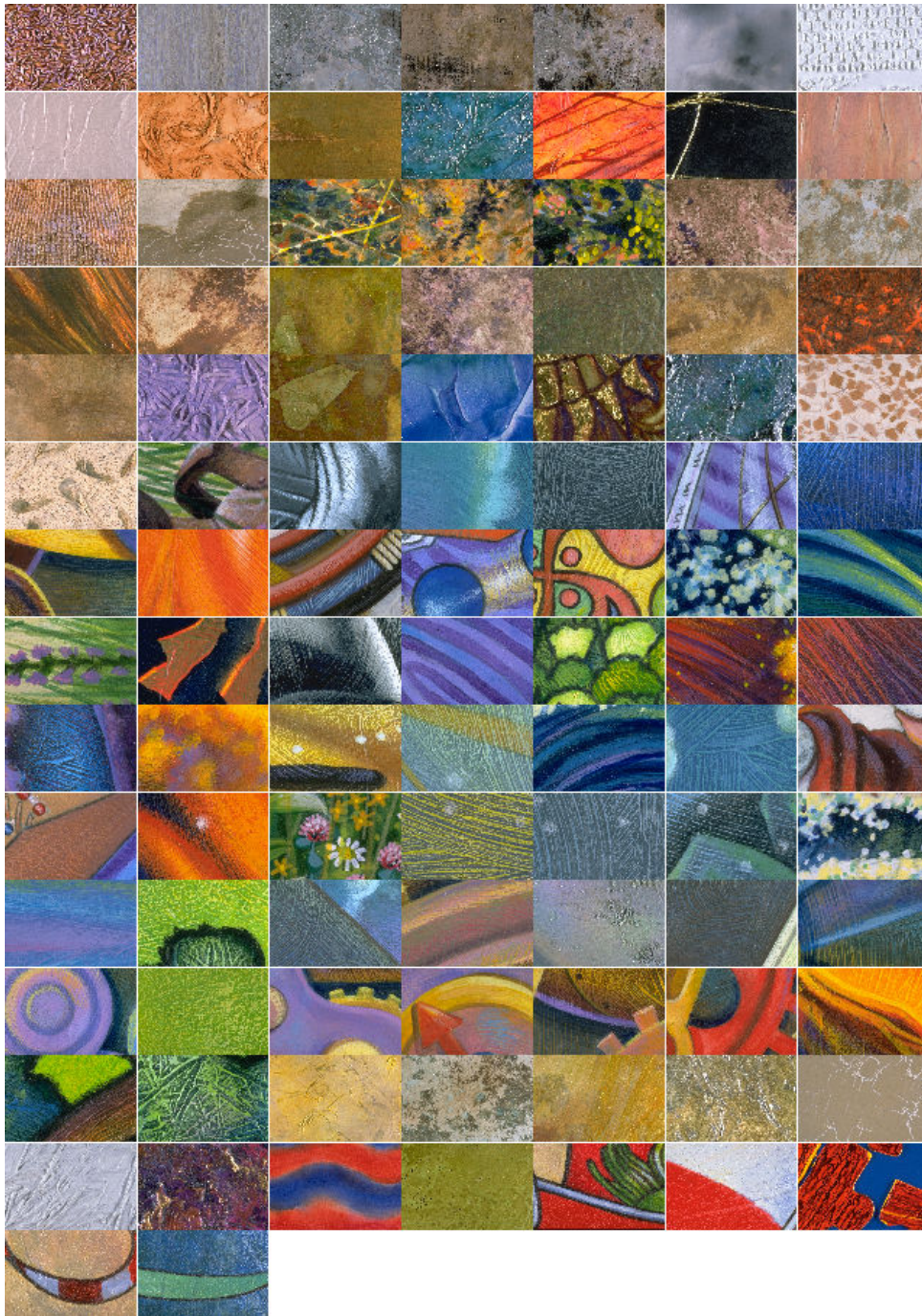


Figura B.17: BD CorelPaint.



Figura B.18: BD CorelShel.

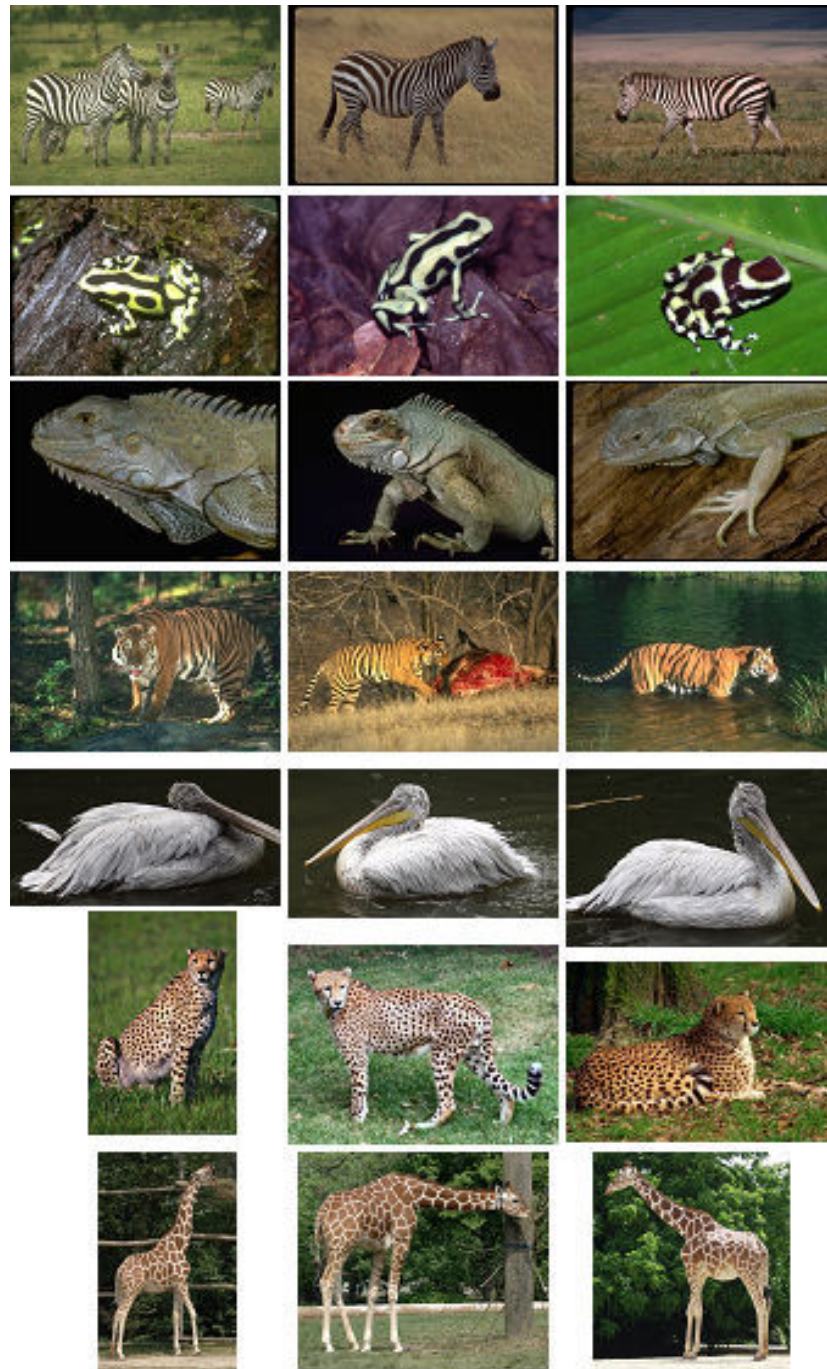


Figura B.19: Imágenes de la BD Animales.

Referencias

- Alvarez, S., Salvatella, A., Vanrell, M., y Otazu, X. (2010a). Lncs: Image analysis and recognition. En (pp. 354–363). A. Campilho and M. Kamel(Eds).
- Alvarez, S., Salvatella, A., Vanrell, M., y Otazu, X. (2010b). Perceptual color texture codebooks for retrieving in highly texture datasets. En *International conference on pattern recognition*.
- Aptoula, E., y Lefevre, S. (2008). On lexicographical ordering in multivariate mathematical morphology. *Pattern Recognition Letters*, 29(2), 109–118.
- Arvis, V., Debain, C., Berducat, M., y Benassi, A. (2004). Generalization of the cooccurrence matrix for colour images: application to colour texture classification. *Image Analysis Stereol.*, 23, 63–72.
- Azencott, R., Graffigne, C., y Labourdette, C. (1990). Edge detection and segmentation of textured plane images. En *Proc. of the special year on image analysis* (pp. 75–87).
- Beck, J., Prazdny, K., y Rosenfeld, A. (1983). Human and machine vision. En (pp. 1–38). J. Beck, B. Hope and A. Rosenfeld (Eds).
- Benavente, R., Vanrell, M., y Baldrich, R. (2004, October). Estimation of fuzzy sets for computational colour categorization. *Color Research and Application*, 29(5), 342-353.
- Benavente, R., Vanrell, M., y Baldrich, R. (2006). A data set for fuzzy colour naming. *Color Research and Application*, 31(1), 48-56.
- Benke, K., Skinner, D., y Woodruff, C. (1988). Convolution operators as a basis for objective correlates of texture perception. *IEEE Trans. SMC*, 18(1), 152–163.
- Bergen, J. (1991). Theories of visual texture perception. En D. Regan (Ed.), *Vision and visual dysfunction* (Vol. 10B, pp. 114–134). New York - MacMillan.
- Berlin, B., y Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley, CA: University of California Press.
- Blostein, D., y Ahuja, N. (1989). Shape from texture: Integrating texture-element extraction and surface estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(12), 1233–1251.
- Brodatz, P. (1966). *Textures: A photographic album for artists and designers*. New York: Dover.
- Burghouts, G., y Geusebroek, J. (2009). Material-specific adaptation of color invariant features. *Pattern Recogn. Lett.*, 30(3), 306–313.
- Dorairaj, R., y Namuduri, K. (2004). Compact combination of MPEG-7 color and texture descrip-

- tors for image retrieval. En *Conference on signals, systems and computers. conference record of the thirty-eighth asilomar* (Vol. 1, p. 387-391).
- Drimbarean, A., y Whelan, P. (2001). Experiments in color texture analysis. *Pattern Recognition Letters*, 22, 1161–1167.
- Evans, R. (1974). The perception of color. En *Industrial color technology* (pp. 43–68). John Wiley & Sons.
- Finlayson, G., Hordley, S., y Hubel, P. (2001). Color by correlation: A simple, unifying framework for color constancy. *IEEE Trans. Pattern Analysis Machine Intelligence*, 23(11), 1209–1221.
- Finlayson, G., y Tian, G. (1999). Investigating colour texture similarity. En *Infotech oulu workshop on texture analysis in machine vision* (p. 127-136).
- Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., y cols. (1995). Query by image and video content: The QBIC system. *Computer*, 28(9), 23–32.
- Forsyth, D. (1990). A novel algorithm for color constancy. *Int. J. Comput. Vision*, 5(1), 5–36.
- Haralick, R. (1979). Statistical and structural approaches to texture. En *Proc. ieee* (Vol. 67, pp. 786–804).
- Haralick, R., Shanmugam, K., y Dinstein, I. (1973). Textural features for image classification. *IEEE Trans. SMC*, 3, 610–621.
- Harvey, L., y Gervais, M. (1978). Visual texture perception and fourier analysis. *Perception and Psychophysics*(24), 534–542.
- Hering, E. (1964). *Outlines of a theory of the light sense*. Harvard University Press.
- Hoang, M., Geusebroek, J., y Smeulders, A. (2005). Color texture measurement and segmentation. *Signal Processing*, 85, 265–275.
- Huang, J., Kumar, S., Mitra, M., Zhu, W., y Zaidi, R. (1999). Spatial color indexing and applications. *International Journal of Computer Vision*, 35(3), 245–268.
- Jain, A., y Farrokhnia, F. (1991). Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, 24(12), 1167–1186.
- Jain, A., y Vailaya, A. (1996). Image retrieval using color and shape. *Pattern Recognition*, 29(8), 1233–1244.
- Julesz, B. (1975). Experiments in the visual perception of texture. *Scientific American*(232), 34–43.
- Julesz, B. (1986). Texton gradients: the texton theory revisited. *Biological Cybernetics*(54), 245–251.
- Julesz, B., y Bergen, J. (1983). Textons, the fundamental elements in preattentive vision and perception of textures. *Bell Systems Technological Journal*, 62(6), 1619–1645.
- Julesz, B., Gilbert, E., Shepp, L., y Frisch, H. (1973). Inability of humans to discriminate between visual textures that agree in second-order statistics. *Perception*, 2(4), 391–405.
- Julesz, B., Gilbert, E., y Victor, J. (1978). Visual discrimination of textures with identical third-order statistics. *Biological Cybernetics*(31), 137–140.
- Kandaswamy, U., y Adjeroh, D. (2005). Color texture retrieval using the collective color texture model. En *Proc. of spiee* (Vol. 5999).

- Kohonen, T. (1997). *Self-organizing maps*. Springer-Verlag.
- Kruskal, J., y Wish, M. (1978). *Multidimensional scaling*. Sage Publications, Inc.
- Laws, K. (1980). *Textured image segmentation* (Inf. Téc. n.º 940). Image Processing Institute. Univ. of Southern California, Los Angeles. (426 citas)
- Lazebnik, S., Schmid, C., y Ponce, J. (2005). A sparse texture representation using local affine regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8), 1265-1278.
- Lee, S., Xin, J., y Wesland, S. (2005). Evaluation of image similarity by histogram intersection. *Color Research and Applications*, 30(4), 265-274.
- Leung, T., y Malik, J. (2001). Representing and recognizing the visual appearance of materials using three-dimensional textons. *Int. Journal Computer Vision*, 43(1), 29-44.
- Liapis, S., y Tziritas, G. (2004). Color and texture image retrieval using chromaticity histograms and wavelet frames. *IEEE Transactions on Multimedia*, 6(5), 676-686.
- Lindeberg, T. (1994). *Scale-space theory in computer vision*. Kluwer Academic Publishers.
- Liu, F., y Picard, R. (1996). Periodicity, directionality, and randomness: Wold features for image modeling and retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(7), 722-733.
- Liu, P., Jia, K., Whang, Z., y Lv, Z. (2007). A new and effective image retrieval method based on combined features. En *Fourth international conference on image and graphics* (pp. 786-790).
- Lohmann, G. (1994). Co-occurrence-based analysis and synthesis of textures. En *Proc. 12th iapr. inter. conf. pattern recognition* (Vol. 1, pp. 449-453).
- Long, F., Zhang, H., y Feng, D. (2003). Multimedia information retrieval and management. En (cap. Fundamentals of Content-Based Image Retrieval). Springer.
- Lovell, R., Uttal, W., Shepherd, T., y Dayanand, S. (1992). A model of visual texture discrimination using multiple weak operators and spatial averaging. *Pattern Recognition*, 25(10), 1157-1170.
- Lowe, D. (1999). Object recognition from local scale-invariant features. En *Iccv '99: Proceedings of the 7th international conference on computer vision-volume 2* (pp. 1150-1157). Washington, DC, USA: IEEE Computer Society.
- Mäenpää, T., y Pietikäinen, M. (2004). Classification with color and texture: jointly or separately? *Pattern Recognition*, 37(8), 1629-1640.
- Mäenpää, T., Pietikäinen, M., y Viertola, J. (2002). Separating color and pattern information for color texture discrimination. En *Icpr '02: Proceedings of the 16th international conference on pattern recognition* (Vol. 1, pp. 668-671).
- Malik, J., y Perona, P. (1990, May). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America*, 7(5), 923-932.
- Maloney, L., y Wandell, B. (1986). Color constancy: A method for recovering surface spectral reflectance. *Journal of the Optical Society of America*, 3(1), 29-33.
- Manjunath, B., y Ma, W. (1996). Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8), 837-842. (Good, basic for texture features)

- Manjunath, B., Ohm, J., Vinod, V., y Yamada, A. (2001, Jun). Color and texture descriptors. *IEEE Trans. Circuits and Systems for Video Technology, Special Issue on MPEG-7*, 11(6), 703–715. Disponible en <http://vision.ece.ucsb.edu/publications/01IEEEManjunath.pdf>
- Manjunath, B., Salembier, P., y Sikora, T. (2003). *Introduction to mpeg-7*. John Wiley & Sons.
- Manjunath, B., Wu, P., Newsam, S., y Shin, H. D. (2000). A texture descriptor for browsing and similarity retrieval. *Journal of Signal Processing: Image Communication*, 16(2), 33–43.
- Martínez, D. P. (1995). Modelos psicofísicos de discriminación de texturas visuales: evolución y aspectos críticos. *Anuario de psicología*(64), 3–20.
- Mojsilovic, A., Kovacevic, J., Hu, J., Safranek, R., y Ganapathy, S. (2000, January). Matching and retrieval based on the vocabulary and grammar of color patterns. *IEEE Transactions on Image Processing*, 9(1), 38–53. (Good)
- Mojsilovic, A., Kovacevic, J., Kall, D., Safranek, R., y Ganapathy, S. (2000, March). The vocabulary and grammar of color patterns. *IEEE Transactions on Image Processing*, 9(3), 417–431.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Ojala, T., Mäenpää, T., Pietikäinen, M., Viertola, J., Kyllönen, J., y Huovinen, S. (2002). Outex - new framework for empirical evaluation of texture analysis algorithms. En *Proc. 16th international conference on pattern recognition* (Vol. 1, pp. 701–706). Disponible en <http://www.outex.oulu.fi/>
- Ojala, T., Pietikäinen, M., y Harwood, D. (1994). Performance evaluation of texture measures with classification based on kullback discrimination of distributions. En *Proc. 12th icpr. inter. conf. pattern recognition* (Vol. 1, pp. 582–585).
- Ojala, T., Pietikäinen, M., y Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *PAMI*, 24(7), 971–987.
- Otazu, X., Vanrell, M., y Párraga, C. (2008). Multiresolution wavelet framework models brightness induction effects. *Vision Research*, 48, 733–751.
- Paschos, G. (2001). Perceptually uniform color spaces for color texture analysis: an empirical evaluation. *IEEE Trans. Image Processing*, 10(6), 932–937.
- Pereira(Ed.), F. (1999). *Mpeg-7 requirements document v.9, iso/iec jtc1/sc29/wg11/document n2859*. Vancouver, Canada.
- Petrou, M., Mirmehdi, M., y Coors, M. (1998). Perceptual smoothing and segmentation of colour textures. En *Eccv '98: Proceedings of the 5th european conference on computer vision* (Vol. 1, pp. 623–639).
- Pietikäinen, M., Rosenfeld, A., y Davis, L. (1983). Experiments with texture classification using averages of local pattern matches. *IEEE Trans. PAMI*, 13(3), 421–426.
- Plataniotis, K., y Venetsanopoulos, A. (2000). Color image processing and applications. En (p. 237-277). New York, NY, USA: Springer-Verlag New York, Inc.
- Poirson, A., y Wandell, B. (1996). Pattern color separable pathways predict sensitivity to simple colored patterns. *Vision Research*, 36(4), 515–526.
- Rao, A., y Lohse, G. (1996). Towards a texture naming system: Identifying relevant dimensions

- of texture. *Vision Research*, 36(11), 1649–1669.
- Renninger, L., y Malik, J. (2004). When is scene recognition just texture recognition? *Vision Research*(44), 2301–2311.
- Richards, W., y Polit, A. (1974). Texture matching. *Kybernetik*(16), 152–162.
- Robertson, A. (1992). Color perception. *Physics Today*(12), 24–29.
- Rubner, Y., Tomasi, C., y Guibas, L. (2000, Nov.). The earth mover’s distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2), 99–121.
- Rui, Y., Huang, T., y Chang, S. (1999). Image retrieval: Current techniques, promising directions, and open issues. *Journal of Visual Communication and Image Representation*, 10(1), 39–62.
- Salvatella, A., y Vanrell, M. (2007). *Blob detection and grouping for texture description and other applications* (Inf. Téc. n.º 110). Computer Vision Center, Bellaterra Autònoma University.
- Serra, J. (1988). *Image analysis and mathematical morphology ii. theoretical advances* (Vol. 2). Academic Press.
- Shafer, S. (1985). Using color to separate reflection components. *Color Research and Applications*, 10(4), 210–218.
- Shapiro, L., y Stockman, G. (2001). Computer vision. En (cap. Texture). Prentice Hall.
- Shi, J., y Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 888-905.
- Shih, J., y Chen, L. (2002). Color image retrieval based on primitives of color moments. En *5th international conference, visual 2002* (pp. 88–94).
- Sivic, J., Russell, B., Efros, A., Zisserman, A., y Freeman, W. (2005). Discovering objects and their localization in images. En *Proc. ieee int. conf. on computer vision (iccv)* (Vol. 1, p. 370-377). Los Alamitos, CA, USA: IEEE Computer Society.
- Sivic, J., y Zisserman, A. (2003). Video google: A text retrieval approach to object matching in videos. En *Nineth ieee international conference on computer vision and pattern recognition (iccv 2003)* (p. 1470-1477).
- Sklansky, J. (1978). Image segmentation and feature extraction. *IEEE SMC*, 8(4), 237–247.
- Smeulders, A., Worring, M., Santini, S., Gupta, A., y Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), 1349-1380.
- Smith, A. (1978). Color gamut transform pairs. *Computer Graphics*(12), 12–19.
- Smith, J. (1998). Image retrieval evaluation. En *Proc. ieee workshop on content - based access of image and video libraries* (p. 112-113). Los Alamitos, CA, USA: IEEE Computer Society.
- Stricker, M., y Orengo, M. (1995). Similarity of color images. En *Proc. spie storage and retrieval for image and video databases iii* (Vol. 2420, pp. 381–392).
- Swain, M., y Ballard, D. (1991). Colour indexing. *Int. Journal of Computer Vision*, 7(1), 11–32.
- Takala, V., Ahonen, T., y Pietikäinen, M. (2005). Block-based methods for image retrieval using local binary patterns. En *In proc. 14th scandinavian conference on image analysis (scia)* (pp. 882–891).

- Tamura, H., Mori, S., y Yamawaki, T. (1978). Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics*, 8, 460–473.
- T. Carron, y Lambert, P. (1994). Color edge detector using jointly hue, saturation and intensity. En *Int. conference on image processing* (Vol. 3, pp. 977–981).
- Tomita, F., y Tsuji, S. (1990). *Computer analysis of visual textures*. Kluwer Academic Publishers.
- Tous, F. (2006). *Computational framework for the white point interpretation based on colour matching*. (PhD thesis, Universitat Autònoma de Barcelona)
- Treisman, A., y Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*(12), 97–136.
- Tsuji, S., y Tomita, F. (1973). A structural analyzer for a class of textures. *Computer Graphics and Image Processing*, 2(1973), 216–231.
- Tuceryan, M., y Jain, A. K. (1998). The handbook of pattern recognition and computer vision (2nd edition). En (pp. 207–248). World Scientific Publishing Co.
- Tuytelaars, T., y Mikolajczyk, K. (2007). Local invariant feature detectors: a survey. *Foundations and Trends in Computer Graphics and Vision*, 3(3), 177–280.
- Unser, M. (1986). Local linear transforms for texture measurements. *Signal Processing*, 11, 61–79.
- Valois, R. D., y Valois, K. D. (1993). A multistage color model. *Vision Research*, 33(8), 1053–1065.
- Vanrell, M., Baldrich, R., Salvatella, A., Benavente, R., y Tous, F. (2004). Induction operators for a computational colour-texture representation. *Computer Vision Image Understanding*, 94(1-3), 92–114.
- Varma, M., y Zisserman, A. (2005). A statistical approach to texture classification from single images. *Int. Journal of Computer Vision*, 62(1-2), 61–81.
- Vilnrotter, F., Nevatia, R., y Price, K. (1986). Structural analysis of natural textures. *IEEE Trans. PAMI*, 8(1), 76–89.
- Voorhees, H., y Poggio, T. (1987). Detecting textons and texture boundaries in natural images. En *1st int. conference on computer vision* (pp. 250–258).
- Voorhees, H., y Poggio, T. (1988). Computing texture boundaries from images. *Nature*, 333, 364–367.
- Wang, G., Zhang, Y., y Fei-Fei, L. (2006). Using dependent regions for object categorization in a generative framework. *IEEE Conference on Computer Vision and Pattern Recognition*, 2, 1597–1604.
- Wang, L., y He, D. (1990). Texture classification using texture spectrum. *Pattern Recognition*, 23(8), 905–910.
- Wyszecki, G., y Stiles, W. (1982). *Color science - concepts and methods, quantitative data and formulae*. John Wiley & Sons.
- Xu, Q., y J. Yang, S. D. (2005). Color texture analysis using the wavelet-based hidden markov model. *Pattern Recognition Letters*, 26(11), 1710–1719.
- Yagi, D., Abe, K., y Nakatani, H. (1992). Segmentation of color aerial photographs using hsv color models. En *Mva'92. iapr workshop on machine vision applications* (pp. 367–370).

-
- Yoe, E., y Essen, D. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neurosciences*, *11*, 219–226.
- Yu, H., Li, M., Zhang, H., y Feng, J. (2003). Color texture moments for content-based image retrieval. En *International conference on image processing* (pp. 24–28).
- Zhong, Y., y Jain, A. K. (2000). Object localization using color, texture and shape. *Pattern Recognition*, *33*(4), 671-684.

Publicaciones

Publicaciones relacionadas con la tesis

- S. Alvarez, A. Salvatella, M. Vanrell y X. Otazu. 3D Texton Spaces for color-texture retrieval . *LNCS: Image Analysis and Recognition*. Part I, pp. 354–363. A. Campilho and M. Kamel(Eds). 2010.
- S. Alvarez, A. Salvatella, M. Vanrell y X. Otazu. Perceptual color texture codebooks for retrieving in highly texture datasets. *International Conference on Pattern Recognition*. Agosto 2010. Estambul, Turquía.
- S. Alvarez, X. Otazu y M. Vanrell. Image Segmentation Based on Inter-Feature Distance Maps. Artificial Intelligence Research and Development. *Frontiers in Artificial Intelligence and Applications*, Vol 131. pp. 75- 82. IOS Press. 2005.
- S. Alvarez, A. Salvatella, M. Vanrell y X. Otazu. Low-dimensional and Comprehensive Color texture description. *Computer Vision and Image Understanding*. En proceso de revisión.
- S. Alvarez, A. Salvatella, M. Vanrell y X. Otazu. 3D Perceptual Space for Texture description. *Int. Journal of Pattern Recognition and Artificial Intelligence*. En proceso de revisión.