

TESIS DEFENDIDA POR  
**Jessica Beltrán Márquez**  
Y APROBADA POR EL SIGUIENTE COMITÉ

---

Dr. Hugo Homero Hidalgo Silva  
Director del Comité

---

Dr. Pedro Gilberto López Mariscal  
Miembro del Comité

---

Dr. Pedro Negrete Regagnon  
Miembro del Comité

---

M. en C. José Adrian Munguía Rivera  
Miembro del Comité

---

Dra. Ana Isabel Martínez García  
Coordinadora del programa de  
posgrado en Ciencias de la  
Computación

---

Dr. David Hilario Covarrubias  
Rosales  
Director de Estudios de Posgrado

20 de noviembre de 2009

**CENTRO DE INVESTIGACIÓN CIENTÍFICA Y DE EDUCACIÓN SUPERIOR  
DE ENSENADA**



---

**PROGRAMA DE POSGRADO EN CIENCIAS  
EN CIENCIAS DE LA COMPUTACIÓN**

---

**RECUPERACIÓN DE IMÁGENES DE TATUAJES BASADA EN CONTENIDO  
VISUAL**

**TESIS**

que para cubrir parcialmente los requisitos necesarios para obtener el grado de  
**MAESTRO EN CIENCIAS**

Presenta:

**JESSICA BELTRÁN MÁRQUEZ**

Ensenada, Baja California, México, noviembre del 2009



**RESUMEN** de la tesis de **Jessica Beltrán Márquez**, presentada como requisito parcial para la obtención del grado de MAESTRO EN CIENCIAS en Ciencias de la Computación. Ensenada, Baja California, octubre de 2009.

## **RECUPERACIÓN DE IMÁGENES DE TATUAJES BASADA EN CONTENIDO VISUAL**

Resumen aprobado por:

---

Dr. Hugo Homero Hidalgo Silva  
Director de Tesis

En la presente tesis se presenta un sistema de reconocimiento de personas mediante imágenes de tatuajes. Diversas agencias policiacas han venido almacenado imágenes que capturan de aquellos individuos que se ven involucrados en eventos criminales. Entre estas imágenes se encuentran tatuajes. Los tatuajes están catalogados dentro de la biometría suave, se utilizan como método complementario al reconocimiento duro: imágenes de rostro, huellas digitales, imágenes de iris, de palmas y ADN, entre otras.

Para obtener una recuperación eficiente de imágenes es necesario representarlas mediante información extraída de las mismas que permita la recuperación en base a su contenido visual. En el desarrollo de esta tesis se explican diferentes técnicas para extraer información de las imágenes, así como de formas para reducir la información obtenida y también mejorar las técnicas de búsqueda para reducir el tiempo de recuperación.

Además se presentan los pasos llevados a cabo para la construcción de una base de datos de imágenes reales obtenidas por la policía. Se presentan los resultados obtenidos mediante los diferentes tipos de experimentos que incluyeron: distintas representaciones de las imágenes, diferentes esquemas de búsqueda, uso de varias funciones distancia para las comparaciones. Se presentan las conclusiones obtenidas a partir de los resultados y se proporcionan recomendaciones sobre el trabajo futuro.

Palabras Clave: Recuperación de imágenes, Descriptores SIFT, Biometría, Tatuajes, SMT.

**ABSTRACT** of the thesis presented by **Jessica Beltrán Márquez** as a partial requirement to obtain the MASTER OF SCIENCE degree in COMPUTER SCIENCE. Ensenada, Baja California, México. October 2009.

## **CONTENT-BASED IMAGE RETRIEVAL SYSTEM IMPLEMENTATION FOR PERSON IDENTIFICATION BY TATOOS**

A tattoo-based person identification system is presented in this thesis. Most police agencies possess a lot of imagery from criminal events. Among those are tattoo images. Tattoos are currently used as a complementary identification source for face, fingerprint, iris, palm and DNA information, among others.

Images must be represented with information extracted by scale and rotation invariant features in order to obtain an efficient retrieval. Several feature extraction techniques are studied and implemented in this thesis. Information reduction methods are also necessary because of the large amount of features obtained by the extractors. Search times must also be reduced by using representation models like the “Bag of Words”.

A real tattoo image data base is constructed from law enforcement images, and several experiments are presented for the data representation, recovery and system validation. The results are reported in terms of precision-recall and Cumulative Matching Characteristic figures.

**Keywords:** Content-based image retrieval, SIFT, Feature extraction, Tattoo, SMT

## **Dedicatorias**

Dedico e

## **Agradecimientos**

## CONTENIDO

	<b>Página</b>
<b>Resumen español</b> .....	<b>i</b>
<b>Resumen inglés</b> .....	<b>ii</b>
<b>Dedicatorias</b> .....	<b>iii</b>
<b>Agradecimientos</b> .....	<b>iv</b>
<b>Contenido</b> .....	<b>v</b>
<b>Lista de figuras</b> .....	<b>viii</b>
<b>Lista de tablas</b> .....	<b>xii</b>
<b>Capítulo I Introducción</b> .....	<b>1</b>
<b>I.1 Antecedentes</b> .....	1
<b>I.2 Trabajo previo</b> .....	7
<b>I.3 Definición del problema y objetivos</b> .....	8
<i>I.3.1 Objetivo general</i> .....	8
<i>I.3.2 Objetivos específicos</i> .....	9
<b>I.4 Limitaciones y suposiciones</b> .....	9
<b>I.5 Metodología</b> .....	10
<b>I.6 Importancia de la investigación</b> .....	11
<b>I.7 Organización de la tesis</b> .....	12
<b>Capítulo II Marco Teórico</b> .....	<b>13</b>
<b>II.1 Introducción a la recuperación de imágenes basada en contenido visual</b> .....	13
<i>II.1.1 Primera generación de sistemas de recuperación de información visual</i> .....	14
<i>II.1.2 Segunda generación de sistemas de recuperación de información visual</i> .....	16
<i>II.1.3 Tipos de consulta en sistemas de recuperación de información visual</i> .....	17
<b>II.2 Representación de imágenes para su recuperación por contenido visual</b> .....	19
<i>II.2.1 Extracción de características</i> .....	19
<i>II.2.2 Técnicas de extracción de características</i> .....	20
<i>II.2.3 Extracción de características con SIFT</i> .....	22
<i>II.2.3.1 Algoritmo SIFT</i> .....	24

## CONTENIDO (continuación)

	<b>Página</b>
II.2.3.1.1 Detección extrema en el espacio de escala .....	25
II.2.3.1.2 Localización precisa de puntos de interés. ....	28
II.2.3.1.3 Asignación de orientaciones .....	29
II.2.3.1.4 Descriptor de puntos de interés .....	31
II.2.3.1.5 Comparación entre dos imágenes con SIFT.....	32
II.2.4 Extracción de características con Wavelets .....	34
<b>II.3 Reducción de información .....</b>	<b>36</b>
II.3.1 Vocabularios de palabras (Bolsa de palabras) .....	37
II.3.2 Permutaciones.....	42
<b>II.4 Evaluación de Sistemas de Recuperación.....</b>	<b>46</b>
II.4.1 Gráfica precisión-exhaustividad.....	48
II.4.2 Curva “Cumulative Matching Characteristic” (CMC).....	49
<b>Capítulo III Construcción de la base de datos .....</b>	<b>52</b>
<b>III.1 Obtención de imágenes de tatuajes para base de datos .....</b>	<b>52</b>
<b>III.2 Construcción base de datos principal.....</b>	<b>53</b>
III.2.1 Consulta con imágenes originales .....	57
<b>III.3 Construcción de base de datos y consultas con imágenes transformadas.....</b>	<b>58</b>
<b>III.4 Construcción de base de datos para consultas por imágenes similares .....</b>	<b>60</b>
<b>Capítulo IV Experimentos y resultados.....</b>	<b>63</b>
<b>IV.1 Introducción de experimentos realizados.....</b>	<b>63</b>
<b>IV.2 Experimentos utilizando SIFT .....</b>	<b>64</b>
IV.2.1 Utilizando medida de similitud de número de igualaciones por par de imágenes. ....	64
IV.2.1.1 Imágenes en escala de grises .....	65
IV.2.1.2 Imágenes con ecualización adaptiva .....	66
IV.2.1.3 Imágenes con umbralización adaptiva .....	68

## CONTENIDO (continuación)

	<b>Página</b>
<i>IV.2.1.4 Comparación de resultados con SIFT utilizando medida de similitud de número de igualaciones por par de imágenes .....</i>	<i>69</i>
<i>IV.2.2 Utilizando medida de similitud de número de igualaciones en base de datos completa.....</i>	<i>70</i>
<i>IV.2.2.1 Librería ANN (Approximate Nearest Neighbor).....</i>	<i>72</i>
<i>IV.2.2.2 Librería PNN, Librería utilizando permutaciones.....</i>	<i>74</i>
<i>IV.2.3 Uso de bolsa de palabras.....</i>	<i>76</i>
<i>IV.2.3.1 20 000 puntos aleatorios.....</i>	<i>76</i>
<i>IV.2.3.2 20000 puntos de árbol jerárquico.....</i>	<i>78</i>
<i>IV.2.3.3 40000 puntos centroides de árbol jerárquico .....</i>	<i>79</i>
<i>IV.2.3.4 Uso de distintas distancias.....</i>	<i>81</i>
<i>IV.2.3.4.1 40000 representantes con distancia L0.5 .....</i>	<i>81</i>
<i>IV.2.3.4.2 40000 puntos con distancia L0.75 .....</i>	<i>83</i>
<i>IV.2.3.4.3 40000 puntos con distancia L1 .....</i>	<i>84</i>
<i>IV.2.3.5 Modificación del método de bolsas de palabras.....</i>	<i>85</i>
<i>IV.2.3.6 Experimentos con base de datos de imágenes transformadas .....</i>	<i>89</i>
<b>IV.3 Experimentos utilizando imgseek (características wavelets).....</b>	<b>94</b>
<b>Capítulo V Conclusiones .....</b>	<b>99</b>
<b>V.1 Trabajo futuro.....</b>	<b>100</b>
<b>Bibliografía.....</b>	<b>102</b>
<b>Anexos .....</b>	<b>107</b>

## LISTA DE FIGURAS

Figura 1. Clasificación de tipos de biometría. ....	2
Figura 2. Tatuajes de las bandas: Hermandad Aria, Nuestra Familia. Mafia Mexicana, Mara Salvatrucha. Familia Guerrilla Negra.....	3
Figura 3. Clases de tatuajes definidas en ANSI/NIST-ITL 1-200.....	6
Figura 4. Algunos tatuajes con varias clases o con dificultad para designar clase.....	6
Figura 5. Esbozo de una primer generación de sistemas de recuperación de información visual para imágenes 2D. Fuente, Del Bimbo. (del Bimbo, 2001). ....	14
Figura 6. Esbozo de una nueva generación de sistemas de recuperación de información visual para imágenes 2D. Fuente, Del Bimbo. (del Bimbo, 2001).....	16
Figura 7. Obtención de puntos de interés con SIFT .....	22
Figura 8. Igualación entre puntos de interés de a) 2 imágenes correspondientes, b) 2 imágenes no correspondientes. ....	23
Figura 9. Imagen convolucionada con kernel Gaussiano con diferentes sigmas para producir imágenes de espacio escala para cada octava.....	26
Figura 10. Obtención de diferencias de Gaussianas. Fuente (Lowe, 2004).....	27
Figura 11. Diferencias de Gaussianas de imágenes de la figura 9.....	27
Figura 12. La detección de puntos máximos y mínimos en las diferencias de Gaussianas se lleva a cabo comparando un pixel (marcado con X) con sus 26 vecinos en regiones de 3x3 en la escala actual y en las escalas adyacentes (marcados con círculos). Fuente (Lowe, 2004). ....	28
Figura 13. Obtención de orientación de puntos de interés.....	31
Figura 14. Obtención del descriptor del punto de interés. ....	32

Figura 15. Top 20 de imágenes recuperadas usando SIFT y realizando comparaciones de la imagen de consulta con cada una de las imágenes de la base de datos. ....	33
Figura 16. Extracción de características de imagen en imgseek.....	35
Figura 17. Agrupamiento jerárquico a partir de algunas imágenes de entrenamiento. ....	38
Figura 18. Cuantización de una imagen con los elementos del agrupamiento. arriba) Obtención del centroide más cercano para un descriptor. abajo) Ejemplo de representación con 5 centroides, Se obtiene el centroide más cercano de cada descriptor, posteriormente el vector representante es la frecuencia con que cada centroide fue vecino más cercano de los descriptores de la imagen representada. ....	39
Figura 19. Sistema de recuperación de imágenes utilizando SIFT y bolsa de palabras. ....	41
Figura 20. Obtención de permutaciones en puntos de una base de datos. ....	43
Figura 21. Consultas de más cercanos mediante permutaciones .....	44
Figura 22. Compromiso entre precisión y exhaustividad. ....	49
Figura 23. Arriba) Ejemplo de resultados de recuperación. Abajo) Ejemplo de la curva CMC.....	50
Figura 24.a) Imagen original, b) imagen ecualizada, c) imagen umbralizada d) imagen intervenida manualmente.....	56
Figura 25. a) Original b) e) o) Modificación de canal de color, c) d) difuminado f) iluminación g) ruido h) i) j) k) Cambio de perspectiva, l) m) n) p) q) Rotación r) s) t) u) v) Escalamiento.....	59
Figura 26, a) Imágenes originales, b) Imagen orientada en base a su orientación predominante, c) Imagen orientada 180 grados en base a su orientación predominante.....	61

Figura 27, a) Imágenes umbralizadas originales, b) Imagen umbralizada orientada en base a su orientación predominante, c) Imagen umbralizada orientada 180 grados en base a su orientación predominante. ....	61
Figura 28. Imágenes elegidas para consultas por imágenes similares. ....	62
Figura 29. Esquema de búsqueda con medida de similitud de número de igualaciones exitosas por par de imágenes ....	64
Figura 30. Curva CMC con base de datos de imágenes en escala de grises.....	66
Figura 31. Curva CMC con base de datos de imágenes con ecualización adaptiva .....	67
Figura 32. Curva CMC con base de datos de imágenes con umbralización adaptiva .....	69
Figura 33. Curva CMC con base de datos de imágenes en escala de grises, con ecualización y umbralización adaptiva .....	69
Figura 34. Esquema de recuperación con número de igualaciones exitosas en toda la base de datos. ....	71
Figura 35. Curva CMC para experimentos utilizando librería ANN.....	73
Figura 36. Curva CMC para experimentos utilizando librería PNN .....	75
Figura 37. Curva CMC con 20000 puntos aleatorios. ....	77
Figura 38. Curva CMC con 20000 centroides de árbol jerárquico. ....	78
Figura 39. Curva CMC con 40000 centroides de árbol jerárquico .....	80
Figura 40. Curva CMC con 40000 centroides de árbol jerárquico y distancia L0.5 .....	82
Figura 41. Curva CMC con 40000 centroides de árbol jerárquico y distancia L0.75 .....	83
Figura 42. Curva CMC con 40000 centroides de árbol jerárquico y distancia L1 .....	84

Figura 43. Representación de imágenes mediante modificación de esquema de bolsa de palabras. ....	86
Figura 44. <i>Curva CMC</i> usando modificación de esquema de bolsa de palabras <i>con</i> $\mathbf{W} = 1,1,1,1,1,1,1,1,1,1$ .....	87
Figura 45. <i>Curva CMC</i> usando modificación de esquema de bolsa de palabras <i>con</i> $\mathbf{W} = \{1,12,14,18,116,132,164,1128,1256,1512\}$ .....	88
Figura 46. Curva Precisión-Exhaustividad con consultas de base de datos de imágenes transformadas.....	90
Figura 47. Gráfica de precisión re cada transformación en el rango 20. ....	91
Figura 48. Recuperación de imágenes similares usando SIFT para consulta dada. ....	92
Figura 49. Recuperación de imágenes similares usando SIFT para consulta dada. ....	92
Figura 50. Recuperación de imágenes similares usando SIFT para consulta dada. ....	93
Figura 51. Recuperación de imágenes similares usando SIFT para consulta dada. ....	93
Figura 52. Diez primeras imágenes recuperadas por imgseek sin realizar orientaciones para la consulta dada.....	94
Figura 53. Diez primeras imágenes recuperadas por imgseek realizando orientaciones para la consulta dada.....	95
Figura 54. Diez primeras imágenes recuperadas por imgseek sin realizar orientaciones para la consulta dada.....	95
Figura 55. Diez primeras imágenes recuperadas por imgseek realizando orientaciones para la consulta dada.....	96
Figura 56. Diez primeras imágenes recuperadas por imgseek sin realizar orientaciones para la consulta dada.....	96

Figura 57. Diez primeras imágenes recuperadas por imgseek realizando orientaciones para la consulta dada.....	97
Figura 58. Diez primeras imágenes recuperadas por imgseek sin realizar orientaciones para la consulta dada.....	97
Figura 59. Diez primeras imágenes recuperadas por imgseek realizando orientaciones para la consulta dada.....	98

## LISTA DE TABLAS

Tabla 1. Comparación entre tatuajes recreacionales y tatuajes reales. ....	52
Tabla 2. Ejemplos de tatuajes, marcas y cicatrices.....	53
Tabla 3. Imágenes recortadas a partir de imágenes originales .....	54
Tabla 4. Clases definidas por el estándar ANSI/NIST ITL 1-2000.....	55
Tabla 5. Cantidad de imágenes con contenidos similares en la base de datos. ....	62
Tabla 6. Cantidad de imágenes y precisión acumulada obtenidas en las posiciones 1-20 en las consultas realizadas con las imágenes en escala de grises. ....	65
Tabla 7. Cantidad de imágenes y precisión acumulada obtenidas en las posiciones 1-20 en las consultas realizadas con las imágenes con ecualización adaptiva.....	67
Tabla 8. Cantidad de imágenes y precisión acumulada obtenidas en las posiciones 1-20 en las consultas realizadas con las imágenes con umbralización adaptiva.....	68
Tabla 9. Cantidad de imágenes y precisión acumulada obtenidas en las posiciones 1-20 utilizando librería ANN .....	73
Tabla 10. Cantidad de imágenes y precisión acumulada obtenidas en las posiciones 1-20 utilizando librería PNN.....	74
Tabla 11. Cantidad de imágenes y Precisión acumulada en las posiciones 1-20 con 20000 puntos aleatorios usando bolsa de palabras. ....	76
Tabla 12. Cantidad de imágenes y Precisión acumulada en las posiciones 1-20 con 20000 centroides de árbol jerárquico usando bolsa de palabras.....	78
Tabla 13. Cantidad de imágenes y Precisión acumulada en las posiciones 1-20 con 40000 centroides de árbol jerárquico usando bolsa de palabras y distancia L2. ....	79

Tabla 14. Comparación de tiempos de recuperación usando diferentes esquemas de recuperación.....	80
Tabla 15. Cantidad de imágenes y Precisión acumulada en las posiciones 1-20 con 40000 centroides de árbol jerárquico usando bolsa de palabras y distancia L0.5 .....	82
Tabla 16. Cantidad de imágenes y Precisión acumulada en las posiciones 1-20 con 40000 centroides de árbol jerárquico usando bolsa de palabras y distancia L0.75 .....	83
Tabla 17. Cantidad de imágenes y Precisión acumulada en las posiciones 1-20 con 40000 centroides de árbol jerárquico usando bolsa de palabras y distancia L1 .....	84
Tabla 18. Cantidad de imágenes y Precisión acumulada en las posiciones 1-20 con 40000 usando modificación de esquema de bolsa de palabras con $\mathbf{W} = 1,1,1,1,1,1,1,1,1$ .....	87
Tabla 19. Cantidad de imágenes y Precisión acumulada en las posiciones 1-20 con 40000 usando modificación de esquema de bolsa de palabras con $\mathbf{W} = 1,1,1,1,1,1,1,1,1$ }.....	88
Tabla 20. Exhaustividad y Precisión de consultas con base de datos de imágenes transformadas.....	89

# Capítulo I

---

## Introducción

---

### I.1 Antecedentes

Los tatuajes son marcas corporales artificiales que pueden servir como una característica distintiva, ya sea de una persona o de un grupo. Consisten en una modificación del color de la piel creando un dibujo, figura o texto realizados con tinta o algún otro pigmento bajo la epidermis de un humano o animal.

En el trabajo de Jain *et al.* (2004) observan que los tatuajes pertenecen a ciertos rasgos complementarios que pueden utilizarse para identificar personas y se incluyen dentro de un concepto llamado *biometría suave*.

El reconocimiento biométrico o, simplemente biometría, se refiere al reconocimiento automático de individuos basado en sus atributos físicos y/o en sus características de comportamiento. Usando la biometría es posible confirmar o establecer la identidad de un individuo basado en “quien es” en vez de “que posee” (Jain *et al.*, 2004b; Wayman *et al.*, 2004). La biometría se divide generalmente en las clases presentadas en la Figura 1.

Como se puede apreciar en la Figura 1, en la biometría dura se incluyen características fisiológicas y/o de comportamiento en los humanos las cuales satisfacen los siguientes requerimientos:

- Universalidad: cada persona debería tener dicha característica.
- Distintividad: La extracción debe ser diferente en cada persona.

- Permanencia: la característica deberá ser lo suficientemente invariante por un periodo de tiempo.
- Colectabilidad: la característica puede ser medida cuantitativamente.

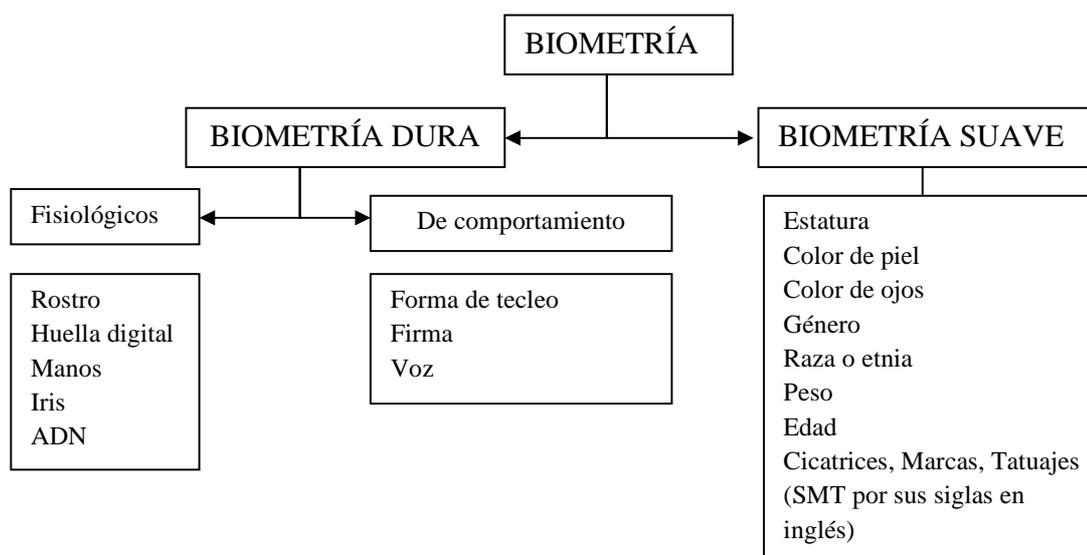


Figura 1. Clasificación de tipos de biometría.

Por otro lado, aunque las características que pertenecen a la biometría suave no cumplen rigurosamente con los requerimientos mostrados, pueden ayudar a complementar la decisión a tomar por un sistema biométrico, de hecho en Jain *et al.* (2004) se muestra que el usar información de biometría suave adicional con un sistema biométrico basado en huellas digitales incrementa el desempeño de reconocimiento en un aproximado de 6%. Además, los rasgos de biometría suave ayudan a reducir drásticamente el tiempo requerido para la identificación.

Es posible aprovechar el hecho de que los tatuajes son cada vez más populares, incluso se han utilizado para reconocer cadáveres después de algún desastre natural, ya que sus pigmentos se encuentran a una profundidad en la piel que los hace permanecer aún después de quemaduras (Jain *et al.*, 2007). Además brindan información sobre asociación a bandas,

clubes, creencias religiosas, estancia en cárceles o crímenes específicos cometidos (Valentine y Schober, 2000).ver Figura 2.



Figura 2. Tatuajes de las bandas: a) Hermandad Aria, b) Nuestra Familia. c) Mafia Mexicana, d) Mara Salvatrucha. e) Familia Guerrilla Negra.

Con la intención de averiguar el interés de las instituciones en México para utilizar los tatuajes en aplicaciones forenses, en mayo del 2009 se realizó una entrevista con el ingeniero Gilberto Gradías Enríquez, Director General de la Dirección General de Sistemas de Información y Política Criminal en la Procuraduría General de Justicia del Estado de Sonora. Este departamento tiene una visión multi-biométrica en el que se tiene como meta utilizar toda la información disponible para la identificación de personas. Actualmente, se utilizan las tecnologías biométricas de reconocimiento facial, identificación de voz, reconocimiento de huellas digitales, además se almacena información personal y también un historial de casos criminales, esto con la intención de poder establecer conexiones entre personas según las similitudes en dichos casos. Las tecnologías que se desean introducir próximamente incluyen el reconocimiento de la palma de la mano, el reconocimiento del iris y el uso de tatuajes para identificar sujetos.

Hasta la fecha, no hay un sistema comercial robusto de identificación de personas mediante tatuajes de forma automática, sino que se encuentra en vías de desarrollo (White, 2008). En el caso del gobierno del estado de Sonora, la idea de utilizar los tatuajes para poder realizar búsquedas, surgió aproximadamente hace 5 años y a partir de ese momento se empezaron a coleccionar tatuajes como parte del procedimiento establecido por el Registro Estatal de Identificación de Sonora (RESIDES). En este sistema se han llegado a acumular en la actualidad un aproximado de 36000 imágenes de tatuajes.

Según el Ingeniero Gilberto Gradías Enriquez, el análisis de las imágenes de tatuajes colabora a relacionar dibujos o imágenes con diferentes tendencias delictivas, coopera a ligar personas con el mismo tipo de tatuaje indicando pertenencia a alguna pandilla o grupo (pertenencia al ejército, marina,...), así como también para vincular formas de tatuajes con zonas específicas (modas que se vinculen al fenómeno delictivo).

Para el ingeniero Gradías, un sistema de recuperación de imágenes de tatuajes debería al menos de contar con las siguientes características:

- Permitir la búsqueda de tatuajes alterados geoméricamente.
- Permitir introducir bosquejos o tatuajes similares para recuperar aquellos tatuajes que se parecen a la consulta.
- Permitir identificar si existen tatuajes encimados.
- Permitir navegar de forma manual sobre los tatuajes para tener la opción de realizar reconocimiento visual.
- Contener un mapa geográfico en donde se agrupen o asocien las formas de tatuajes relacionados a zonas específicas.
- Permitir relacionar el lugar de origen de los tatuajes de la base de datos (cárcel en donde fue realizado, artistas por quien fue realizado, etc.).
- Cumpla con los estándares internacionales.

Es necesario establecer la estandarización lo más antes posible ya que mientras no se realiza, se continúan capturando imágenes las cuales posiblemente no cumplan con las características necesarias para el sistema en donde se realicen las búsquedas.

Desde abril 1997 el Instituto Nacional de Estándares y Tecnología de Estados Unidos (NIST por sus siglas en inglés) incluyó a los tatuajes dentro del documento “*Data Format for the Interchange of Fingerprint, Facial, y Scar, Mark, y Tattoo (SMT) Information*” (ITL, 2009) Este documento presenta los estándares para el intercambio de información

biométrica aprobados por el Instituto Americano Nacional de Estándares (ANSI por sus siglas en inglés).

Este estándar, que actualmente se encuentra en su versión ANSI-NIST ITL 1-2007, menciona que para la captura de imágenes de tatuajes se incluyen cámaras digitales o cámaras de video que produzcan imágenes provenientes de la cabeza o cuerpo de algún sujeto. En el caso de requerirse digitalizar imágenes de fotografías, dibujos o bocetos se utiliza un escáner. Las representaciones de dichas imágenes son ya sea a color o en escala de grises dependiendo del equipo utilizado. Se debe especificar la ubicación corporal del tatuaje así como, de manera opcional, su tamaño.

Las imágenes se pueden almacenar tanto en su versión original como comprimidas, junto con una anotación textual que describa la información de cada imagen (metadatos). Este texto descriptivo debe consistir de las clases y/o subclases definidas por el estándar ANSI-NIST ITL 1-2007 así como del color del tatuaje. Se presenta en la Figura 3 las clases que definen la información de las imágenes de tatuajes y las subclases de la clase ANIMAL.

Son precisamente los metadatos los que se utilizan para realizar consultas de imágenes en una base de datos. Un experto se encarga de etiquetar las imágenes para almacenarlas en la base de datos y al momento de realizar una búsqueda se introduce texto referente a las clases o subclases mencionadas en el estándar, donde las imágenes recuperadas son aquellas que coinciden con el campo textual introducido.

Sin embargo, en la asignación de los metadatos a las imágenes de tatuajes se puede presentar que sea difícil describir a qué clase o subclase pertenece, pudiendo ocurrir que no pertenezca a alguna clase establecida por el estándar. Ver Figura 4.

ANSI/NIST ITL 1-2007 Clases de Tatuajes		ANSI/NIST ITL 1-2007 Subclases de Tatuajes de Animales	
Descripción de clase	Código de clase	Subclase	Código de subclase
Formas humanas y Características	HUMAN	Gatos y Cabezas de gatos	CAT
Animales y Características de animal	ANIMAL	Perros y Cabezas de perros	DOG
Plantas	PLANT	Otros animales domésticos	DOMESTIC
Banderas	FLAG	Animales feroces (Leones, Tigres, etc.)	VICIOUS
Objetos	OBJECT	Caballos (Burros, Mulas, etc.)	HORSE
Abstracciones	ABSTRACT	Otros animales feroces	WILD
Insignias y Símbolos	SYMBOL	Serpientes	SNAKE
Otras imágenes	OTHER	Dragones	DRAGON
		Aves (Cardenales, Halcones, etc.)	BIRD
		Arañas, bichos e insectos	INSECT
		Animales abstractos	ABSTRACT

Figura 3. Clases de tatuajes definidas en ANSI/NIST-ITL 1-200.



Figura 4. Algunos tatuajes con varias clases o con dificultad para designar clase.

Existen pocos trabajos (Jain *et al.*, 2007 ; Lee *et al.*, 2008; Acton y Rossi, 2008) que proponen identificación de tatuajes usando contenido de las imágenes y no solo los metadatos. Esta forma de identificación responde a un área conocida como “Recuperación

de información visual basada en contenido” la cual se muestra con mayor detalle en el siguiente capítulo.

## **I.2 Trabajo previo**

En el trabajo de Jain *et al.* (2007) se realiza un preprocesamiento de la imagen para separar los tatuajes del fondo, el cual corresponde a la piel. Para esto utilizan un operador Sobel y operaciones morfológicas. Después proceden a extraer información de las imágenes tal como color, forma y textura.

Para color, se utiliza el espacio RGB, se extrae un histograma y un correlograma de color. Para forma, se usa representación 2D utilizando momentos invariantes. Para textura se utiliza el vector coherente de dirección de bordes. Una vez obtenida la información se realizan comparaciones entre las imágenes de consulta y las de la base de datos, mediante una función distancia basada en el método de intersección de histogramas.

Por otro lado Acton y Rossi(2008), utilizan contornos activos para extraer segmentos de tatuajes. Después procesan dichos segmentos para obtener información de forma mediante descriptores de forma con Fourier. Para la textura, usan descomposición en wavelets de Haar. Para color, utilizan un histograma con un modelo de color basado en tinta-saturación-valor (HSV). La desventaja de este método es que los contornos activos fallan al recuperar imágenes no conectadas, además los experimentos que muestran son con imágenes artificiales de alta calidad y no resultan tan efectivos como la propuesta de Lee *et al.* (2008);

En el trabajo de Lee *et al.* (2008) se utiliza el algoritmo SIFT ( Scale Invariant Feature Transform) (Lowe, 2004) el cual se describe en el siguiente capítulo. La comparación se lleva a cabo con puntos de interés entre dos imágenes. Se busca la cantidad de puntos de

interés que dos imágenes tienen semejantes y a partir de la cantidad de puntos de interés semejantes se realizan las comparaciones entre una consulta y la base de datos.

En esta última publicación, se presentan experimentos con imágenes obtenidas de la web, que corresponden a tatuajes de alta calidad y ricas en color, así como experimentos con tatuajes proporcionados por la policía del estado de Michigan en Estados Unidos. Los resultados presentados en el caso de la base de datos de imágenes de la web presentan una alta efectividad, mientras que los resultados obtenidos con las imágenes de la policía bajan en efectividad. El tiempo que se reporta es de 0.3 segundos por cada comparación entre par de imágenes, lo cual resulta un tiempo grande si se realizan muchas comparaciones.

Como se menciona, solo han sido publicadas estas investigaciones para la identificación de tatuajes. Exceptuando la última publicación, en estas propuestas se ha trabajado principalmente con imágenes de tatuajes de alta calidad que no corresponden con las imágenes reales. Además, en dichas propuestas solo se ataca el problema de consultar tatuajes los cuales estén alterados geométricamente en la base de datos.

### **I.3 Definición del problema y objetivos**

Aún es necesario desarrollar un sistema de identificación de tatuajes más eficiente, en donde se obtengan buenos resultados logrando un compromiso entre tiempo de búsqueda y exactitud, además que no se requiera almacenar demasiada información para representar las imágenes. También falta por tratar el problema de la búsqueda de tatuajes que se parecen a los de la consulta, así como los demás requerimientos que recomienda el ingeniero Gradías.

#### *I.3.1 Objetivo general*

Se busca construir un sistema de recuperación de imágenes de tatuajes el cual se base en el contenido visual y no solamente en las anotaciones textuales. Para realizar esto, será

necesario representar la imagen de tal forma que el algoritmo de búsqueda permita hacer comparaciones entre las representaciones de las imágenes y recuperar aquellas que corresponden a una consulta dada. Además se pretende evaluar el desempeño y compararlo con los sistemas ya propuestos.

### *1.3.2 Objetivos específicos*

- Entender diferentes métodos de extracción de información de las imágenes.
- Encontrar y entender las formas que existen para analizar la eficiencia de los sistemas de recuperación de información visual basada en contenido.
- Construir una base de datos que permita llevar a cabo las pruebas necesarias.
- Investigar y proponer una forma de reducir la información almacenada de las imágenes e incrementar la velocidad de recuperación.
- Realizar pruebas con imágenes dibujadas en algún lienzo digital para observar los primeros N (“top-N”) tatuajes recuperados. Esto utilizando diferentes representaciones de las imágenes.
- Mejorar la representación en cuanto espacio y acelerar el tiempo de búsqueda sin perder efectividad en los resultados.
- Evaluar la representación propuesta en cuanto a velocidad de respuesta en búsqueda, espacio requerido y recuperación de imágenes por contenido.

## **I.4 Limitaciones y suposiciones**

Los experimentos se enfocan en su mayoría a buscar imágenes alteradas geométricamente sin atacar el resto de las recomendaciones mencionadas por el ingeniero Gradías. Aunque en un apartado se trabaja con consultas de imágenes parecidas mediante el uso de una aplicación ya existente.

En la construcción de la base de datos para los experimentos, se manejan imágenes en las cuales han sido recortados los tatuajes manualmente, de esta forma se logran separar de fondos que introduzcan ruido, así como otros elementos como ropa, partes del cuerpo, etc.

Para poder realizar una comparación, lo ideal es que los algoritmos trabajen bajo la misma base de datos. Sin embargo, debido a que no se cuenta ni con el algoritmo ni con la base de datos que utilizaron (Jain *et al.*, 2007), la comparación no se puede llevar a cabo de manera estricta, pero se utiliza el mismo tipo de gráficas, además que se recrea el algoritmo a partir de lo mencionado en su artículo.

En los experimentos del presente trabajo no se realiza realimentación de relevancia, un concepto presentado en una sección posterior. Tampoco se realizan experimentos con anotaciones textuales como metadatos, sino únicamente en base al contenido de las imágenes.

## **I.5 Metodología**

A continuación se hace mención sobre las fases que se llevaron a cabo en la realización del presente trabajo:

- Revisión bibliográfica. En esta etapa se buscó obtener información sobre el trabajo desarrollado en el área de la recuperación de información, así como la recuperación de imágenes de tatuajes específicamente.
- Estudio de diferentes métodos para extraer información utilizados en la recuperación de imágenes. Se analizaron técnicas como el algoritmo SIFT, la transformada Haar y otros extractores de características.
- Estudio de las pruebas y análisis de resultados que se realizan en la recuperación de imágenes. En esta parte se investigó la forma de evaluar los sistemas biométricos y los de recuperación de información visual.

- Estudio del lenguaje de programación a utilizar (C, C++).
- Construcción de la base de datos con imágenes de tatuajes de una agencia policiaca. En esta etapa se consiguieron imágenes provenientes de una agencia de policía del estado de California de Estados Unidos, donde se inspeccionó y recortó cada imagen para extraer únicamente el tatuaje, el cual se almacenó con un número identificador y una palabra clave de acuerdo a estándar ANSI-NIST ITL 1-2007.
- Estudio de técnicas de reducción de dimensionalidad. En esta etapa se investigó sobre formas de reducir la cantidad de información almacenada, así como la reducción del tiempo de búsqueda.
- Planeación y construcción del algoritmo. El objetivo de esta fase fue la de construir un algoritmo basado en el conocimiento obtenido.
- Pruebas y evaluación de resultados. Aquí se realizaron las pruebas sobre la base de datos y se analizaron los resultados haciendo comparaciones con las propuestas de otros autores mediante curvas CMC y de precisión-exhaustividad, así como comparaciones con la propuesta de este trabajo utilizando distintos parámetros tales como diferente cantidad de puntos utilizados.
- Escritura de la tesis. En esta fase se procede a documentar la información recopilada, las ideas propuestas, los resultados obtenidos, así como las conclusiones y recomendaciones surgidas a partir de este trabajo.

## **I.6 Importancia de la investigación**

La contribución del presente trabajo consiste en la investigación y análisis de distintos métodos de extracción de características para utilizarse en el área de la recuperación de información, así como la reducción de información necesaria para representar la imagen de tal forma que la recuperación continúe siendo efectiva.

En este caso específico de aplicación de imágenes de tatuajes, permite conocer las características típicas de este tipo de imágenes, así como también el detectar diversas

necesidades para construir el sistema de recuperación, el cual como se menciona, vendría a ser una herramienta útil en la identificación biométrica en las agencias policíacas. Por ejemplo, el FBI (Federal Bureau of Investigation por sus siglas en inglés) tiene un programa llamado “Next Generation Identification” (NGI) el cual consiste en desarrollar avances en los sistemas de identificación biométrica de tal manera que se incrementen las capacidades de los sistemas actuales. Entre los desarrollos previstos, se encuentra el mejoramiento en los sistemas de fotografías, esta mejora involucra la realización de búsquedas más eficientes, así como el poder incluir imágenes de cicatrices, marcas y tatuajes (SMT) (FBI, 2009).

## **I.7 Organización de la tesis**

En el resto del documento se presenta lo siguiente: En el capítulo 2 se presenta información sobre los sistemas de recuperación de imágenes basada en contenido visual, también se menciona sobre la forma de extraer información de las imágenes de tal forma que sea posible representarlas para que sean identificadas de manera automática. Además se presentan esquemas de búsqueda que tienen el objetivo de mejorar la eficiencia de recuperación.

En el capítulo 3 se presenta la base de datos que es utilizada para los experimentos, así como el proceso seguido para su creación a partir de ciertas imágenes originales.

En el capítulo 4 se muestran los experimentos realizados así como los resultados obtenidos con sus respectivos análisis y comparaciones.

Finalmente, en el capítulo 5 se presentan las conclusiones obtenidas en el desarrollo de la presente tesis, así como las recomendaciones de trabajo futuro.

## Capítulo II

---

### II Marco Teórico

---

#### I.1 Introducción a la recuperación de imágenes basada en contenido visual

La recuperación de información visual consiste en localizar imágenes relevantes en algún repositorio de base de datos dada una consulta específica. Los sistemas de recuperación de información visual reciben como entrada una imagen de consulta a la cual se le extrae información para que la imagen pueda compararse con los archivos de la base de datos, esto mediante una medida de similitud previamente establecida. Una vez obtenidos resultados de las comparaciones, se recuperan aquellas imágenes consideradas relevantes.

Una forma de realizar consultas, es mediante el uso de los metadatos, estos consisten en cadenas alfanuméricas que se refieren a anotaciones textuales para etiquetar las imágenes al momento de almacenar los archivos, se pueden referir a los objetos que contiene la imagen o al contenido general de la misma (del Bimbo, 2001). El uso de los metadatos para la recuperación de información se clasifica dentro de una primera generación de sistemas de recuperación de información visual. Es importante notar que la anotación textual para describir la imagen, puede no resultar una tarea sencilla ya que en ella interviene la interpretación personal de quien haga las anotaciones.

### II.1.1 Primera generación de sistemas de recuperación de información visual

En la Figura 5, se muestra un esbozo de los sistemas de recuperación de información visual de primera generación. El usuario incluye los metadatos mediante anotación textual correspondiente a cada elemento de la base de datos. Posteriormente, el sistema extrae información del texto que fue introducido. A partir de esta información, el sistema se encarga de formar un índice con todos los elementos que conforman el repositorio de la base de datos.

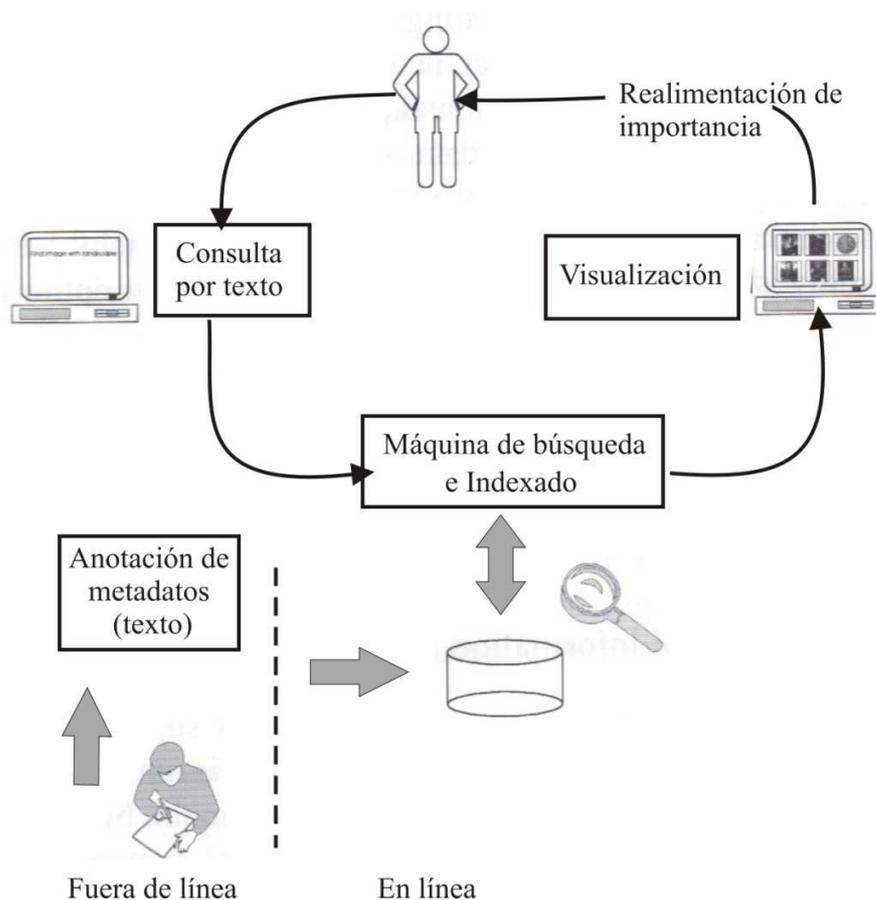


Figura 5. Esbozo de una primer generación de sistemas de recuperación de información visual para imágenes 2D. Fuente, del Bimbo(2001).

Cuando se quiere realizar la recuperación de alguna imagen, se introduce una consulta, la cual consiste en una cadena de caracteres introducida por el usuario que indica el contenido de la imagen a recuperar. A dicha consulta, el sistema le extrae el mismo tipo de información que extrajo de los elementos de la base de datos. Una vez que se tiene la información necesaria de la consulta, se entra a la etapa de la máquina de búsqueda, en donde se hacen comparaciones con los elementos indexados mediante alguna medida de similitud previamente elegida.

Ya que se han hecho las comparaciones necesarias para determinar cuáles son los elementos más relevantes para la consulta, se procede a desplegar la imagen que resultó más parecida a la consulta según las comparaciones realizadas. Los sistemas de recuperación de información visual permiten flexibilidad para visualizar una cantidad de imágenes ordenadas por su relevancia (Gunther y Beretta, 2001).

Dependiendo de la implementación del sistema de recuperación, es posible incluir una etapa de realimentación de importancia. Esto consiste en que una vez visualizadas las N imágenes relevantes según la medida de similitud establecida, el usuario indica cuáles son las que considera como relevantes según su percepción y a partir de ahí inicia nuevamente el ciclo de búsqueda en donde el sistema ya ha “aprendido” cuáles imágenes debe buscar y afina la relevancia de los resultados arrojados.

Como muestra la Figura 5, en esta primera generación de sistemas de recuperación de información visual, no se toma en cuenta el contenido visual de las imágenes, sino que la medida de similitud se lleva a cabo a partir de la información extraída de las anotaciones textuales. Sin embargo, en la segunda generación de sistemas de recuperación de información visual, sí se toma en cuenta el contenido de las imágenes (CBIR por las siglas en inglés de “Content-Based Image Retrieval”).

## II.1.2 Segunda generación de sistemas de recuperación de información visual

En la Figura 6, se muestra un esbozo de un sistema de recuperación de imágenes basado en contenido (del Bimbo, 2001). En este sistema, a diferencia del anterior, se lleva a cabo una extracción automática de propiedades o características visuales de las imágenes. Dicha extracción de características se realiza mediante alguna técnica previamente establecida. En la sección II.2, se detallan técnicas de extracción por medio de color, textura, forma y combinaciones de las mismas.

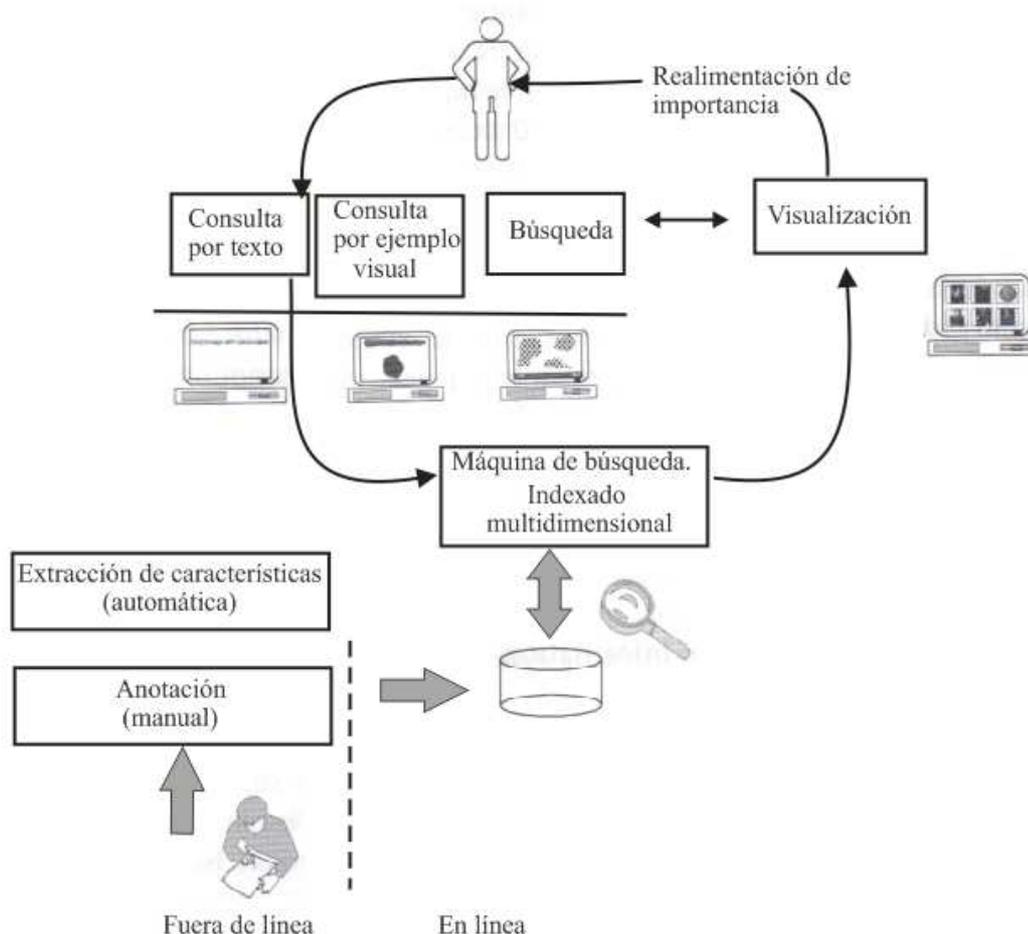


Figura 6. Esbozo de una nueva generación de sistemas de recuperación de información visual para imágenes 2D. Fuente, del Bimbo(2001).

Como se observa en la Figura 6, se pueden incluir metadatos mediante anotaciones textuales, además se extraen automáticamente características visuales para todas las imágenes que estarán indexadas en la base de datos. Debido a que la introducción de metadatos con texto y la extracción de características proveen diferentes tipos de representación para las imágenes, entonces una consulta puede realizarse de distintas maneras utilizando la información disponible que representa a la imagen.

### *II.1.3 Tipos de consulta en sistemas de recuperación de información visual*

Los tipos de entrada de una consulta en un sistema CBIR incluyen (del Bimbo, 2001) (Marchand-Maillet y Beretta, 2005):

**Consulta por anotaciones textuales.**- En esta consulta se introduce una cadena de caracteres que indica el contenido de la imagen que se quiere recuperar.

**Consulta por imagen ejemplo.**- Aquí se busca si una imagen específica se encuentra en la base de datos.

Consulta con una copia exacta de alguna imagen que este en la base de datos.

Consulta con parte de una imagen en la base de datos.

Consulta con una versión alterada geoméricamente de una imagen de la base de datos.

Consulta con una versión comprimida de una imagen en la base de datos.

**Consulta por similitud de imágenes.**- Aquí se introduce una imagen y se buscan imágenes parecidas en la base de datos, dichas imágenes no son necesariamente la misma imagen o alguna transformación de la consulta. La similitud entre imágenes se da según cierto criterio que depende de la métrica utilizada. Este tipo de recuperación no es sencilla ya que involucra diversos factores de interpretación, ya que la medida de similitud que se elija

puede no correlacionarse con los juicios de similitud de los seres humanos, a esto se le conoce como brecha semántica (Vasconcelos, 2007).

**Consulta con esbozo de imagen.** En este tipo de consulta se introduce un dibujo semejante según las capacidades artísticas del usuario, de aquella imagen que se quiere recuperar de la base de datos. Al igual que en la consulta por similitud, en este caso también se involucra la brecha semántica.

**Consultas combinadas.** Dependiendo de la imagen, también existe la opción de combinar la información tanto de contenido visual como de las anotaciones textuales de las imágenes. Esto permite hacer un filtrado de búsqueda mediante las anotaciones textuales, ya que las características obtenidas de las cadenas de caracteres suelen ser de menor dimensión que las características de contenido visual, lo cual incrementa la rapidez de recuperación.

Una vez que se ha introducido una consulta en un sistema CBIR, se procede a extraer las características correspondientes, las cuales deben ser del mismo tipo que las extraídas de las imágenes en la base de datos. Posteriormente se hace una comparación de la consulta con el índice de imágenes según la medida de similitud definida, para posteriormente desplegar una cantidad N de imágenes relevantes a la consulta. También se presenta la posibilidad de utilizar realimentación de importancia para afinar la calidad de los resultados según el juicio del usuario.

Como se ha visto, en los sistemas CBIR se toma en cuenta el contenido visual de las imágenes que se quieren recuperar. Una parte importante de estos sistemas es la extracción de características, ya que las imágenes que se indexarán en la base de datos estarán representadas mediante estas características extraídas y es mediante estas mismas con las que se realizarán las comparaciones, entonces los resultados obtenidos dependen fuertemente de ellas. En la siguiente sección se presenta la extracción de características en las imágenes para un sistema de recuperación de información visual basado en contenido.

## **II.2 Representación de imágenes para su recuperación por contenido visual**

Una imagen se puede definir como una función de dos dimensiones para  $I$ , teniendo a  $I(x, y)$  donde  $x$  e  $y$  son coordenadas espaciales (planas), la amplitud de  $I$  en cualquier par de coordenadas  $(x, y)$  se conoce como intensidad en ese punto. Esta definición incluye solo cambios de nivel de intensidad y por ello para este tipo de imágenes se utiliza el término de imagen con niveles de gris. Una imagen a color está formada por una combinación de imágenes de dos dimensiones individuales. Por ejemplo, en el sistema RGB, una imagen a color consiste de 3 imágenes individuales las cuales corresponden a los componentes rojo, verde y azul) (Gonzalez *et al.*, 2004).

En vez de utilizar toda la información que provee esta representación, la cual puede resultar redundante, se utiliza información relevante y descriptiva de las imágenes, donde las imágenes generalmente se transforman a una representación reducida que consta de un conjunto de características. Al proceso de transformar los datos de entrada en un conjunto de características, se le conoce como extracción de características (Guyon y Elisseeff, 2003).

### *II.2.1 Extracción de características*

Existen diversas técnicas de extracción de características (Smeulders *et al.*, 2000; Datta *et al.*, 2008), generalmente las técnicas están basadas en las propiedades perceptuales de las imágenes que incluyen color, textura y forma (del Bimbo, 2001). Una vez que se tiene el conjunto de características de una imagen, entonces la imagen se ha representado de otra manera. Por ejemplo, si se realiza extracción de características visuales de la imagen utilizando color, entonces se tiene una representación de la imagen con propiedades de color.

Una vez representadas las imágenes con los conjuntos de características a partir de alguna propiedad visual, entonces se pueden realizar comparaciones utilizando dichos conjuntos. Para realizar estas comparaciones, se utiliza una medida de similitud, propia de la técnica utilizada. Una medida de similitud entre una consulta y una imagen de la base de datos indica el parecido entre ambas mediante algún valor o porcentaje según sea el caso.

La elección del extractor de características para un sistema depende de las necesidades del mismo, ya que algunos extractores poseen invariancia a transformaciones, con otros se tiene la ventaja de rapidez en la búsqueda y otros pueden poseer otro tipo de ventajas, por lo que es importante conocer las fortalezas y debilidades de cada método. A continuación se muestran algunos ejemplos breves de algunas de las técnicas.

### *II.2.2 Técnicas de extracción de características*

Color.- Una de las formas más comunes de representar imágenes mediante color es con histogramas, de esta forma se tiene la ventaja que no se depende del tamaño de la imagen o de la orientación, aunque tiene la desventaja que no incluye información espacial, por lo que imágenes con diseños muy distintos pueden tener representaciones similares. Para hacer comparaciones, usualmente se utiliza la medida de similitud comparando los histogramas de color (Jain y Vailaya, 1996).

Textura.- Una imagen o cierta región de la misma posee textura si un conjunto de estadísticas locales u otras propiedades locales de la imagen son constantes, varían lentamente o son aproximadamente periódicas. Muchas texturas comunes se componen por pequeños elementos estructurales o textones, los cuales se encuentran generalmente en cantidades muy grandes para poder considerarse como regiones separadas. Una forma de representar imágenes mediante la propiedad perceptual textura es tomando en cuenta el tamaño espacial de los textones y representándolos como variaciones de niveles de gris bidimensionales. También se puede buscar textura en imágenes utilizando características

extraídas por filtros de Gabor (Ma y Manjunath, 1996) o usando estadísticas locales como la matriz de co-ocurrencia (Newsam y Kamath, 2005).

Forma.- Una imagen se puede representar en un espacio de formas mediante la obtención de formas en regiones particulares de la misma. Una forma se puede caracterizar mediante su área, elongación, orientación, esquinas, puntos característicos. Se puede utilizar un histograma de las direcciones de los bordes para representar este atributo, así como también se pueden utilizar momentos invariantes a transformaciones afines (Jain y Vailaya, 1998).

Combinaciones y otras técnicas.- No es necesario representar la imagen mediante solo un atributo, sino que incluso es conveniente combinar distintos tipos de características ya que la recuperación basada en un solo atributo puede no ser lo suficientemente discriminadora y el combinar las características provee eficiencia y efectividad en las consultas por contenido visual (Jain y Vailaya, 1996). Además se continúan generando nuevos extractores de características basados en transformaciones de imágenes (Jacobs *et al.*, 1995), y en representación en series de imágenes (Korikana y Chandrasekaran, 2008).

Hasta el momento se ha mostrado que las imágenes pueden representarse de diferentes maneras según las propiedades perceptuales que se hayan elegido. La representación contiene información descriptiva de la imagen la cual permite hacer comparaciones entre imágenes mediante alguna medida de similitud. Se quiere que una comparación entre dos imágenes arroje alta similitud cuando se trata de dos imágenes que según juicios humanos también posea alta similitud.

No todas las técnicas son capaces de ser efectivas cuando las imágenes han sufrido transformaciones u obstrucciones, en realidad aún se dista de tener sistemas de recuperación de información visual altamente eficientes. El algoritmo Scale Invariant Feature Transform (SIFT) (Lowe, 2004), el cual sirve para detectar características locales en imágenes, ha resaltado cuando se trata de localizar objetos en imágenes (Mikolajczyk y

Schmid, 2005), por lo que se ha elegido en el desarrollo de esta tesis y se explica a continuación.

### II.2.3 Extracción de características con SIFT

La idea general para utilizar SIFT en recuperación de imágenes según (Lee *et al.*, 2008) y (Bakken, 2007), es utilizar el algoritmo para detectar ciertos puntos que contienen información relevante de una imagen, a los cuales se les conoce como puntos de interés, ver Figura 7. Posteriormente, se hacen comparaciones a partir de información extraída alrededor de los puntos de interés tanto de la imagen de consulta como en aquellas que se encuentran en la base de datos. A esta información extraída del vecindario de cada punto de interés, se le conoce como descriptor. Por cada punto de interés se obtiene un descriptor que es un vector de dimensión 128 el cual se describe más adelante en este mismo capítulo.

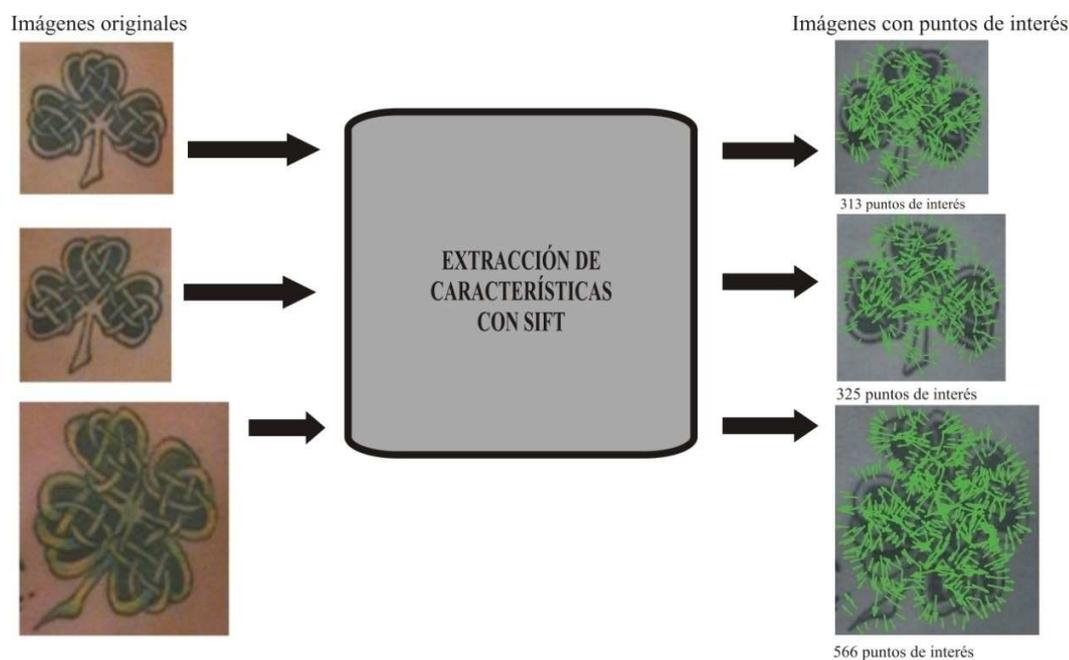
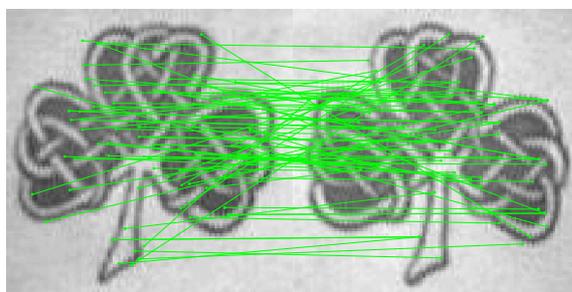
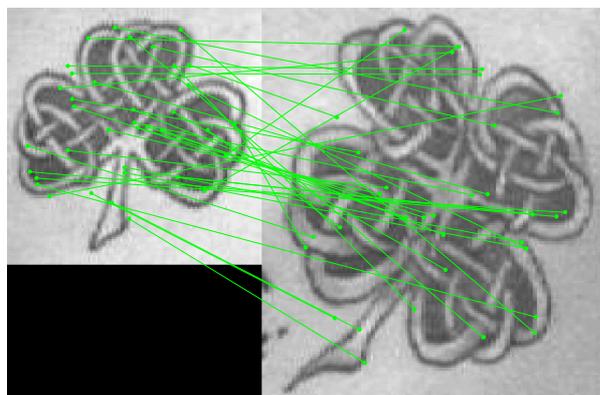


Figura 7. Obtención de puntos de interés con SIFT.

Para comparar dos imágenes, se obtiene una igualación entre los descriptores de los puntos de interés de ambas imágenes, ver Figura 8. En esta idea de recuperación de imágenes, la medida de similitud utilizada es la cantidad de igualaciones exitosas entre dos imágenes (más adelante se explica sobre igualaciones exitosas). Por lo que la consulta es igualada con los elementos de la base de datos para de esta manera obtener el número de igualaciones exitosas con todos los elementos, donde aquella imagen con la que se haya conseguido un mayor número de igualaciones exitosas se coloca en primer lugar de recuperación.



- a) 79 igualaciones exitosas entre imagen de consulta e imagen correspondiente en la base de datos dos imágenes similares.



- b) 38 igualaciones exitosas entre imagen de consulta e imagen no correspondiente en la base de datos .

Figura 8. Igualación entre puntos de interés de a) 2 imágenes correspondientes, b) 2 imágenes no correspondientes.

La forma en que el algoritmo SIFT extrae los puntos de interés de las imágenes es todo un proceso que involucra una serie de pasos. A continuación se explica la extracción de características mediante SIFT, así como la propuesta del autor del algoritmo para realizar igualación entre imágenes.

### *II.2.3.1 Algoritmo SIFT*

El algoritmo SIFT consta de cuatro etapas principales que se utilizan para la generación o extracción de lo que se conoce como el conjunto de características de las imágenes. A continuación se muestra una introducción de estas etapas:

**Detección extrema en el espacio de escala.** La primera etapa busca sobre toda la imagen puntos de interés potenciales que sean invariantes a cambios de escala y orientación.

**Localización de puntos de interés.** A cada punto que es candidato a punto de interés, se le aplica un modelo para determinar su ubicación y escala. Se entra en una etapa en donde se verifica qué tan estables son los puntos de interés, para de esta manera hacer un filtro de aquellos puntos que no son estables según ciertos criterios.

**Asignación de orientación.** Se asignan una o más orientaciones a cada posición de puntos de interés, basándose en las direcciones de los gradientes locales de imagen. A partir de este momento, todas las operaciones son llevadas a cabo en datos de la imagen que han sido transformados en base a la orientación que fue encontrada en el punto de interés, así como de su escala y posición. De ese modo, se provee de invariancia a estas transformaciones.

**Descriptor de puntos de interés.** En cada punto de interés se obtienen gradientes locales de la imagen en una región establecida alrededor del mismo. Con estos gradientes se forma un vector, representación conocida como descriptor del punto de interés.

### II.2.3.1.1 Detección extrema en el espacio de escala

La primera etapa es la detección de los puntos de interés, esta consiste en identificar todas aquellas características que se mantienen estables a cambios de escala en la imagen, se puede llevar a cabo haciendo una búsqueda a través de todas las escalas posibles usando una función de escala continua, en la representación conocida como espacio-escala.

Una noción del concepto espacio-escala en el caso de imágenes bidimensionales (Lindeberg, 1994) es que para una imagen dada  $I(x, y)$ , su representación en el espacio escala lineal (Gaussiano) está dado por una familia de señales derivadas  $L(x, y, \sigma)$  las cuales están definidas por la convolución de la imagen  $I(x, y)$  con el kernel Gaussiano  $G(x, y, \sigma)$ :

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \quad (1)$$

donde  $*$  es la operación convolución en  $x$  e  $y$ ,  $y$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}. \quad (2)$$

Esta definición de  $L$  implica que funciona de forma continua para  $\sigma \geq 0$ , pero típicamente solo se considera un conjunto finito de niveles en la representación espacio-escala.

Para detectar los puntos de interés SIFT, la imagen se convoluciona con filtros Gaussianos para producir el conjunto de imágenes del espacio escala, lo anterior se ilustra en la Figura 9, donde además se puede observar que se obtiene un conjunto de imágenes de espacio escala para cada imagen la cual es submuestreada a la mitad varias veces, cada vez que una imagen se submuestra a la mitad entonces se tiene una nueva octava. La primera octava es la imagen original con sus imágenes de espacio escala, una vez que se ha procesado esta imagen, entonces se submuestra cada imagen de la etapa anterior por un factor de dos, y se convoluciona con una Gaussiana que tiene el doble del valor inicial de  $\sigma$  en dicha octava y entonces se tiene la imagen inicial para la siguiente octava y así sucesivamente.

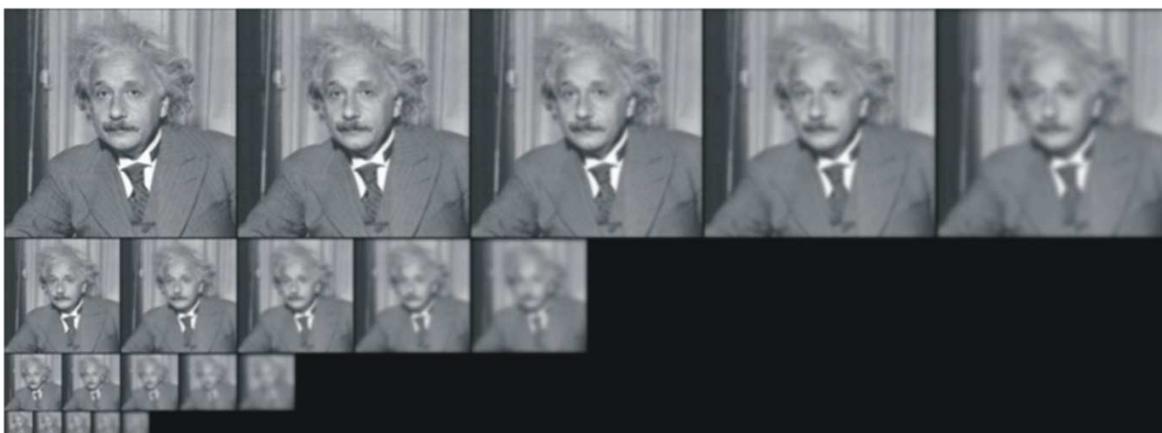


Figura 9. Imagen convolucionada con kernel Gaussiano con diferentes sigmas para producir imágenes de espacio escala para cada octava.

Una vez que se tiene lo anterior, se procede a tomar una diferencia sucesiva de las imágenes suavizadas del espacio escala. A esto se le conoce como diferencia de Gaussianas (DoG), la cual está dada por:

$$D(x, y, \sigma) = L(x, y, k_i\sigma) - L(x, y, k_j\sigma) \quad (3)$$

Donde la diferencia de Gaussianas entre las escalas  $k_i\sigma$  y  $k_j\sigma$  representa la diferencia de las imágenes suavizadas con el kernel Gaussiano en las escalas  $k_i\sigma$  y  $k_j\sigma$ .

En la Figura 10 se ilustra la obtención de las diferencias de Gaussianas. En la parte izquierda se tienen las imágenes del espacio escala en cada octava, mientras que en la parte derecha se tiene las diferencias de Gaussianas obtenidas de la diferencia de 2 imágenes suavizadas adyacentes. En la Figura 11 se ilustra el resultado aplicar el operador diferencia de Gaussianas a las imágenes de la Figura 9.

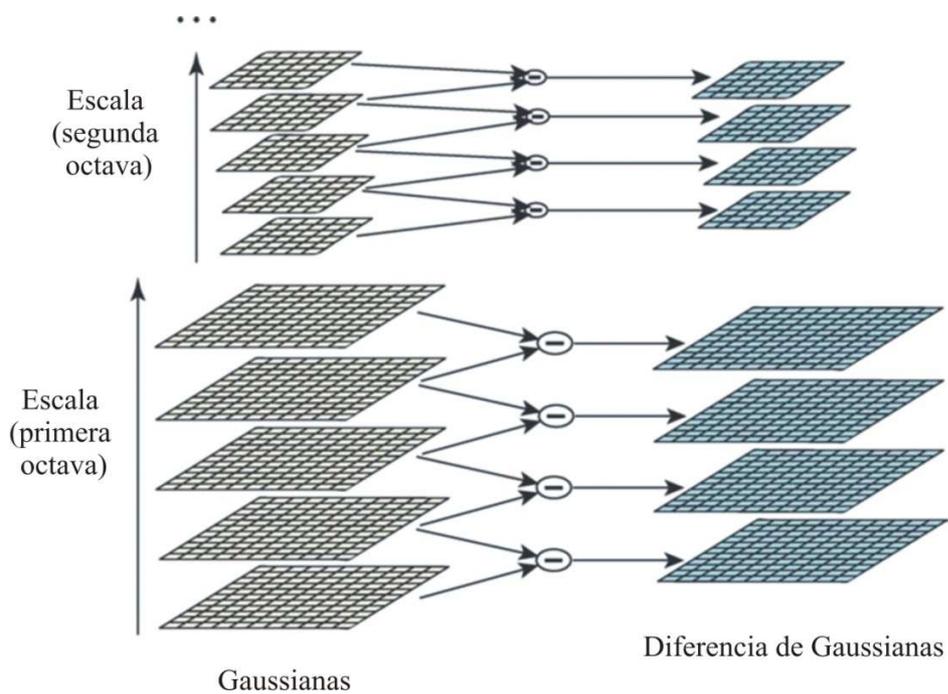


Figura 10. Obtención de diferencias de Gaussianas. Fuente (Lowe, 2004).



Figura 11. Diferencias de Gaussianas de imágenes de la Figura 9.

Ya que se tienen las imágenes  $D(x, y, \sigma)$ , lo que se busca ahora es encontrar los extremos o máximos/mínimos. Para esto, cada punto es comparado con sus 8 vecinos en la imagen

DoG actual y nueve vecinos en la escala de arriba y nueve en la de abajo (ver Figura 12). Se seleccionan como puntos de interés candidatos solo aquellos puntos que son mayores o menores que sus vecinos.

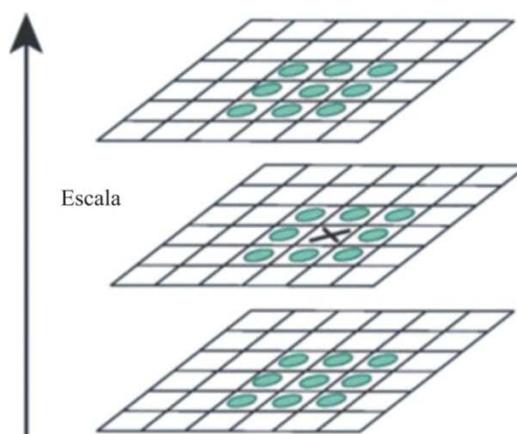


Figura 12. La detección de puntos máximos y mínimos en las diferencias de Gaussianas se lleva a cabo comparando un píxel (marcado con X) con sus 26 vecinos en regiones de 3x3 en la escala actual y en las escalas adyacentes (marcados con círculos). Fuente (Lowe, 2004).

#### II.2.3.1.2 Localización precisa de puntos de interés.

El procedimiento anterior para obtener los puntos de interés candidatos, produce muchos puntos de interés, algunos de los cuales son inestables. El siguiente paso en el algoritmo, es ajustar para posición, escala y proporción de las curvaturas principales aquellos datos cercanos. Esta información permite que se rechacen los puntos que tienen bajo contraste o los que son localizados pobremente.

Originalmente solo se calculaba cada punto de interés en la posición y escala de los puntos de interés candidatos. Pero recientemente, se usa una interpolación en los datos cercanos para determinar precisamente su posición. Se calcula la posición interpolada del máximo, lo cual mejora sustancialmente la estabilidad y la búsqueda de correspondencias

(igualaciones exitosas). La interpolación se realiza usando la expansión de Taylor cuadrática de la función de diferencia de Gaussianas, ec. (3), de tal manera que los puntos candidatos son el origen. Esta expansión de Taylor, está dada por:

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x} \quad (4)$$

Donde  $D$  y sus derivadas se evalúan en el punto candidato y  $\mathbf{x} = (x, y, \sigma)$  es la nueva posición. La posición del extremo  $\hat{\mathbf{x}}$  se determina tomando la derivada de esta función con respecto a  $\mathbf{x}$  y poniéndola a cero. Entonces resolviendo

$$\hat{\mathbf{x}} = - \left( \frac{\partial^2 D}{\partial \mathbf{x}^2} \right)^{-1} \frac{\partial D}{\partial \mathbf{x}}. \quad (5)$$

Si la compensación  $\hat{\mathbf{x}}$  es más grande que 0.5 en cualquier dimensión, esto indica que el extremo está cerca de otro punto candidato. En este caso, el punto candidato se cambia y la interpolación se lleva a cabo ahora en ese punto. El desplazamiento final  $\mathbf{x}$  se agrega a la posición de su punto candidato para obtener la estimación interpolada para la posición del extremo.

El valor de la función en el extremo  $D(\hat{\mathbf{x}})$  es útil para rechazar extremos inestables con bajo contraste. Esto se puede obtener, sustituyendo la ecuación (5) en (4), obteniendo

$$D(\hat{\mathbf{x}}) = D + \frac{1}{2} \frac{\partial D^T}{\partial \mathbf{x}} \hat{\mathbf{x}}. \quad (6)$$

### II.2.3.1.3 Asignación de orientaciones

En este paso, a cada punto de interés se le asigna una orientación consistente, basada en las propiedades locales de la imagen. Este es un paso clave para lograr la invariancia a la rotación, ya que el punto de interés se puede representar relativo a su orientación.

Primero, se toma la imagen suavizada por el filtro Gaussiano  $L(x, y, \sigma)$  en la escala  $\sigma$  donde fue detectado el punto de interés, para hacer el procedimiento invariante a escala. Para cada muestra de la imagen  $L(x, y)$  en esta escala, se pre-calculan la magnitud del gradiente,  $m(x, y)$  y la orientación  $\theta(x, y)$  usando diferencias finitas.

$$m(x, y) = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2}, \quad (7)$$

$$\theta(x, y) = \tan^{-1} \left( \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right). \quad (8)$$

A partir de los cálculos realizados en una región alrededor del punto de interés, se toman las orientaciones de gradientes resultantes y se forma un histograma. Este histograma, el cual se conoce como histograma de orientación, cuenta con 36 compartimientos, donde en cada compartimiento se cubren 10 grados, de esta forma el histograma cubre el rango de 360 grados. Cada muestra de esa región que es agregada al histograma se pondera con su magnitud del gradiente y por una ventana circular ponderada Gaussiana que tiene una escala  $\sigma$  que es 1.5 veces la escala del punto de interés. Los picos en el histograma, corresponden a las direcciones dominantes de los gradientes locales. El pico más alto en el histograma corresponde a la orientación del punto de interés, si existe cualquier otro pico local que está dentro del 80% del pico más alto, se usa para crear otro punto de interés con esa orientación. En el caso de múltiples orientaciones se crean puntos de interés adicionales con la misma escala y posición para cada orientación adicional (ver Figura 13).

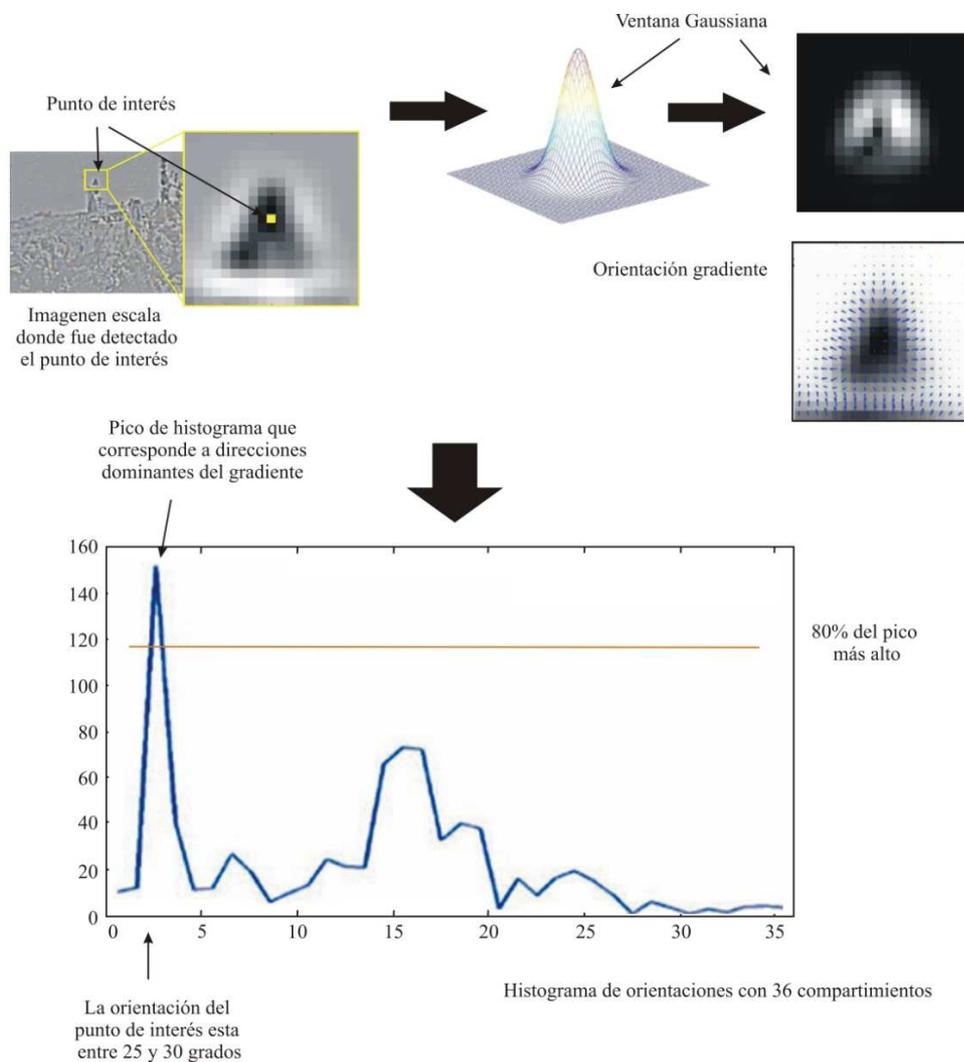


Figura 13. Obtención de orientación de puntos de interés.

#### II.2.3.1.4 Descriptor de puntos de interés

Hasta el momento, se ha mostrado de qué forma se encuentran por SIFT los puntos de interés en una imagen y se tiene de ellos su posición, escala y orientación. Ahora se desea que estos puntos de interés posean vectores descriptores los cuales sean altamente distintivos y parcialmente invariantes a otros parámetros, como iluminación y cambio de punto de vista en 3D.

El descriptor de características del punto de interés consiste de un conjunto de histogramas de orientación en vecindarios de pixel de 4x4. Los histogramas de orientación son relativos a la orientación del punto de interés y los datos de orientación vienen de la imagen Gaussiana con la escala más cercana a la escala del punto de interés. Tal como en el paso anterior, la contribución de cada pixel se pondera por la magnitud del gradiente, y por una Gaussiana con  $\sigma = 1.5$  veces la escala del punto de interés. Los histogramas contienen 8 compartimientos cada uno, y cada descriptor contiene un arreglo de 4x4, es decir 16 histogramas alrededor del punto de interés. Esto deja un vector de características de  $4 \times 4 \times 8 = 128$  elementos. Este vector se normaliza para mejorar la invariancia a cambios en la iluminación (ver Figura 14).

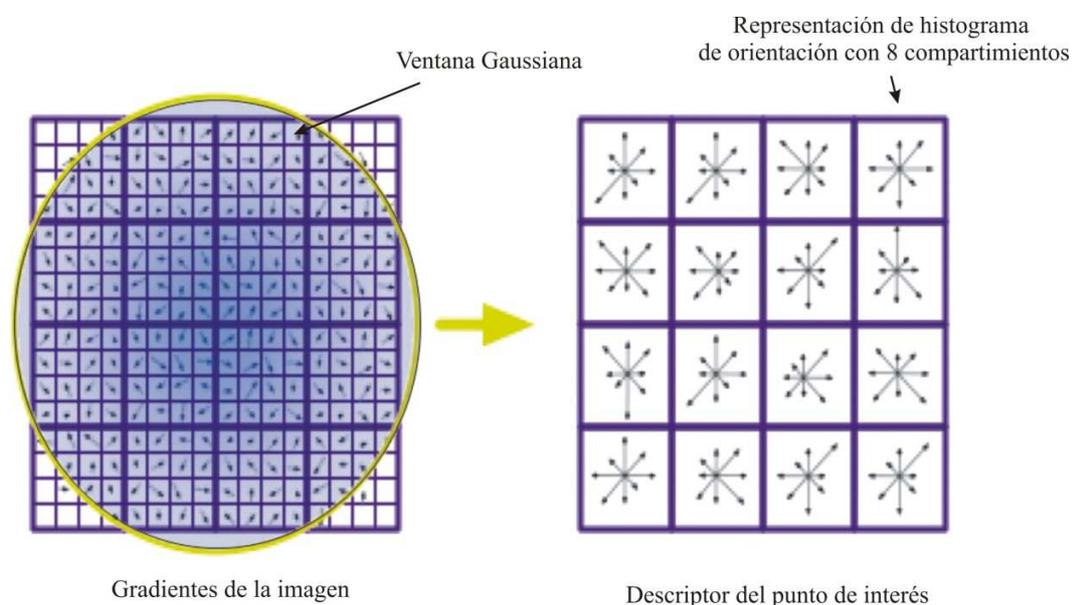


Figura 14. Obtención del descriptor del punto de interés.

#### II.2.3.1.5 Comparación entre dos imágenes con SIFT

Para comparar dos imágenes (A y B) utilizando la información que se provee al utilizar el algoritmo SIFT sobre ellas, se debe encontrar el vecino más cercano de cada punto de interés de la imagen A en la imagen B, donde el vecino más cercano se define como el

punto con la distancia Euclidiana mínima. Este punto será una igualación exitosa entre las dos imágenes si la proporción entre la distancia del vecino más cercano con el segundo vecino más cercano es menor que 0.8. La razón para rechazar aquellas igualaciones que resultaron con una proporción mayor de 0.8 entre el primero y segundo vecino más cercano, es porque se espera que en una igualación exitosa, el vecino más cercano se encuentre lo suficientemente cerca, mientras que el segundo más cercano lo suficientemente lejos. Este criterio, el cual fue obtenido del trabajo de Lowe (2004) elimina el 90% de los falsos positivos, aunque también se elimina el 5% de los verdaderos positivos, sin embargo, según experimentos esta proporción presenta los mejores resultados. Los falsos positivos pueden haber surgido de puntos de interés de imágenes con información adicional o incluso porque puede que haya puntos de interés que no hayan sido detectados.

En la Figura 15 se observa la recuperación de imágenes utilizando los descriptores SIFT, usando como medida de similitud entre dos imágenes el número de igualaciones exitosas.

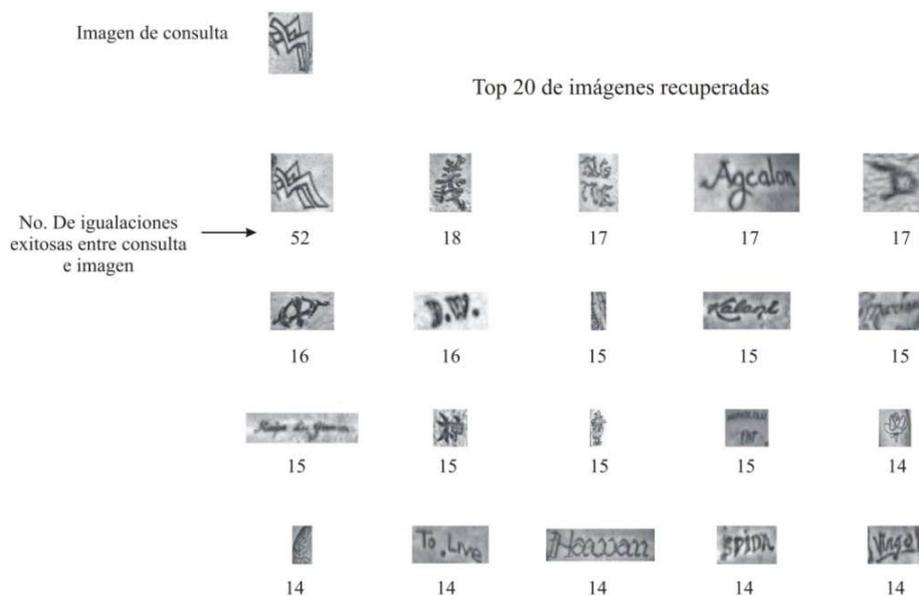


Figura 15. Top 20 de imágenes recuperadas usando SIFT y realizando comparaciones de la imagen de consulta con cada una de las imágenes de la base de datos.

Como se ha mostrado, las imágenes pueden representarse mediante descriptores SIFT. Con este algoritmo es posible realizar consultas por imagen ejemplo (Lee *et al.*, 2008). Sin embargo, el algoritmo no funciona efectivamente para realizar consulta por similitud de imágenes.

Una aplicación que se puede utilizar para realizar consultas por similitud, es imgSeek, que es un manejador y visor de colecciones de fotografías con búsqueda por contenido visual y algunas otras características (Niederberger C., 2007). Esta aplicación se basa en un artículo que propone representar las imágenes mediante descomposición wavelet (Jacobs *et al.*, 1995). A continuación se explica un poco más de este método:

#### *II.2.4 Extracción de características con Wavelets*

La idea es que la imagen de consulta se pueda dibujar a mano en algún lienzo digital, o que haya sido escaneada incluso con baja calidad y que el sistema sea capaz de realizar la recuperación de forma exitosa. Tanto la imagen de consulta, como las imágenes de la base de datos se representan mediante coeficientes obtenidos con descomposición wavelet bi-dimensional multi-resolución. La medida de similitud consiste en comparar los coeficientes wavelet significativos que tienen en común las imágenes. El algoritmo resultante es sencillo, requiere de poco almacenamiento para la representación de las imágenes además de proporcionar velocidad en la recuperación. Sin embargo, tiene la desventaja de no ser invariante a la rotación ni ser robusto ante obstrucciones en las imágenes.

En la Figura 16 se muestra el procedimiento para extraer características en las imágenes mediante esta propuesta. Se obtiene el color promedio de la imagen  $I$  y se representa mediante  $I^c[0,0]$  en cada canal del modelo de color utilizado. Se obtiene la transformada wavelet bi-dimensional Haar de la imagen en cada canal del modelo de color, de la cual se seleccionan  $m$  que posean la mayor magnitud. Se cuantiza cada coeficiente truncando a solo dos niveles, +1 que representa los coeficientes positivos y -1 representando los coeficientes negativos.  $\tilde{I}^c[i,j]$  representa el  $[i,j]$ -ésimo coeficiente truncado cuantizado.

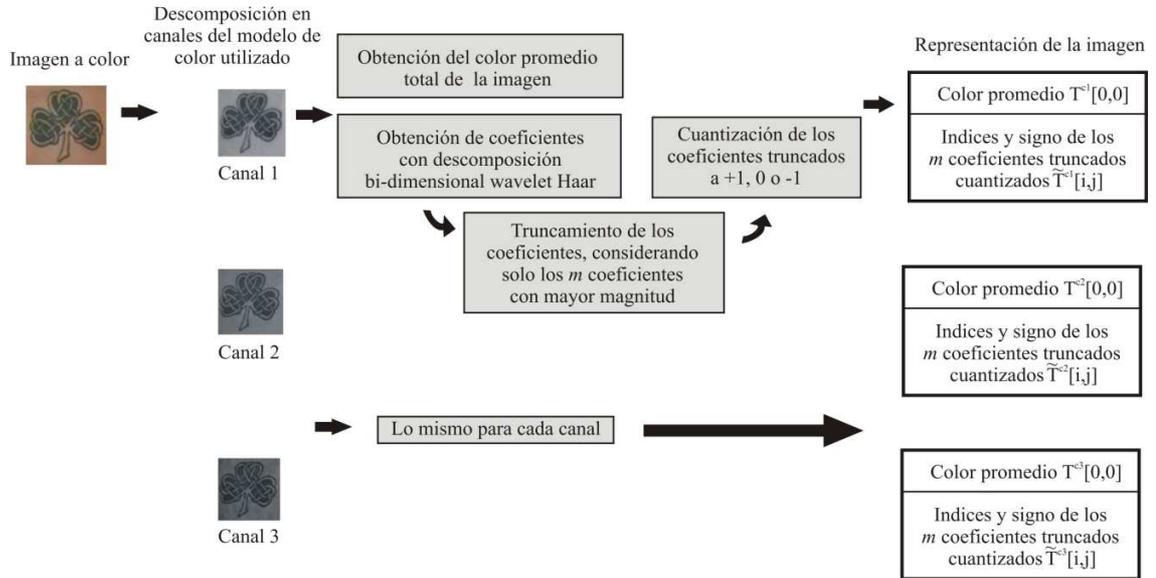


Figura 16. Extracción de características de imagen en Imgseek.

Cuando se desea recuperar una consulta dada, primero se extraen las mismas características a la imagen de consulta  $Q$  de tal forma que se obtenga el color promedio de cada canal en la imagen  $Q^c[0,0]$  y los coeficientes wavelets truncados cuantizados  $\tilde{Q}^c[i,j]$ . Posteriormente se procede a utilizar la medida de similitud la cual consiste en la diferencia del promedio de cada canal de color entre la consulta y la imagen de la base de datos y el número de coeficientes wavelet correspondientes que son iguales en ambas imágenes:

$$d(Q, I) = w_0 |Q[0,0] - I[0,0]| - \sum_{i,j: \tilde{Q}[i,j] \neq 0} w_{bin(i,j)} (\tilde{Q}[i,j] = \tilde{I}[i,j]) \quad (9)$$

Los pesos  $w_{bin(i,j)}$  en la ecuación 9 proveen un mecanismo de afinación para la medida de similitud en diferentes bases de datos y estilos de consultas, se obtienen experimentalmente con cierta cantidad de imágenes de entrenamiento.

Para optimizar el proceso de búsqueda, se constituye un índice mediante seis arreglos llamados arreglos de búsqueda, donde cada uno corresponde a cada combinación de signos ( + o - ) y un canal de color (tal como R,G y B). Por ejemplo, sea  $D_+^c$  un arreglo de búsqueda para el color en el canal  $c$ . Cada elemento  $D_+^c[i,j]$  de este arreglo contiene una lista de todas las imágenes  $I$  que tienen coeficientes wavelet  $\tilde{I}^c[i,j]$  en el canal de color  $c$ .

Como se menciona, las imágenes se pueden representar de diversas maneras donde cada opción tiene ventajas y desventajas, por ejemplo la representación con los descriptores SIFT permite recuperar imágenes incluso bajo transformaciones en las imágenes, sin embargo tiene la desventaja de que el almacenamiento de la representación es alto además de que la implementación de la medida de similitud puede ocasionar que la recuperación sea lenta.

Existen técnicas que tienen el objetivo de disminuir la cantidad de información a utilizar, para de esta manera reducir el almacenamiento y acelerar la velocidad de búsqueda. También es posible utilizar estructuras de índices de tal forma que se eviten búsquedas secuenciales y se aumente la eficiencia del sistema de recuperación. Para elegir el método de reducir la información y acelerar la búsqueda es necesario conocer la técnica de extracción de características a utilizar, así como la medida de similitud para realizar comparaciones.

### **II.3 Reducción de información**

Cuando una base de datos de donde se espera recuperar información, es formada por miles o millones de imágenes, entonces el almacenamiento y la velocidad de recuperación se vuelven puntos importantes.

De los métodos mostrados en la sección anterior, en la presente tesis se utiliza la extracción de características con SIFT para encontrar imágenes mediante consultas por imagen ejemplo. Este algoritmo arroja bastante información por cada imagen (886 descriptores en

promedio por imagen en la base de datos utilizada en el presente trabajo). Cada descriptor SIFT tiene una dimensión de 128, por lo tanto el encontrar una manera de reducir la cantidad de información que se utiliza para la medida de similitud vendría a cooperar con la reducción de almacenamiento y a acelerar el proceso de búsqueda.

Existen diferentes técnicas para obtener una representación de menor dimensionalidad, algunos métodos que reducen el tiempo de acceso y el tamaño de la base de datos incluyen: Análisis de Componentes Principales (PCA), Análisis de Componentes Independientes (ICA), Análisis de Factores, aprendizaje de variedades (manifold learning) (Laplacian EigenMAP, ISOMAP, LLE, LPP) (He y Niyogi, 2003) (Belkin y Niyogi, 2003), Algoritmos genéticos, redes neuronales y proyecciones aleatorias (Fodor, 2002) entre otras.

En el trabajo de Nistér y Stewenius (2006) se propone cuantizar los descriptores SIFT en palabras visuales utilizando un árbol de vocabulario. En este trabajo se utilizará una versión modificada de la representación de Nistér y Stewenius (2006). A continuación se describen los pasos para representar los descriptores SIFT de una imagen en “palabras visuales”.

### *II.3.1 Vocabularios de palabras (Bolsa de palabras)*

A partir de los descriptores SIFT de las imágenes de entrenamiento, se realiza un agrupamiento (*cluster*) jerárquico para formar un árbol de vocabulario. Este agrupamiento jerárquico consiste en elegir un valor  $k$  y aplicar un procedimiento inicial de  $k$ -medias sobre los datos de entrenamiento para de esta forma obtener  $k$  centroides de los agrupamientos. Entonces los datos de entrenamiento son particionados en  $k$  grupos, donde cada grupo consiste en los vectores descriptores más cercanos al centroide del agrupamiento particular. Se repite el mismo proceso de forma recursiva a cada uno de los grupos de vectores hasta tener un número  $L$  de niveles (ver Figura 17).

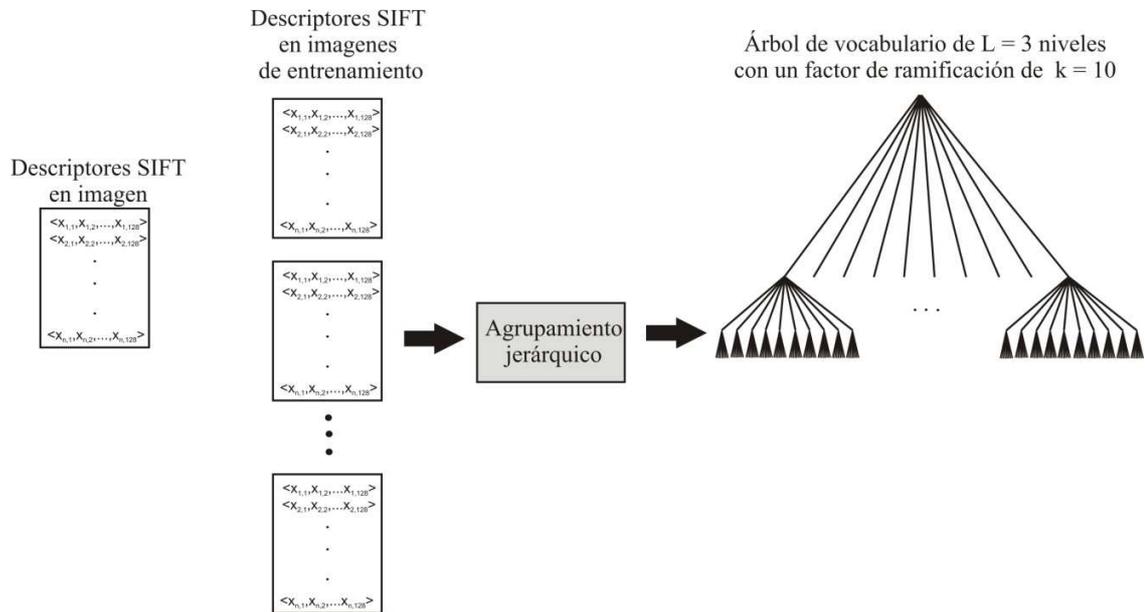
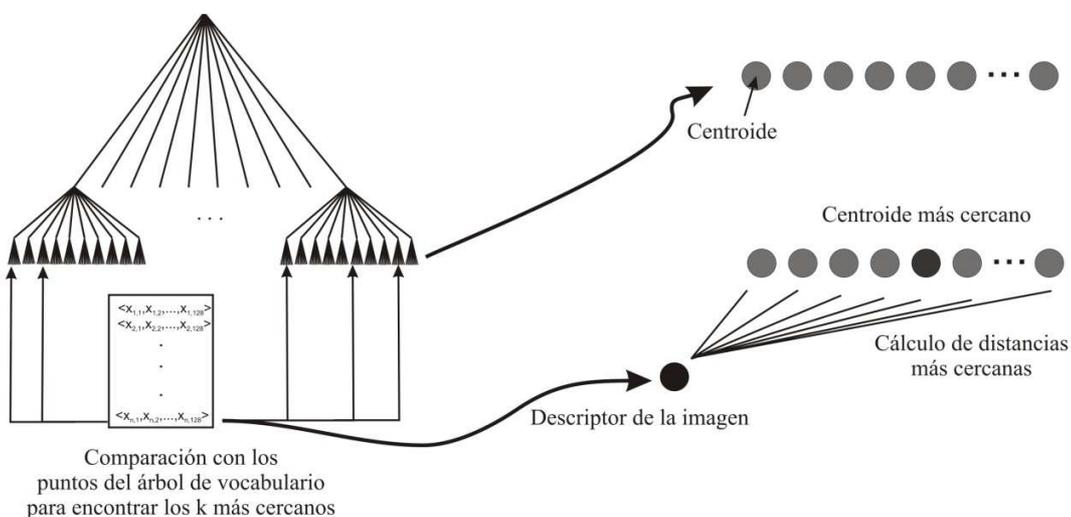


Figura 17. Agrupamiento jerárquico a partir de algunas imágenes de entrenamiento.

Una vez que se tiene el árbol de vocabulario, entonces se procede a representar las imágenes mediante el mismo. Como se muestra en la Figura 18, de cada descriptor de cada imagen se obtienen los puntos del vocabulario más cercanos, a este paso se le conoce como cuantización, ya que cada descriptor de dimensión 128 se representará ahora mediante el índice del centroide más cercano.



#### Ejemplo con 5 centroides

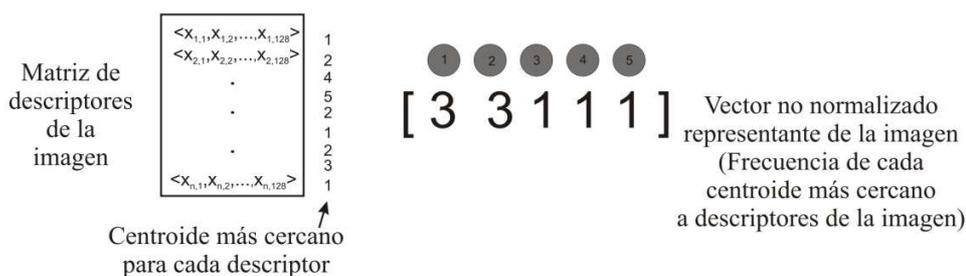


Figura 18. Cuantización de una imagen con los elementos del agrupamiento. (arriba) Obtención del centroide más cercano para un descriptor. (abajo) Ejemplo de representación con 5 centroides. Se obtiene el centroide más cercano de cada descriptor, posteriormente el vector representante es la frecuencia con que cada centroide fue vecino más cercano de los descriptores de la imagen representada.

Los componentes de la imagen de consulta  $Q$  y la imagen de la base de datos  $I$ , se representan así:

$$Q_i = n_i w_i, \quad (10)$$

$$I_i = m_i w_i. \quad (11)$$

Donde  $n_i$  y  $m_i$  son el número de descriptores de las imágenes de consulta y de la base de datos respectivamente, que tienen un elemento en el nodo  $i$ , mientras que  $w_i$  es un peso

asignado a cada nodo  $i$  el cual en el caso más simple es un valor constante o puede ser basado en la entropía, por ejemplo:

$$w_i = \ln \frac{N}{N_i} \quad (12)$$

En donde  $N$  es el número de imágenes en la base de datos y  $N_i$  es el número de imágenes con al menos un punto por el nodo  $i$ . En nuestra implementación, se asignaron  $w_i$  constantes, ya que se realizó un experimento utilizando la ecuación 12 y no se obtuvo mejora en los resultados.

La distancia que se utiliza para realizar comparaciones y determinar la relevancia de las imágenes en la base de datos, compara las frecuencias de los centroides en las representaciones de las imágenes:

$$d(Q, I) = \left\| \frac{Q}{\|Q\|} - \frac{I}{\|I\|} \right\|, \quad (13)$$

donde  $\|\cdot\|$  representa la norma. En este trabajo utilizamos la norma  $L_1$ .

En los experimentos realizados se observa lo mencionado por Nistér y Stewenius (2006) que consiste en que entre mayor sea la cantidad de hojas en el árbol de vocabulario, mejores son los resultados obtenidos al momento de la recuperación de imágenes.

En la Figura 19, se muestra el esquema para realizar consultas y agregar imágenes a una base de datos utilizando representación de imágenes por medio de descriptores SIFT y la utilización de bolsas de palabras.

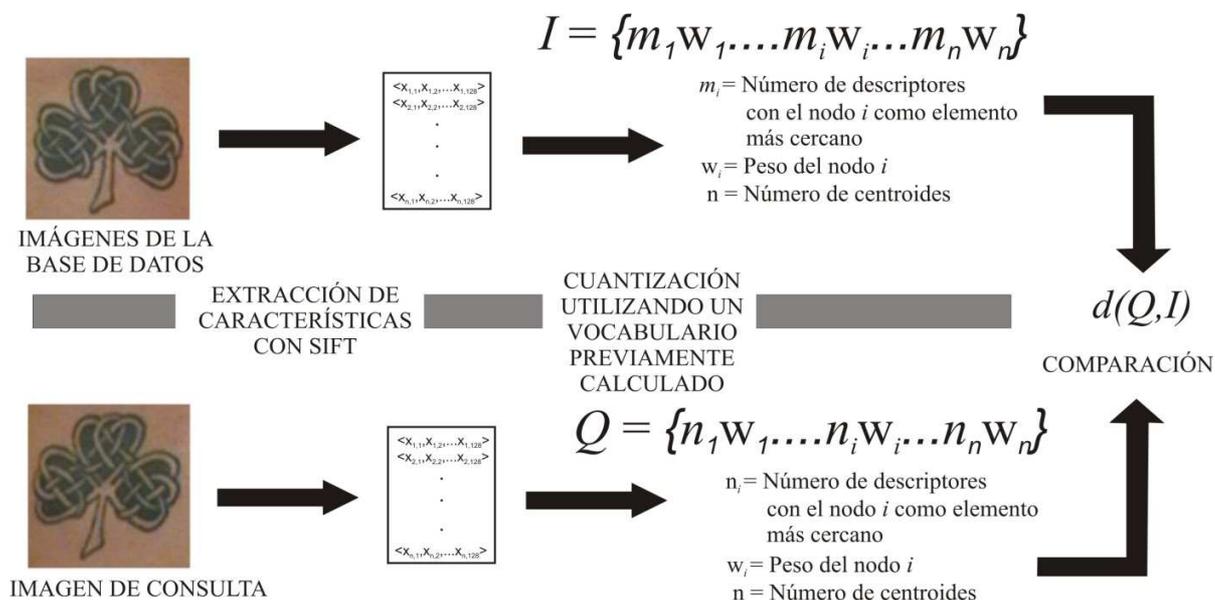


Figura 19. Sistema de recuperación de imágenes utilizando SIFT y bolsa de palabras.

Primero se tiene la imagen a la cual se le extraen los vectores descriptores con el algoritmo SIFT, posteriormente se utilizan vectores del árbol de vocabulario para cuantizar las imágenes a partir de los descriptores. A partir de dicha cuantización se procede a realizar comparaciones entre las imágenes de consulta y de la base de datos, de donde se obtendrá el puntaje que servirá para realizar la recuperación de las  $N$  imágenes más similares.

Como se puede observar, gracias a la cuantización mediante el árbol de vocabulario se reduce la cantidad de información utilizada para representar las imágenes, ya que alguna imagen en vez de representarse con una cantidad  $N$  de descriptores de dimensión 128, se representará mediante un vector que contendrá la frecuencia con que cada nodo del árbol de vocabulario fue más cercano a alguno de los  $N$  descriptores, donde no es necesario almacenar aquellos nodos que no fueron más cercanos a algún descriptor.

Otra técnica con la cual se puede reducir la cantidad de información para realizar búsquedas en una base de datos con la cual se realizaron experimentos durante el desarrollo de la presente tesis, es la utilización de permutaciones ordenadas (Chávez *et al.*, 2008).

### II.3.2 Permutaciones

Las permutaciones ordenadas consisten en un algoritmo de búsqueda por proximidad probabilístico que tiene la idea central de predecir la cercanía entre cualquier dos objetos en un espacio métrico, por medio de comparar la forma en que dichos objetos ordenan sus distancias hacia un conjunto de objetos anclados llamados *permutantes*.

Sean  $\mathbf{U}$  el conjunto de los elementos de la base de datos,  $\mathbf{P} \subseteq \mathbf{U}$  los puntos *permutantes*,  $x \in \mathbf{X}$  los elementos que definen una permutación y sea  $d$  una distancia, en la Figura 20 se ilustra la obtención de permutaciones en dos dimensiones. Los puntos circulares representan a los puntos anclados *permutantes*, los cuales se pueden elegir aleatoriamente, mientras que los asteriscos son puntos de la base de datos. En la tabla se muestra el orden en que cada objeto “ve” a los puntos *permutantes*, es decir, el orden de los puntos *permutantes* de menor a mayor distancia con respecto a cada objeto. Entonces, cada objeto se representa mediante una permutación  $\Pi_x$  que indica a los elementos de  $\mathbf{P}$  escritos de menor a mayor distancia de  $x$ , donde se tiene que para toda  $1 < i < k$ , se mantiene que  $d(p_{\Pi_x(i)}, x) < d(p_{\Pi_x(i+1)}, x)$  o  $d(p_{\Pi_x(i)}, x) = d(p_{\Pi_x(i+1)}, x)$  y  $\Pi_x(i) < \Pi_x(i + 1)$ .

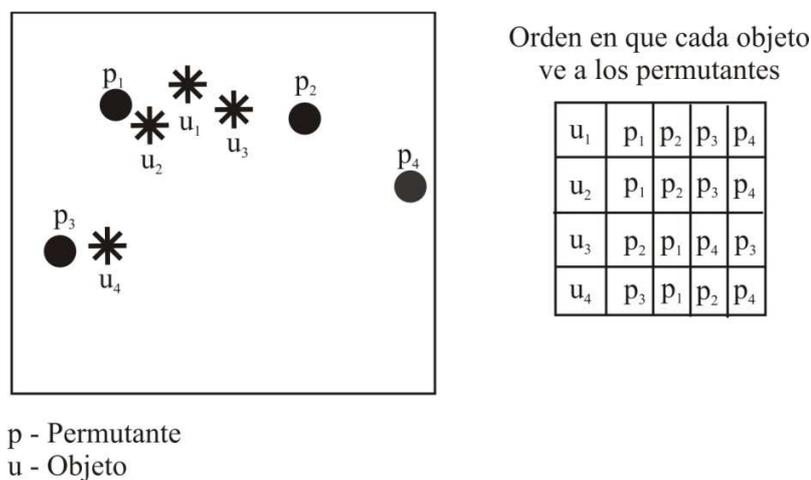


Figura 20. Obtención de permutaciones en puntos de una base de datos.

El algoritmo se basa en que de la misma forma de que dos elementos cercanos tienen distancias similares a puntos permutantes, dos elementos cercanos “verán” a los permutantes en un orden similar de cercanía. En la figura anterior, se puede visualizar que los objetos  $u_1$  y  $u_2$  ven en el mismo orden a los puntos permutantes por lo que se puede esperar que sean elementos cercanos en la base de datos, mientras que una diferencia en la forma en que dos objetos ven a los puntos permutantes puede señalar que dichos elementos no son tan cercanos como el caso de  $u_1$  con  $u_4$ . Sin embargo, el algoritmo no permite probar las distancias entre dos elementos y se tiene que realizar un paso de búsqueda con los objetos en su representación original.

Los pasos que se llevan a cabo consisten en la elección de un parámetro  $k$  el cual indica la cantidad de permutantes a utilizar, posteriormente se elige un conjunto de  $k$  permutantes  $\mathbf{P} = \{p_1, p_2, \dots, p_k\}$  de forma aleatoria a partir de  $\mathbf{U}$ . Después se obtiene la permutación  $\Pi_u$  para cada elemento  $u \in \mathbf{U}$ .

Para realizar la búsqueda de una consulta, se computa la permutación  $\Pi_q$  con los mismos  $k$  permutantes con los que se calcularon las permutaciones de la base de datos. Después se procede a encontrar aquellos elementos que son más similares a  $\Pi_q$  mediante una distancia  $S$  (más adelante se presenta esta fórmula).

Una vez que se han obtenido las distancias entre las permutaciones, se realiza un ordenamiento de  $\mathbf{U}$  de acuerdo a sus distancias  $S(\Pi_u, \Pi_q)$  ordenadas de menor a mayor. Lo que se obtiene con este ordenamiento son los puntos más cercanos a una consulta según la representación con permutaciones, sin embargo es necesario utilizar la representación original y obtener las distancias  $d(u, q)$  para un cierto número de puntos  $u$  que se encuentren entre los más cercanos según las distancias obtenidas con  $S$ . ver Figura 21.

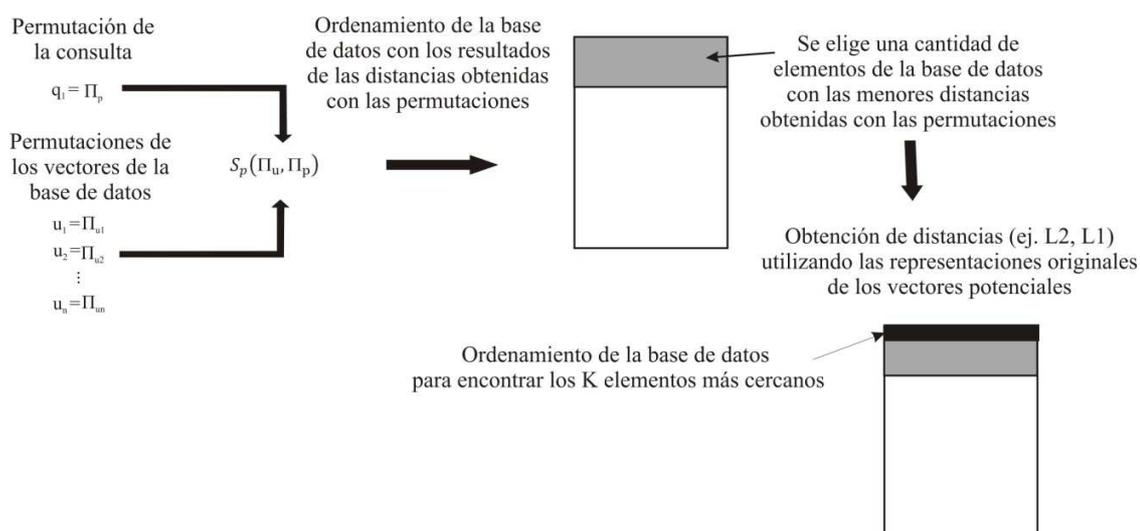


Figura 21. Consultas de más cercanos mediante permutaciones.

El hecho de obtener distancias con las representaciones originales es porque como ya se ha mencionado, este algoritmo no garantiza encontrar los vecinos más cercanos en el espacio de los permutantes. Por lo tanto, primero se obtienen aquellos elementos potenciales a tener

las menores distancias con la consulta y después se obtienen las distancias con los vectores originales de las cuales se garantizará encontrar un número  $K$  de vecinos más cercanos.

La distancia utilizada  $S_p(\Pi_u, \Pi_p)$  para comparar dos permutaciones requiere que para cada  $p_i \in \mathbf{P}$  se compute su posición en  $\Pi_u$  y  $\Pi_q$ , y se obtenga  $\Pi_u^{-1}$  y  $\Pi_q^{-1}$ . Una vez conocidas las posiciones se obtiene la suma de los cuadrados de las diferencias en las posiciones, tal como lo muestra la ecuación 14.

$$S_p(\Pi_u, \Pi_p) = \sum_{1 \leq i \leq k} (\Pi_u^{-1}(i) - \Pi_q^{-1}(i))^2 \quad (14)$$

Un ejemplo de la diferencia entre dos permutaciones es el siguiente: Sea  $\Pi_q = 6, 2, 3, 1, 4, 5$  la permutación de la consulta y  $\Pi_u = 3, 6, 2, 1, 5, 4$  la permutación de un elemento  $u$ , las diferencias entre las posiciones de los elementos son: 4-4 para  $p_1$ , 2-3 para  $p_2$ , 3-1 para  $p_3$ , 5-6 para  $p_4$ , 6-5 para  $p_5$  y 1-2 para  $p_6$ . Donde la suma de los cuadrados de dicha diferencia es  $S_p(\Pi_q, \Pi_u) = 8$ .

Para facilitar la obtención de la distancia, lo que se almacena de cada punto son las permutaciones inversas, las cuales consisten en las posiciones de los elementos permutantes con respecto al índice del elemento permutante, por ejemplo, a partir de  $\Pi_q$  se tiene que  $\Pi_q^{-1} = 4, 2, 3, 5, 6, 1$  ya que  $p_1$  se encuentra en la posición 4,  $p_2$  en la segunda,  $p_3$  en la tercera,  $p_4$  en la quinta,  $p_5$  en la cuarta y  $p_6$  en la primera. De la misma manera  $\Pi_u^{-1} = 4, 3, 1, 6, 5, 2$ .

Ya con las permutaciones inversas, la distancia se obtiene sacando las diferencias de los elementos de  $\Pi_q^{-1}$  y  $\Pi_u^{-1}$  que en este caso serían (4-4), (2-3), (3-1), (5-6), (6-5) y (1-2) las cuales coinciden con la diferencia entre las posiciones de los elementos en  $\Pi_q$  y  $\Pi_u$  mostradas arriba.

El uso de los elementos permutantes, así como de las bolsas de palabras e incluso cualquier técnica que reduce la información, provoca que se pierdan datos, lo cual puede afectar en el desempeño del sistema. Por lo que es necesario realizar una evaluación sobre el compromiso entre la efectividad del sistema y las ventajas de reducir la información utilizada. A continuación se explica sobre la evaluación en los sistemas de recuperación.

## **II.4 Evaluación de sistemas de recuperación**

Mediante una evaluación de los sistemas de recuperación, es posible detectar aquellos que arrojan mejor efectividad en los resultados. Si los sistemas son construidos utilizando diferentes técnicas de extracción de características, reducción de información o indexación, entonces es necesario evaluar el desempeño de dichas técnicas.

El progreso en el desarrollo de sistemas de recuperación de información basada en contenido visual, así como en cualquier campo, depende de la habilidad de evaluar los resultados de los métodos de recuperación, de tal manera que no se dependa de una opinión subjetiva del evaluador (Jermyn *et al.*, 2002).

Un concepto importante el cual es usado en evaluación, es el de relevancia. Este concepto tiene diversas implicaciones que complican su definición (Saracevic, 2007). En el contexto de recuperación de información y como es utilizada en esta tesis, un documento recuperado es relevante si cumple con las necesidades del usuario o si es recuperado de acuerdo a lo que es correcto recuperar según cierta información necesaria (Baeza-Yates y Ribeiro-Neto, 2004).

Para evaluar un sistema de recuperación de información, se realizan consultas de las cuales hay que conocer cuales elementos de la base de datos son relevantes, es decir, cuales son aquellos que se deben recuperar. En el caso de las consultas por imagen ejemplo, los elementos relevantes son todos aquellos que corresponden a la misma imagen sujeta a alguna modificación como se muestra en la sección II.1.3.

En el caso de las consultas por similitud de imágenes, la definición del conjunto de elementos relevantes se vuelve más complicada, aunque se han propuesto diversas técnicas, por ejemplo, Black *et al.* (2002) y Jain y Vailaya (1998) comparan resultados arrojados por un sistema de recuperación automático con aquellos elementos que serían recuperados por humanos.

Un recurso utilizado para definir un conjunto de elementos relevantes, es el realizar transformaciones en las imágenes de la base de datos. Mediante este recurso, se tiene que las imágenes correctas a recuperar según una consulta dada, son todas las transformaciones a las cuales pertenece la consulta. Se pueden ver ejemplos en Jain *et al.* (2007) y Acton y Rossi (2008).

Para poder realizar verdaderas comparaciones entre diferentes sistemas de recuperación propuestos por distintos grupos de investigación, no solo es necesario evaluar los sistemas con las mismas mediciones, sino también el utilizar la misma base de datos. Ante tal problema, un grupo de investigadores tuvo una iniciativa que surgió de la conferencia “*The Internet Imaging Conference 2000*” que consiste en proporcionar un marco de referencia común para la evaluación de los sistemas de recuperación de información basada en contenido visual, además de proporcionar una base de datos de imágenes abierta que contenga anotaciones y se encuentren definidos los elementos relevantes para las consultas (Marchand-Maillet y Beretta, 2005).

En cuanto a las evaluaciones realizadas para los sistemas biométricos, se tienen aquellas medidas sugeridas por el grupo “*The Biometric Consortium*” (Dunn y Podio, 2005) y las que realiza la competencia “*International Fingerprint Verification Competition*” (Maio, *et al.*, 2006). El grupo biométrico sirve como un punto focal para la investigación, desarrollo, pruebas, evaluación y aplicaciones de las tecnológicas de identificación/verificación basadas en biometría, este consorcio organiza conferencias, realiza actividades y se pueden encontrar recursos en su sitio web: [www.biometrics.org](http://www.biometrics.org). La competencia de verificación de las huellas digitales recibe trabajos de los participantes y realiza evaluaciones para

determinar cuál es el ganador, más información sobre el concurso y las evaluaciones que realiza se pueden encontrar en su página de internet: <http://bias.csr.unibo.it/fvc2004>.

Las métricas a utilizar para evaluar el desempeño por los sistemas biométricas según el consorcio biométrico son: Tasa de falsos positivos, tasa de falsos negativos, curva ROC (*Receiver Operating Characteristic*) y curva CMC (*Cumulative Match Characteristic*).

En (Jain *et al.*, 2007) y (Lee *et al.*, 2008) se realiza la evaluación del sistema de recuperación utilizando el tiempo de recuperación, así como el uso de las gráficas de precisión-exhaustividad y CMC las cuales se explican a continuación ya que fueron elegidas para la evaluación del presente trabajo.

#### *II.4.1 Gráfica precisión-exhaustividad*

Precisión es la habilidad de recuperar al inicio los documentos que son más relevantes.

$$\text{Precisión} = \frac{\text{No.de documentos relevantes recuperados}}{\text{No.Total de elementos recuperados}} \quad (15)$$

La exhaustividad (*recall*) es la habilidad del sistema de encontrar todos los elementos relevantes en la base de datos.

$$\text{Exhaustividad} = \frac{\text{No.de documentos relevantes recuperados}}{\text{No.Total de documentos relevantes en la colección}} \quad (16)$$

Una gráfica de precisión-exhaustividad, incluye la precisión en la ordenada y exhaustividad en la abscisa. Mediante esta gráfica es posible observar el compromiso existente entre precisión y exhaustividad para recuperar todos los elementos relevantes en las primeras posiciones del resultado de la consulta (Manning, Raghavan, & Schütze, 2008). Ver Figura 22.

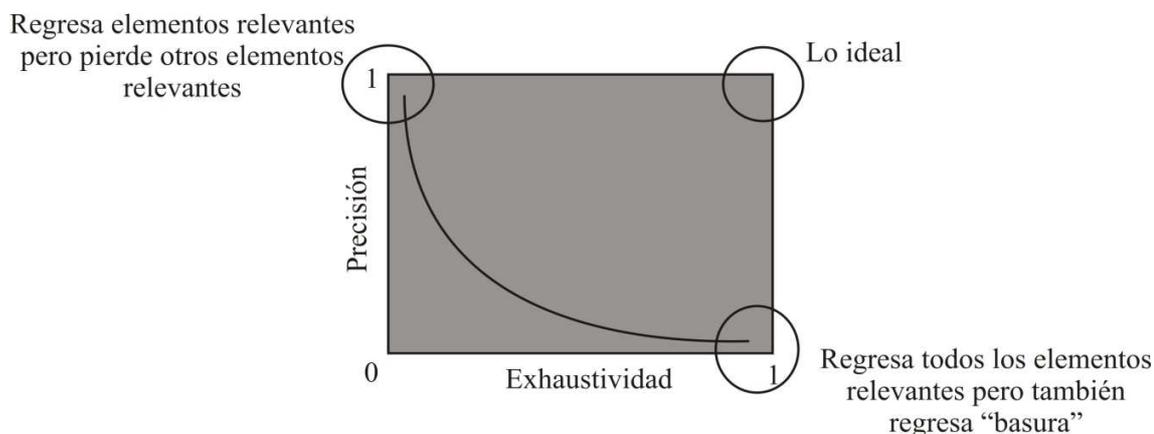


Figura 22. Compromiso entre precisión y exhaustividad.

#### II.4.2 Curva "Cumulative Matching Characteristic" (CMC)

La recuperación de una consulta tiene rango  $k$  si la respuesta correcta es recuperada en la posición  $k$ , por ejemplo, la recuperación tiene rango 5 si la respuesta correcta se encuentra en la quinta posición. La *tasa o probabilidad de identificación* en el rango  $k$  o precisión acumulada es la fracción de las respuestas de varias consultas que han sido encontradas en la quinta posición o antes. El desempeño de identificación se grafica en la curva "Cumulative Matching Characteristic".

La curva CMC es un método para mostrar la exactitud del desempeño de un sistema biométrico operando en una tarea de identificación en un conjunto cerrado (que el sujeto de la consulta este dentro de la base de datos) (Moon y Phillips, 2001). La curva CMC muestra que tan frecuentemente se encuentra a un individuo en ciertos rangos (1, 5, 10, 100, etc.), después de aplicar una medida de similitud a características extraídas de las imágenes. La curva CMC comprara el rango (1, 5, 10, 100, etc.) (abscisa) contra la tasa de identificación (ordenada) tal como se muestra en la Figura 23.

En la parte de arriba de la Figura 23 se muestran ejemplos de resultados relevantes en cada posición (de la 1 a la 13). En este ejemplo se considera que existe sólo un documento relevante para cada consulta (si existieran más relevantes se considera solo el primero). La probabilidad de identificación o precisión acumulada indica cuantos documentos se recuperaron hasta la posición indicada sobre la cantidad de documentos relevantes que debieron ser recuperados.

Consulta	Rango												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	O	X	X	X	X	X	X	X	X	X	X	X	X
2	X	O	X	X	X	X	X	X	X	X	X	X	X
3	O	X	X	X	X	X	X	X	X	X	X	X	X
4	O	X	X	X	X	X	X	X	X	X	X	X	X
5	X	X	O	X	X	X	X	X	X	X	X	X	X
6	X	O	X	X	X	X	X	X	X	X	X	X	X
7	X	O	X	X	X	X	X	X	X	X	X	X	X
8	X	X	X	O	X	X	X	X	X	X	X	X	X
9	X	X	X	X	X	X	X	X	O	X	X	X	X
10	O	X	X	X	X	X	X	X	X	X	X	X	X
Probabilidad de identificación	.4	.7	.8	.9	.9	.9	.9	.9	1.	1.	1.	1.	1.

O – Sujeto correcto

X – Sujeto incorrecto

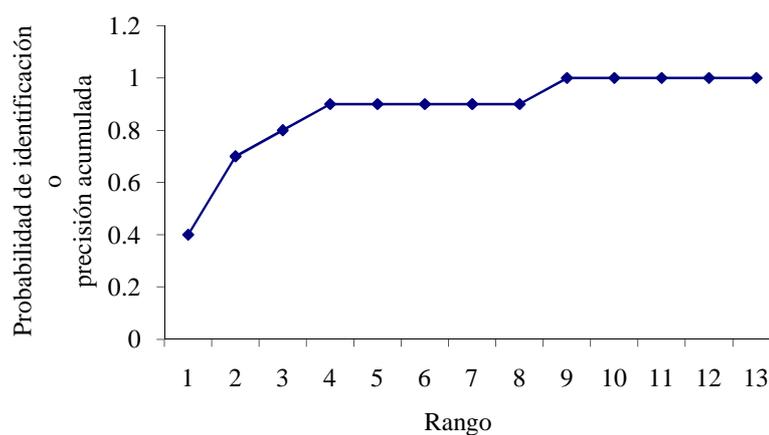


Figura 23. (Arriba) Ejemplo de resultados de recuperación. (Abajo) Ejemplo de la curva CMC.

Como se menciona, las evaluaciones se llevan a cabo mediante consultas sobre una base de datos, en donde es importante conocer el conjunto de elementos relevantes para cada consulta, en el siguiente capítulo se presenta sobre la base de datos utilizada en la presente tesis, así como las consultas elegidas y sus elementos relevantes.

## Capítulo III

### III Construcción de la base de datos

#### III.1 Obtención de imágenes de tatuajes para base de datos

En el presente trabajo, se tiene el objetivo de detectar imágenes de tatuajes en bases de datos de agencias policíacas, por lo que se utilizan imágenes obtenidas bajo circunstancias reales, la cuales generalmente son de calidad muy distinta de aquellos tatuajes que son denominados como tatuajes *recreacionales*. En la Tabla I se muestran ejemplos que comparan estos dos tipos de tatuajes.

Tabla I. Comparación entre tatuajes recreacionales y tatuajes reales.

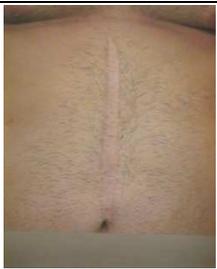
Tatuajes recreacionales	Tatuajes en condiciones reales de agencias policíacas
	
	
	

### III.2 Construcción base de datos principal

La base de datos que se utilizó para realizar los experimentos, consta de imágenes que fueron proporcionadas por el departamento de policía de Los Angeles, California. Se facilitaron un total de 2098 imágenes, de las cuales 1000 de ellas son de 800 x 1000 pixeles de resolución y 1098 son de 400 x 600 pixeles.

1960 imágenes incluyen por lo menos un tatuaje, mientras que 138 incluyen tatuajes que no se visualizan correctamente o incluyen marcas y/o cicatrices. En la Tabla II, se muestran ejemplos de imágenes con tatuajes e imágenes con marcas y cicatrices.

Tabla II. Ejemplos de tatuajes, marcas y cicatrices.

Tatuajes	Tatuajes no visualizados correctamente	Marcas	Cicatrices
			
			

Para poder realizar los experimentos con los métodos mencionados, fue necesario trabajar preferentemente con el área correspondiente a los tatuajes individuales, por este motivo fue necesario hacer un recorte en las imágenes de la base de datos. A partir de dicho recorte se obtuvo que de las 1960 imágenes originales, se extrajeran 2818 imágenes de tatuajes individuales ya que existen fotografías con más de un tatuaje. Ver tabla III.

Tabla III. Imágenes recortadas a partir de imágenes originales.

IMÁGENES ORIGINALES	IMÁGENES RECORTADAS
	
Resolución: 800x1000 pixeles	Resolución: 360x411 pixeles
	
Resolución: 800x1000	Resolución: 604x223 y 224x107 pixeles
	
Resolución: 800x1000	Resolución: 231x692 , 176x171 y 231x692 pixeles

Entre las ventajas que brinda el recortar las imágenes se puede mencionar el eliminar el ruido ocasionado por elementos que no forman parte del tatuaje, tales como fondos, ropa, cabellos, artefactos, sombras e incluso partes del cuerpo. Así como la identificación de cada tatuaje por separado lo cual favorece en la ejecución de los experimentos.

De las 2818 imágenes recortadas, se hizo una clasificación por medio de las clases establecidas por el estándar ANSI/NIST ITL 1-2000 (ITL, 2009), En la Tabla IV se muestran las cantidades que se asignaron a cada clase. La etiqueta TEXTO no forma parte de las clases del estándar ANSI/NIST ITL 1-2000 sino que forma parte de una subclase de la clase OTHER, sin embargo se asignó esta etiqueta debido a la gran cantidad de imágenes que cayeron en dicha clasificación.

Tabla IV. Clases definidas por el estándar ANSI/NIST ITL 1-2000.

CLASE	NUMERO DE IMÁGENES
ABSTRACT	327
ANIMAL	289
FLAG	9
HUMAN	186
OBJECT	316
OTHER	564
PLANT	138
SYMBOL	160
TEXT	829
TOTAL	2818

Todas las imágenes recortadas pasaron por una etapa de procesamiento, esta etapa de procesamiento consiste en una ecualización adaptiva limitada por contraste (CLAHE por sus siglas en inglés) la cual fue realizada mediante una función de MATLAB llamada *adapthisteq* la cual hace dicha ecualización. Posteriormente se aplicó umbralización adaptiva local mediante una función en matlab llamada *adaptivethreshold* la cual fue descargada de la web (Xiong, 2006) en donde se utilizó un tamaño de ventana (*ws*) de 100

y una media de umbralización ( $C$ ) de 0.1, ya que con dichos parámetros se obtuvo mejor descripción visual en las imágenes resultantes. Aparte del procesamiento automático aplicando las funciones mencionadas, también se aplicó un procesamiento de forma manual utilizando las aplicaciones *Microsoft paint* y *Corel photo-paint*, el cual consistió en un borrado de los artefactos introductores de ruido a las imágenes de tatuajes umbralizadas. Todo el procesamiento mencionado se puede apreciar en la Figura 24.

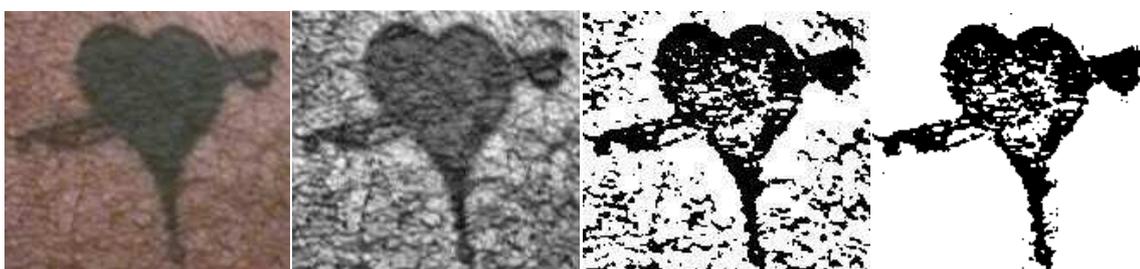


Figura 24.a) Imagen original, b) imagen ecualizada, c) imagen umbralizada d) imagen intervenida manualmente.

A partir de este procesamiento fueron detectados los elementos considerados como artefactos que introducen ruido a los tatuajes los cuales son los siguientes:

- Cabello
- Cambios de iluminación
- Venas resaltadas
- Piel porosa
- Lunares
- Obstrucción con ropa y joyas
- El perfil de las partes del cuerpo
- Dedos
- Vello

La importancia del conocer estos elementos radica al momento de la toma de fotografías en las agencias policíacas, ya que se pueden establecer acciones que disminuyan el efecto

causado para los diferentes tipos de artefactos, lo cual incrementa la eficiencia del sistema utilizado para la recuperación de información.

Una vez construida esta base de datos, se eligieron imágenes de consulta para de esta manera poder realizar los experimentos. Los experimentos relacionados con las consultas por imagen ejemplo, constan de dos etapas; las consultas con las imágenes en su estado original y las consultas con imágenes transformadas las cuales se explican a continuación.

### *III.2.1 Consulta con imágenes originales*

Debido a que se realizan consultas por imagen ejemplo, se requiere que por cada consulta, haya al menos una imagen en la base de datos que cuente con las siguientes características las cuales hacen que dichas imágenes sean relevantes:

- Una imagen exacta a la consulta
- Parte de la imagen de consulta
- Una versión alterada geoméricamente de la imagen de consulta
- Una versión comprimida de la consulta

Mediante inspección visual fueron detectadas 130 imágenes que tenían al menos una imagen repetida con las características anteriores. Tales imágenes se pueden observar en el anexo junto con aquellas imágenes que se espera sean recuperadas al realizar la consulta. El resto de las imágenes son únicas por lo que se complica el conocer cuales imágenes del resto de la base de datos son relevantes, por lo tanto no se utilizaron como consultas.

Una vez formada la base de datos y elegidas imágenes de consulta, estas se utilizaron para realizar experimentos cuyos resultados se presentan en la sección IV.3. Cabe mencionar que con las mismas imágenes se generaron tres tipos de bases de datos, que contienen imágenes en escala de grises, ecualizadas y umbralizadas.

Con el objetivo de tener una mayor cantidad de consultas y poder obtener distintos tipos de resultados, se construyó otra base de datos la cual se menciona a continuación.

### **III.3 Construcción de base de datos y consultas con imágenes transformadas**

Para poder obtener gráficas de precisión y exhaustividad, es necesario conocer un determinado número de imágenes relevantes (ver sección II.4), en el caso de la base de datos principal mostrada en la sección anterior, las imágenes se repiten a lo más 4 veces. Por lo que fue necesario crear otra base de datos donde se tenga un mayor número de imágenes relevantes conocidas.

Para la construcción de esta nueva base de datos, se eligieron 100 imágenes de la base de datos anterior, buscando que fueran de buena calidad y los tatuajes tuvieran los límites definidos. A partir de estas imágenes iniciales, se obtuvieron 2100 imágenes, las cuales surgieron de aplicar 21 transformaciones diferentes a cada una.

Las transformaciones que fueron aplicadas (Ver Figura 25) son las siguientes.:

- Rotación con 5,10,25,-10,-25 grados
- Escalamiento horizontal con relación de 1 - 0.7
- Escalamiento horizontal con relación de 1 - 1.3
- Escalamiento vertical con relación de 1 - 0.7
- Escalamiento vertical con relación de 1 - 1.3
- Escalamiento horizontal con relación 1 - 0.8 y vertical 1 - 1.1
- Cambio de perspectiva horizontal, con  $x = w + \alpha z$  y  $y = z$  donde  $\alpha=0.15$
- Cambio de perspectiva horizontal, con  $x = w + \alpha z$  y  $y = z$  y  $\alpha=-0.15$
- Cambio de perspectiva vertical:  $x = w$  y  $y = \beta w + z$  donde  $\beta=0.15$
- Cambio de perspectiva vertical:  $x = w$  y  $y = \beta w + z$  donde  $\beta=-0.15$

- Modificación del canal R (RGB) multiplicando cada pixel en este canal por 0.5
- Modificación del canal G (RGB) multiplicando cada pixel en este canal por 0.5
- Modificación del canal B (RGB) multiplicando cada pixel en este canal por 0.5
- Aplicación de difuminado con una ventana gaussiana de 3x3 y sigma de valor 3
- Aplicación de difuminado con una ventana gaussiana de 4x4 y sigma de valor 4
- Aplicación de ruido sal y pimienta con densidad de 0.01
- Aplicación de iluminación con una ventana gaussiana de radio 140 con una mínima intensidad de 0.4 y una máxima de 1

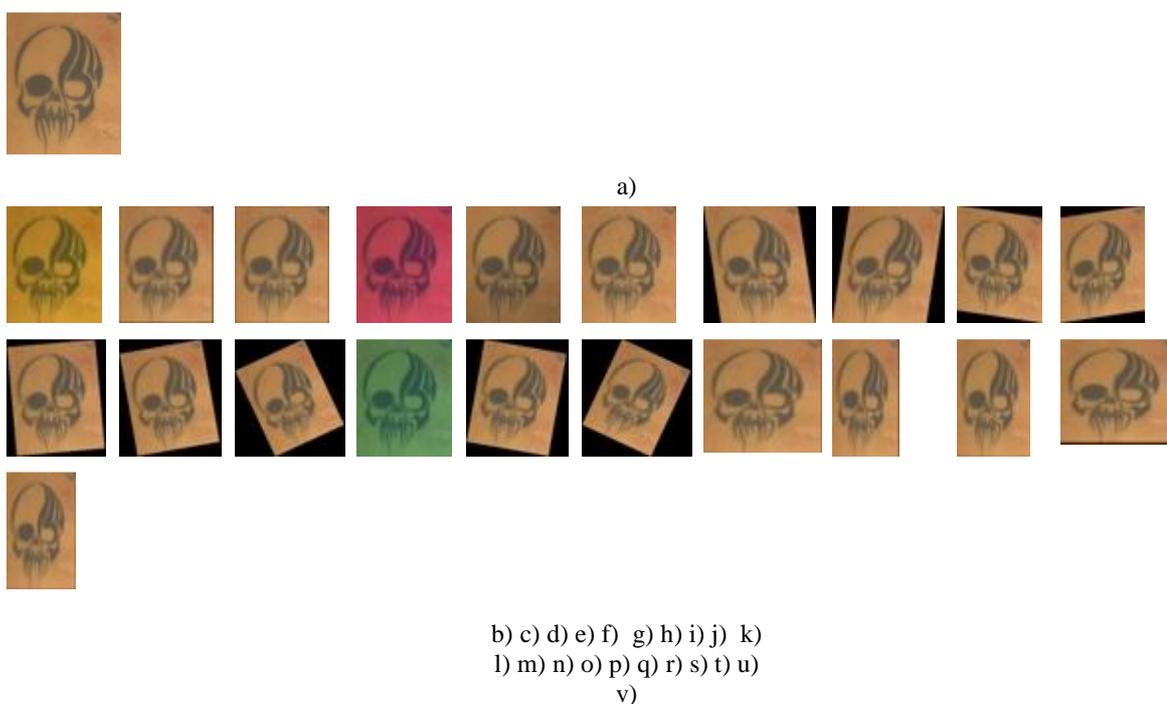


Figura 25. a) original b) e) o) modificación de canal de color, c) d) difuminado f) iluminación g) ruido h) i) j) k) cambio de perspectiva, l) m) n) p) q) rotación r) s) t) u) v) escalamiento.

Una vez realizadas las transformaciones se procedió a aplicar ecualización adaptiva a todas las imágenes, ya que como se menciona más adelante, las consultas se realizan mediante las imágenes ecualizadas.

Las bases de datos mencionadas se utilizan en los experimentos con imágenes ejemplo, sin embargo, como se comenta en la sección II.2.4, los experimentos para las consultas por imágenes similares son de otra naturaleza y por lo tanto fue necesario construir otra base de datos.

### **III.4 Construcción de base de datos para consultas por imágenes similares**

Para construir la base de datos para los experimentos con consultas usando imágenes similares, se eligieron 1100 imágenes aleatorias a partir de la base de datos principal. Debido a que la técnica utilizada no posee invariancia a la rotación, se procedió a aplicar un procesamiento a las imágenes el cual consiste en orientarlas en base a su orientación predominante para de esta manera observar si la técnica resulta invariante a esta transformación.

La orientación predominante se obtiene a partir de la transformada de Hough (Duda y Hart, 1972), la cual fue aplicada a las imágenes por medio de una función de MATLAB llamada *houghpeaks*. Una vez obtenida la orientación predominante de cada imagen, se guarda una copia de la imagen con la orientación obtenida y también se guarda otra versión correspondiente a 180 grados de la orientación predominante obtenida. Ver Figura 26.

Debido a las características propias de la técnica utilizada para este tipo de consultas, los fondos negros que se aprecian en la Figura 26, causados por el efecto de la rotación, no son deseables ya que interfieren con los resultados de la recuperación. Por lo tanto fue necesario procesar las imágenes de la base de datos para remover dichos fondos.

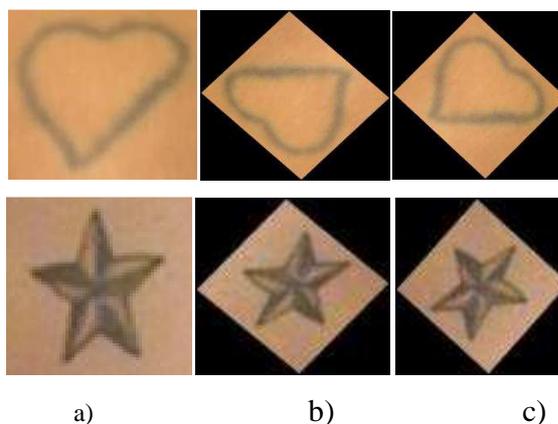


Figura 26, a) Imágenes originales, b) Imagen orientada en base a su orientación predominante, c) Imagen orientada 180 grados en base a su orientación predominante.

Para resolver el problema de los fondos, se utilizaron las imágenes umbralizadas con intervención manual, las cuales fueron rotadas en base a la orientación predominante como ya se menciona, ver Figura 27. Al utilizar estas imágenes se pierde la característica de color, la cual es soportada por esta técnica de recuperación, sin embargo, la naturaleza de las imágenes de tatuajes reales es que no poseen diversidad en color, si no que este se presenta generalmente de manera uniforme en todas las imágenes donde domina el color negro y verde oscuro correspondientes a las tintas utilizadas para su creación.

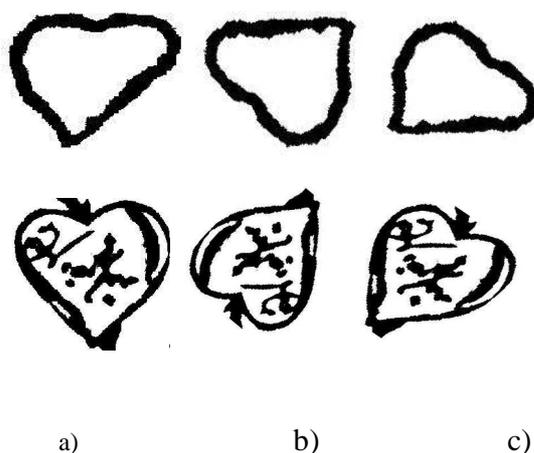


Figura 27, a) Imágenes umbralizadas originales, b) Imagen umbralizada orientada en base a su orientación predominante, c) Imagen umbralizada orientada 180 grados en base a su orientación predominante.

La elección de las consultas para este tipo de experimentos consistió en encontrar imágenes parecidas visualmente. Inicialmente se buscaron imágenes en la base de datos que correspondieran a etiquetas de la misma clase, posteriormente se agruparon imágenes de contenidos semánticos similares, donde en la Tabla V se muestran las cantidades de imágenes encontradas.

Tabla V. Cantidad de imágenes con contenidos similares en la base de datos.

Contenido imagen	Cantidad de imágenes en base de datos	Contenido imagen	Cantidad de imágenes en base de datos
Cruces	82	Jesús	10
Posición de rezo	10	Caras de calavera	31
Corazones	45	Rosas	33
Símbolo de triángulo	14	Símbolo yin yang	6
Estrellas	29	Conejo Playboy	5

Las imágenes correspondientes al contenido de posición de rezo, Jesús, caras de calavera, rosas, símbolo de yin yang y conejo Playboy no poseen similitud visual, por lo que no fueron elegidas como consultas. Las consultas elegidas se presentan en la Figura 28.



Figura 28. Imágenes elegidas para consultas por imágenes similares.

En el siguiente capítulo se presentan los experimentos y resultados obtenidos donde se utilizan las consultas elegidas y las bases de datos construidas como se explica en este capítulo.

## Capítulo IV

---

### IV Experimentos y resultados

---

#### IV.1 Introducción de experimentos realizados

Se realizaron diferentes tipos de experimentos en donde se usaron distintos conjuntos de bases de datos y diferentes métodos. La extracción de características para consultas por imágenes ejemplo, se llevo a cabo mediante el algoritmo SIFT, del cual se obtuvo una versión en MATLAB de internet (Vedaldi, 2006). A partir de las características extraídas, se implementaron diversas formas de representar y recuperar los datos basádonos en las ideas propuestas en el capítulo II.

La extracción de características y recuperación para consultas por imágenes similares se llevó a cabo mediante una aplicación existente llamada Imgseek (Niederberger C., 2007) la cual fue implementada a partir del trabajo de Jacobs, Finkelstein, y Salesin, (1995)

A continuación se mencionan los esquemas de recuperación con los cuales se realizaron experimentos para posteriormente explicar en qué consistió cada experimento y mostrar los resultados obtenidos.

Experimentos utilizando SIFT

- Utilizando medida de similitud de número de igualaciones por par de imágenes.
- Utilizando medida de similitud de número de igualaciones en base de datos completa.
- Uso de bolsa de palabras.

Experimentos utilizando imgseek

## IV.2 Experimentos utilizando SIFT

Como se menciona, la extracción de los descriptores SIFT se llevó a cabo mediante un algoritmo previamente codificado (Vedaldi, 2006). A partir de esta información, se utilizaron diferentes esquemas de búsqueda, los cuales se mencionan en esta sección.

### IV.2.1 Utilizando medida de similitud de número de igualaciones por par de imágenes

Este esquema de recuperación fue el propuesto por Lee *et al.* (2008) para la recuperación de imágenes de tatuajes. Este consiste en que la medida de similitud a utilizar, es el número de igualaciones exitosas entre la consulta y cada imagen de la base de datos. En donde un mayor número de igualaciones exitosas representa mayor similitud, tal como se explica en la sección II.2.3.1.5. Ver Figura 29.

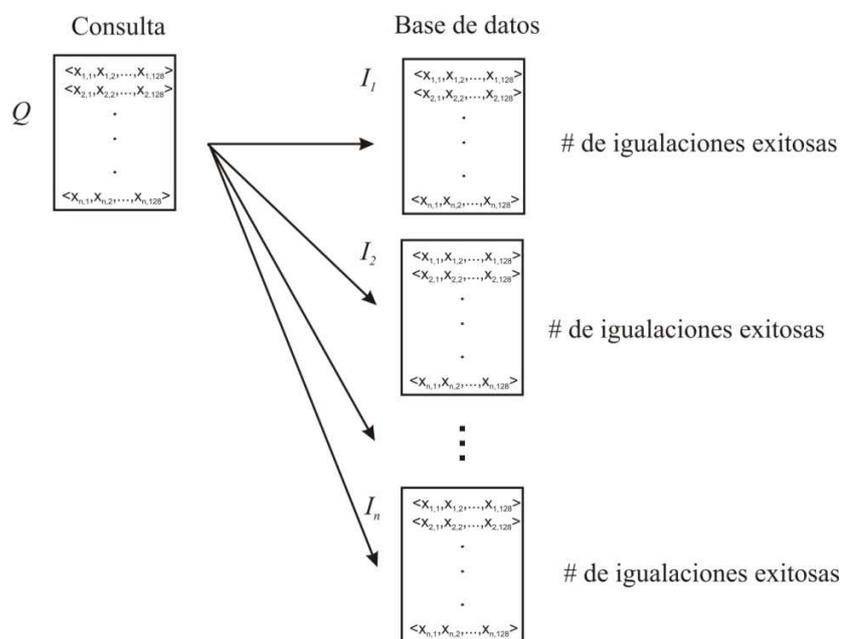


Figura 29. Esquema de búsqueda con medida de similitud de número de igualaciones exitosas por par de imágenes.

Este experimento se realizó sobre tres bases de datos diferentes, estas bases de datos incluyen a las imágenes originales en escala de grises, a las imágenes con ecualización adaptiva y las imágenes umbralizadas con intervención manual las cuales se pueden observar en la Figura 24.

Al momento de realizar la consulta, la imagen también fue tratada de la misma manera que los elementos de la base de datos de la cual se espera recuperar. A continuación se presentan los resultados obtenidos para cada experimento.

#### IV.2.1.1 Imágenes en escala de grises

El experimento fue realizado con 130 consultas sobre 2773 imágenes de las 2818 recortadas, ya que una vez utilizado el algoritmo SIFT, 45 imágenes resultaron con solo 1 o 2 descriptores debido a la falta de contraste o detalle en las imágenes de los tatuajes, lo que hace que no puedan utilizarse en la búsqueda.

Tabla VI. Cantidad de imágenes y precisión acumulada obtenidas en las posiciones 1-20 en las consultas realizadas con las imágenes en escala de grises.

<b>Rango</b>	<b>Cantidad</b>	<b>Precisión acumulada</b>	<b>Rango</b>	<b>Cantidad</b>	<b>Precisión acumulada</b>
<b>1</b>	62	0.476923	<b>11</b>	0	0.9
<b>2</b>	35	0.746154	<b>12</b>	0	0.9
<b>3</b>	8	0.807692	<b>13</b>	0	0.9
<b>4</b>	4	0.838462	<b>14</b>	0	0.9
<b>5</b>	1	0.846154	<b>15</b>	2	0.915385
<b>6</b>	4	0.876923	<b>16</b>	0	0.915385
<b>7</b>	1	0.884615	<b>17</b>	0	0.915385
<b>8</b>	1	0.892308	<b>18</b>	1	0.923077
<b>9</b>	0	0.892308	<b>19</b>	0	0.923077
<b>10</b>	1	0.9	<b>20</b>	1	0.930769

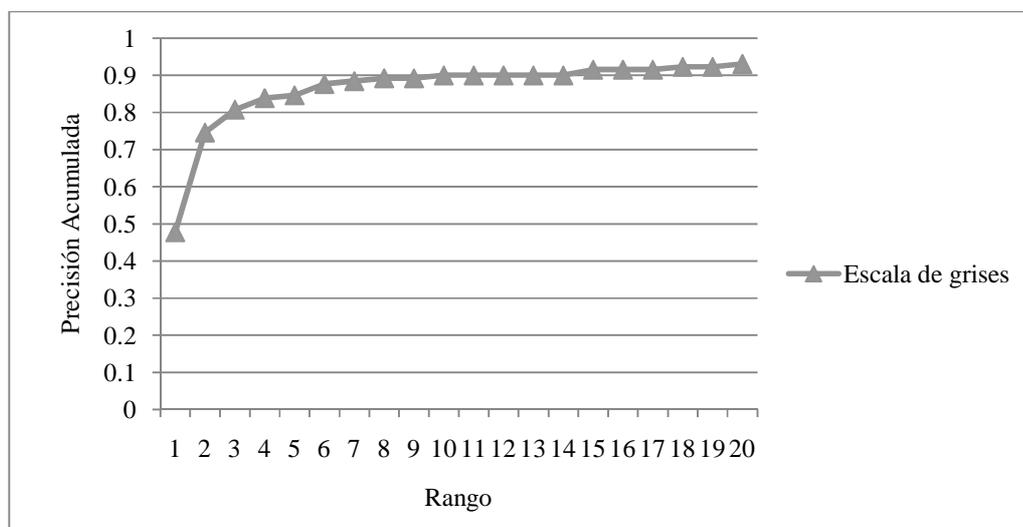


Figura 30. Curva CMC con base de datos de imágenes en escala de grises.

En la Tabla VI y la Figura 30 se muestran los resultados, usando 130 consultas, de las cuales 9 no se encontraron en las primeras 20 posiciones.

#### IV.2.1.2 Imágenes con ecualización adaptiva

El experimento fue realizado con 130 consultas sobre 2803 imágenes de las 2818 recortadas, ya que una vez utilizado el algoritmo SIFT, 15 imágenes resultaron con solo 1 o 2 descriptores debido a la falta de contraste o detalle en las imágenes de los tatuajes, lo que hace que no puedan utilizarse en la búsqueda.

La Tabla 7 y la Figura 31 muestran los resultados usando 130 consultas, de las cuales 5 no se encontraron en las primeras 20 posiciones. Estos resultados resultan mejores que los mostrados en la Tabla VII y la Figura 31 ya que se encuentran más imágenes en las primeras posiciones.

Tabla VII. Cantidad de imágenes y precisión acumulada obtenidas en las posiciones 1-20 en las consultas realizadas con las imágenes con eualización adaptiva.

Rango	Cantidad	Precisión acumulada	Rango	Cantidad	Precisión acumulada
1	113	0.869231	11	0	0.946154
2	2	0.884615	12	0	0.946154
3	3	0.907692	13	0	0.946154
4	3	0.930769	14	0	0.946154
5	0	0.930769	15	1	0.953846
6	1	0.938462	16	1	0.961538
7	0	0.938462	17	0	0.961538
8	1	0.946154	18	0	0.961538
9	0	0.946154	19	0	0.961538
10	0	0.946154	20	0	0.961538

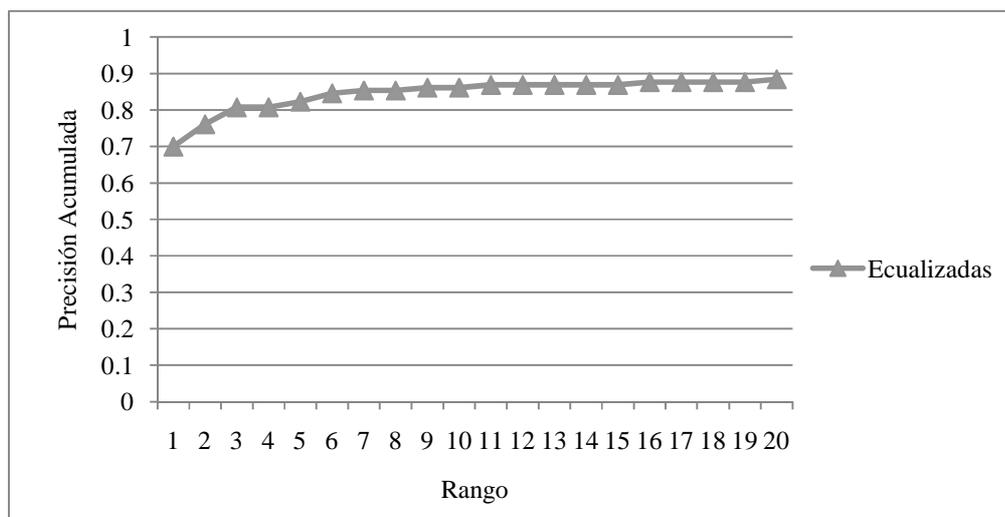


Figura 31. Curva CMC con base de datos de imágenes con eualización adaptiva.

### IV.2.1.3 Imágenes con umbralización adaptiva

El experimento fue realizado con 130 consultas sobre 2815 imágenes de las 2818 recortadas, ya que una vez utilizado el algoritmo SIFT, 3 imágenes resultaron con solo 1 o 2 descriptores debido a las características de las imágenes, lo que hace que no puedan utilizarse en la búsqueda.

La Tabla VIII y la Figura 32 muestran los resultados usando 130 consultas, de las cuales 15 no se encontraron en las primeras 20 posiciones. Estos resultados no resultan mejores que en los experimentos anteriores a pesar de que es necesario un mayor pre-procesamiento en la base de datos utilizada.

Tabla VIII. Cantidad de imágenes y precisión acumulada obtenidas en las posiciones 1-20 en las consultas realizadas con las imágenes con umbralización adaptiva.

<b>Rango</b>	<b>Cantidad</b>	<b>Precisión acumulada</b>	<b>Rango</b>	<b>Cantidad</b>	<b>Precisión acumulada</b>
<b>1</b>	91	0.7	<b>11</b>	1	0.869231
<b>2</b>	8	0.761538	<b>12</b>	0	0.869231
<b>3</b>	6	0.807692	<b>13</b>	0	0.869231
<b>4</b>	0	0.807692	<b>14</b>	0	0.869231
<b>5</b>	2	0.823077	<b>15</b>	0	0.869231
<b>6</b>	3	0.846154	<b>16</b>	1	0.876923
<b>7</b>	1	0.853846	<b>17</b>	0	0.876923
<b>8</b>	0	0.853846	<b>18</b>	0	0.876923
<b>9</b>	1	0.861538	<b>19</b>	0	0.876923
<b>10</b>	0	0.861538	<b>20</b>	1	0.884615

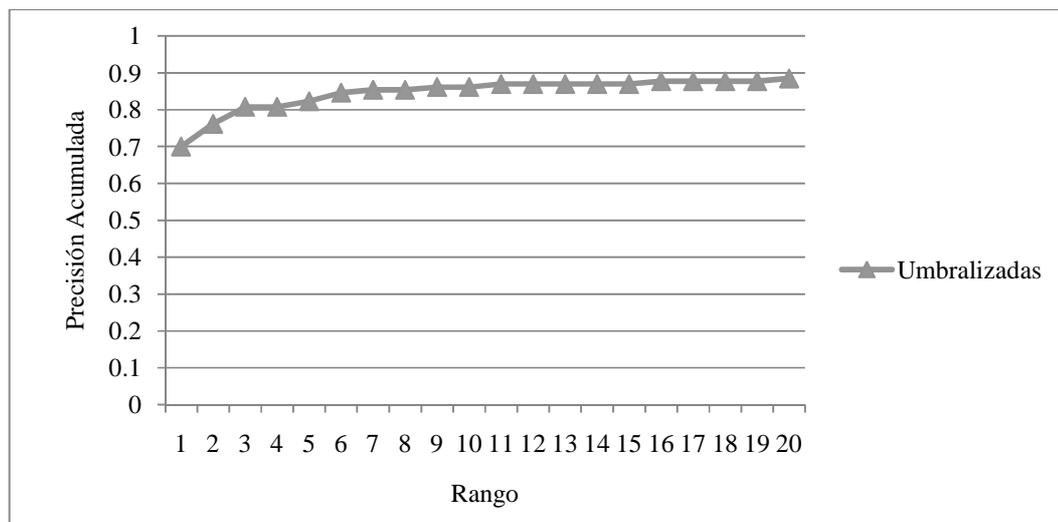


Figura 32. Curva CMC con base de datos de imágenes con umbralización adaptativa.

#### IV.2.1.4 Comparación de resultados con SIFT utilizando medida de similitud de número de igualaciones por par de imágenes

A continuación se presenta en la Figura 33 una gráfica que muestra una comparación de los resultados de la recuperación de imágenes utilizando las bases de datos mencionadas.

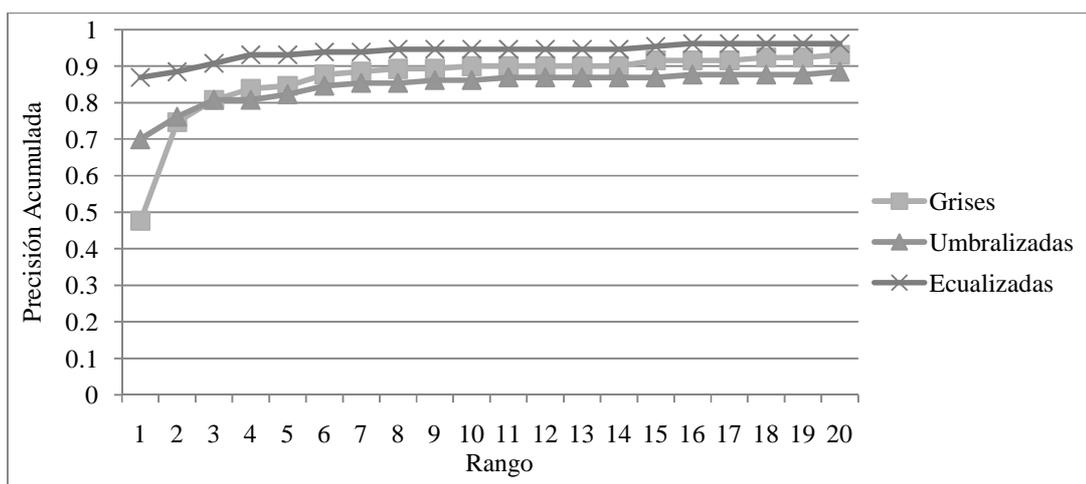


Figura 33. Curva CMC con base de datos de imágenes en escala de grises, con ecualización y umbralización adaptativa.

Como se puede apreciar en la Figura 33, el hecho de umbralizar las imágenes no trae una ventaja consigo, además que esto trae la problemática de que la elección de los parámetros de la umbralización adaptiva (tamaño de ventana y media de umbralización) puede causar que algunas mejoren su apariencia mediante la eliminación de ruido, sin embargo también puede causar que haya descriptores en donde no debería ya que puede resaltar objetos que no pertenecen a los tatuajes o hacer que los tatuajes se dividan en varias regiones.

Para los próximos experimentos del presente trabajo, se utilizará únicamente la base de datos de imágenes ecualizadas, ya que produjo los mejores resultados, además de que el hecho de ecualizar incrementa el contraste, lo que provoca que en algunas imágenes aparezcan descriptores en donde normalmente (para bajos contrastes) no hay. Bakken (2007) explica sobre el efecto de contraste en las imágenes.

Este esquema de búsqueda resulta lento, ya que el tiempo promedio de búsqueda de una consulta es de 476.51 segundos, lo que representa que cada comparación con las imágenes de la base de datos tarda 0.17 segundos. A continuación se presenta otro esquema de búsqueda que utiliza también los descriptores SIFT.

#### *IV.2.2 Utilizando medida de similitud de número de igualaciones en base de datos completa*

Otra forma de realizar búsquedas utilizando la definición de igualaciones exitosas con el algoritmo SIFT mencionada en Lowe (2004) y en la sección II.2.3.1.5, es la de establecer la relación entre el primero y segundo vecino más cercano de cada descriptor de la consulta, con todos los descriptores de todas las imágenes de la base de datos. Donde si la relación es mayor de 0.8 con esos dos vectores provenientes de dos imágenes distintas, entonces se considera una igualación exitosa para aquella imagen que tiene el índice del vector descriptor más cercano. Ver Figura 34.

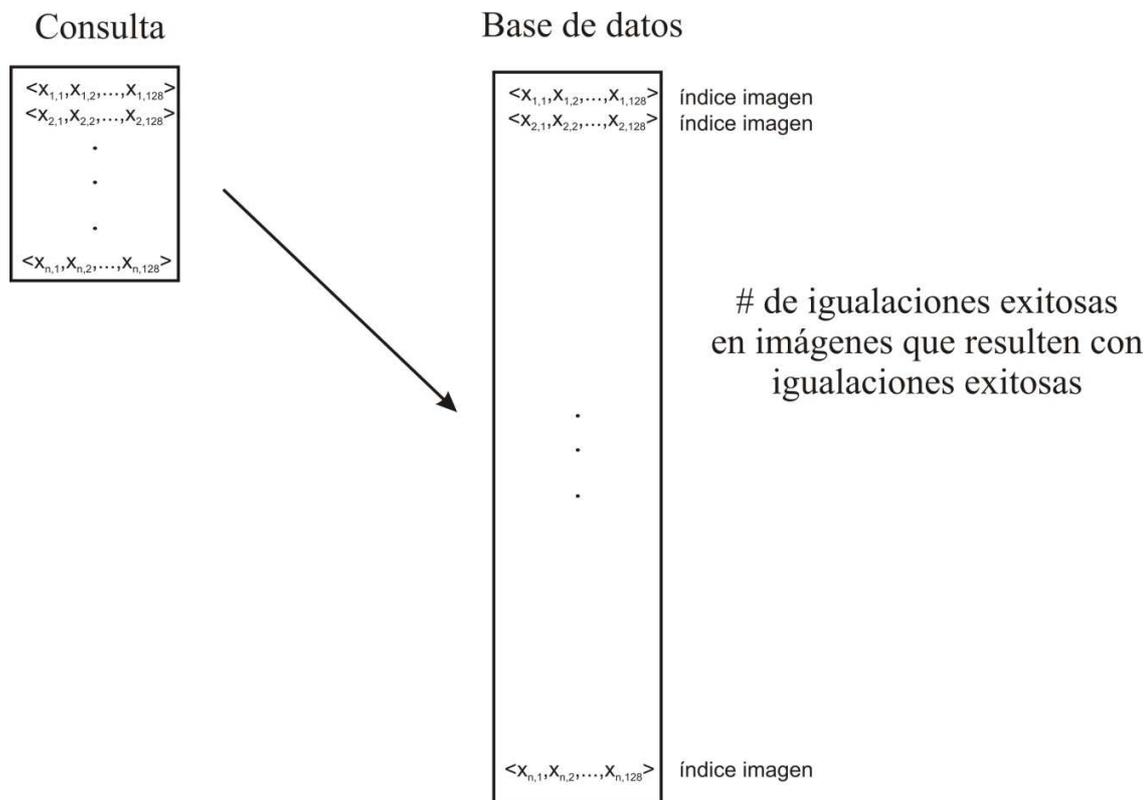


Figura 34. Esquema de recuperación con número de igualaciones exitosas en toda la base de datos.

A continuación, se presentan resultados de experimentos que se realizaron sobre la base de datos con imágenes ecualizadas utilizando este esquema de búsqueda. Para encontrar los vecinos más cercanos se utilizaron dos bibliotecas distintas:

- Biblioteca ANN (Aproximate Nearest Neighbor Library).
- Biblioteca utilizando permutaciones (PNN).

#### IV.2.2.1 Biblioteca ANN (*Aproximate Nearest Neighbor Library*)

La biblioteca ANN está escrita en C++ e implementa estructuras de datos y algoritmos para búsquedas de vecinos más cercanos de forma exacta y aproximada con diferentes dimensiones. La biblioteca utiliza estructuras tales como árboles *K-d tree* o árboles *box-decomposition trees*, con el fin de hacer más eficiente la búsqueda (Mount & Arya, 2006) .

Cuando se desean encontrar de forma exacta los vecinos más cercanos en dimensiones mucho más altas que 8, la tarea de búsqueda se vuelve complicada. Mediante la búsqueda aproximada es posible obtener resultados mucho más rápidos que con búsqueda exacta, pero esto trae consigo errores pequeños. La biblioteca ANN permite establecer un parámetro de límite de error el cual permite un control entre el compromiso de exactitud y tiempo. Además también permite ajustar el parámetro de máximo número de visitas en las hojas del árbol lo cual también ayuda a reducir el tiempo de búsqueda. Este último parámetro corresponde a una forma de búsqueda conocida como *best bin first* propuesta por Beis y Lowe (1997) la cual se explica en el artículo de Lowe (2004) para la realización de comparaciones entre imágenes con características extraídas mediante el algoritmo SIFT.

El siguiente experimento, implica la búsqueda del primero y segundo descriptor vecino más cercano en una base de datos de 2803 imágenes ecualizadas, las cuales tienen en total 2,482,301 descriptores. Para esto se ajustó el parámetro de máximo número de visitas de la librería ANN a 50000. En la Tabla IX y la Figura 35 se muestran los resultados, usando 130 consultas, de las cuales 4 no se encontraron en las primeras 20 posiciones.

Tabla IX. Cantidad de imágenes y precisión acumulada obtenidas en las posiciones 1-20 utilizando la biblioteca ANN.

Rango	Cantidad	Precisión acumulada	Rango	Cantidad	Precisión acumulada
1	122	0.938462	11	0	0.969231
2	4	0.969231	12	0	0.969231
3	0	0.969231	13	0	0.969231
4	0	0.969231	14	0	0.969231
5	0	0.969231	15	0	0.969231
6	0	0.969231	16	0	0.969231
7	0	0.969231	17	0	0.969231
8	0	0.969231	18	0	0.969231
9	0	0.969231	19	0	0.969231
10	0	0.969231	20	0	0.969231

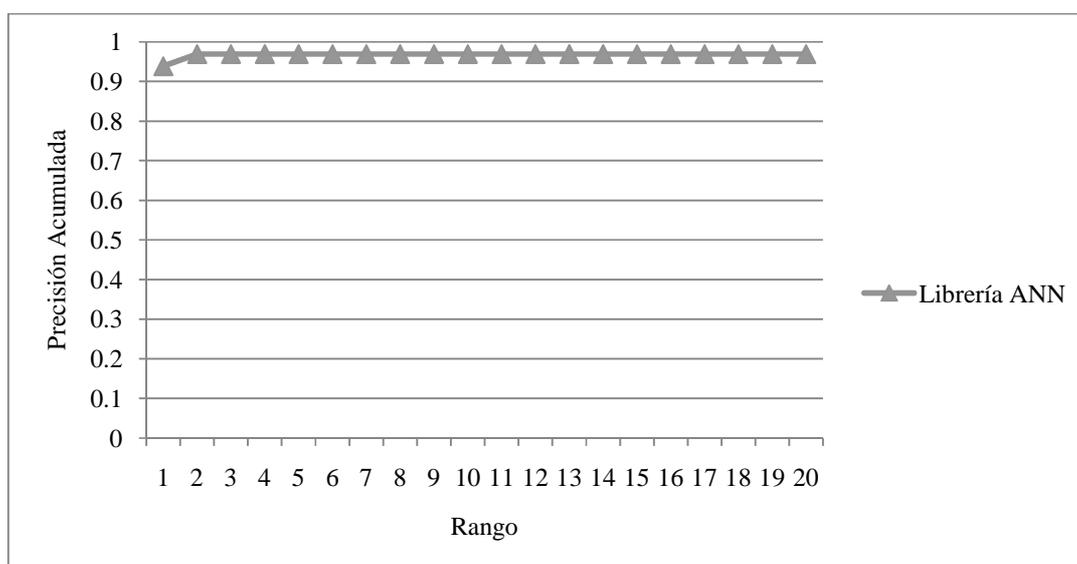


Figura 35. Curva CMC para experimentos utilizando la biblioteca ANN.

Este esquema de búsqueda arroja un tiempo promedio de búsqueda de una consulta de 54.72 segundos, lo que representa que cada comparación con las imágenes de la base de datos tarda 0.0204 segundos. A continuación se presenta el experimento sobre este esquema de búsqueda utilizando ahora la biblioteca PNN.

#### IV.2.2.2 Biblioteca PNN, utilizando permutaciones

Esta biblioteca fue creada en el desarrollo de este trabajo de investigación en lenguaje C++ a partir de la descripción de un procedimiento propuesto en Chávez *et al.* (2008), la búsqueda de los vecinos más cercanos se lleva a cabo tal como se menciona en la sección II.3.2. En la Tabla X y la Figura 36 se muestran los resultados usando 130 consultas, de las cuales 3 no se encontraron en las primeras 20 posiciones.

Tabla X. Cantidad de imágenes y precisión acumulada obtenidas en las posiciones 1-20 utilizando la biblioteca PNN.

<b>Rango</b>	<b>Cantidad</b>	<b>Precisión acumulada</b>	<b>Rango</b>	<b>Cantidad</b>	<b>Precisión acumulada</b>
<b>1</b>	123	0.946154	<b>11</b>	0	0.976923
<b>2</b>	3	0.969231	<b>12</b>	0	0.976923
<b>3</b>	1	0.976923	<b>13</b>	0	0.976923
<b>4</b>	0	0.976923	<b>14</b>	0	0.976923
<b>5</b>	0	0.976923	<b>15</b>	0	0.976923
<b>6</b>	0	0.976923	<b>16</b>	0	0.976923
<b>7</b>	0	0.976923	<b>17</b>	0	0.976923
<b>8</b>	0	0.976923	<b>18</b>	0	0.976923
<b>9</b>	0	0.976923	<b>19</b>	0	0.976923
<b>10</b>	0	0.976923	<b>20</b>	0	0.976923

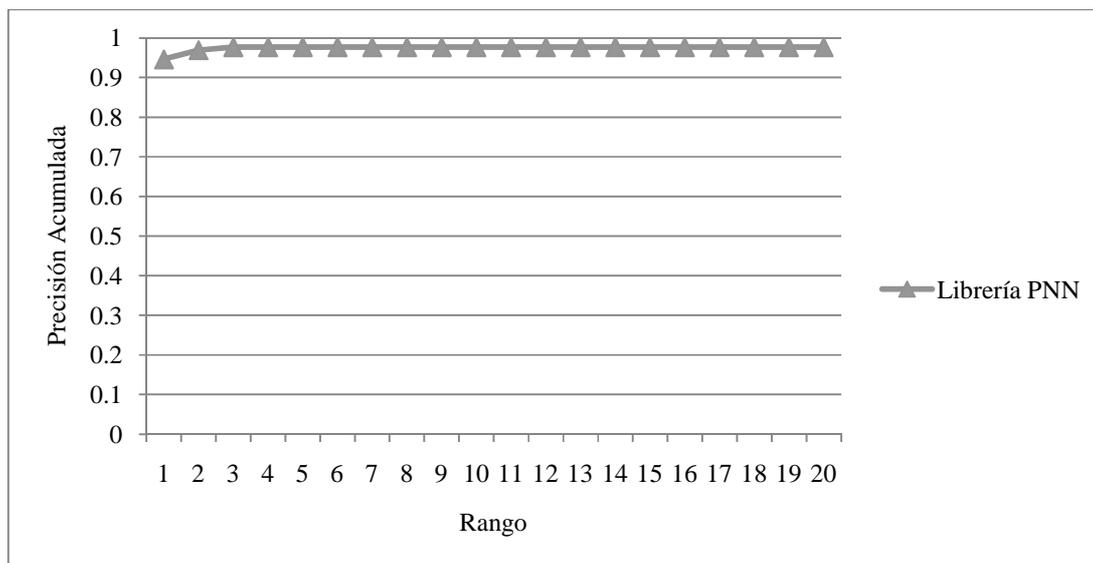


Figura 36. Curva CMC para experimentos utilizando la biblioteca PNN.

Este esquema de búsqueda arroja un tiempo promedio de búsqueda de una consulta de 441.44 segundos, lo que representa que cada comparación con las imágenes de la base de datos tarda 0.165 segundos. Este tiempo representa una menor velocidad de recuperación que con la biblioteca ANN, aunque presenta mejores resultados, sin embargo habría que hacer más experimentos con una base de datos mayor y con más consultas para observar el comportamiento de efectividad y tiempo.

Hasta este momento, los experimentos se han llevado a cabo utilizando todos los descriptores extraídos de cada imagen para representarlas mediante los mismos. Sin embargo, como se menciona en la sección II.3.1 es posible cambiar la representación de las imágenes para reducir la cantidad de información almacenada y utilizar diferentes distancias para de esta manera disminuir el tiempo de búsqueda. A continuación se presenta un esquema de recuperación que se basa en este concepto.

### IV.2.3 Uso de bolsa de palabras

En base a lo mencionado en el trabajo de Nistér y Stewenius (2006), se efectuaron experimentos sobre la base de datos de imágenes ecualizadas utilizando un vocabulario de palabras. Se realizaron diferentes tipos de experimentos mediante este esquema, cambiando la forma de obtener los representantes (centroides), la norma utilizada y una variación en el método. En las siguientes secciones se menciona la forma como fueron realizados los experimentos, así como los resultados obtenidos con cada uno.

Primero se seleccionan aleatoriamente 20000 representantes de los elementos de la base de datos, los cuales toman el papel de los elementos de las hojas del árbol o centroides del vocabulario mencionado en la sección II.3.1. En este experimento se usó la norma L2. El algoritmo de la máquina de búsqueda fue programado en C++. En la Tabla XI y la Figura 37 se presentan resultados utilizando estos veinte mil puntos y 130 consultas donde 30 no fueron encontradas en las primeras 20 posiciones.

#### IV.2.3.1 20 000 puntos aleatorios

Tabla XI. Cantidad de imágenes y precisión acumulada en las posiciones 1-20 con 20000 puntos aleatorios usando bolsa de palabras.

<b>Rango</b>	<b>Cantidad</b>	<b>Precisión acumulada</b>	<b>Rango</b>	<b>Cantidad</b>	<b>Precisión acumulada</b>
<b>1</b>	77	0.592308	<b>11</b>	0	0.746154
<b>2</b>	7	0.646154	<b>12</b>	2	0.761538
<b>3</b>	5	0.684615	<b>13</b>	0	0.761538
<b>4</b>	0	0.684615	<b>14</b>	0	0.761538
<b>5</b>	2	0.700000	<b>15</b>	0	0.761538
<b>6</b>	1	0.707692	<b>16</b>	0	0.761538
<b>7</b>	1	0.715385	<b>17</b>	1	0.769231
<b>8</b>	3	0.738462	<b>18</b>	0	0.769231
<b>9</b>	1	0.746154	<b>19</b>	0	0.769231
<b>10</b>	0	0.746154	<b>20</b>	0	0.769231

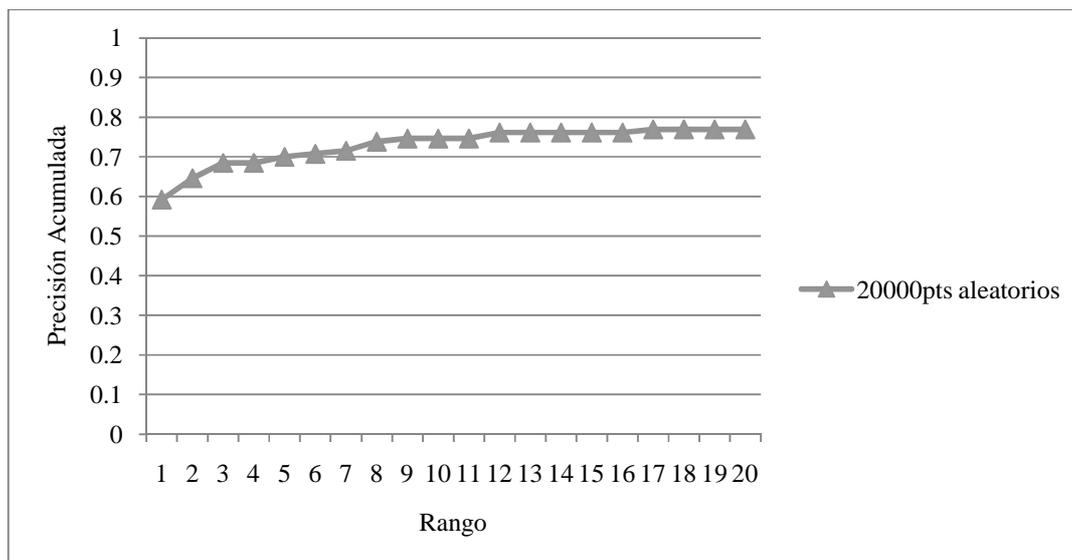


Figura 37. Curva CMC con 20000 puntos aleatorios.

Los resultados mostrados, son menos efectivos que los presentados en la sección IV.2.2, sin embargo el interés por este esquema de recuperación es debido a que el tiempo de recuperación resulta mucho más pequeño. Se tiene un tiempo promedio de recuperación en toda la base de datos de 12.51 segundos, lo que representa un tiempo de 0.004465 segundos por par de imágenes, comparado contra 0.02 segundos del tiempo proporcionado con la recuperación utilizando la librería ANN.

También se realizó un experimento usando la misma cantidad de puntos, pero ahora obtenidos a partir de un árbol jerárquico creado mediante k-medias. El algoritmo de k-medias y la obtención del árbol jerárquico fueron programados en C++.

#### IV.2.3.2 20000 puntos de árbol jerárquico

En la Tabla XII y la Figura 38 se presentan los resultados usando 130 consultas con veinte mil centroides jerárquicos, de las cuales 21 no fueron encontradas en las primeras 20 posiciones.

Tabla XII. Cantidad de imágenes y precisión acumulada en las posiciones 1-20 con 20000 centroides de árbol jerárquico usando bolsa de palabras.

Rango	Cantidad	Precisión acumulada	Rango	Cantidad	Precisión acumulada
1	95	0.730769	11	0	0.792308
2	3	0.753846	12	1	0.800000
3	2	0.769231	13	3	0.823077
4	0	0.769231	14	0	0.823077
5	1	0.776923	15	0	0.823077
6	1	0.784615	16	0	0.823077
7	0	0.784615	17	0	0.823077
8	1	0.792308	18	2	0.838462
9	0	0.792308	19	0	0.838462
10	0	0.792308	20	0	0.838462

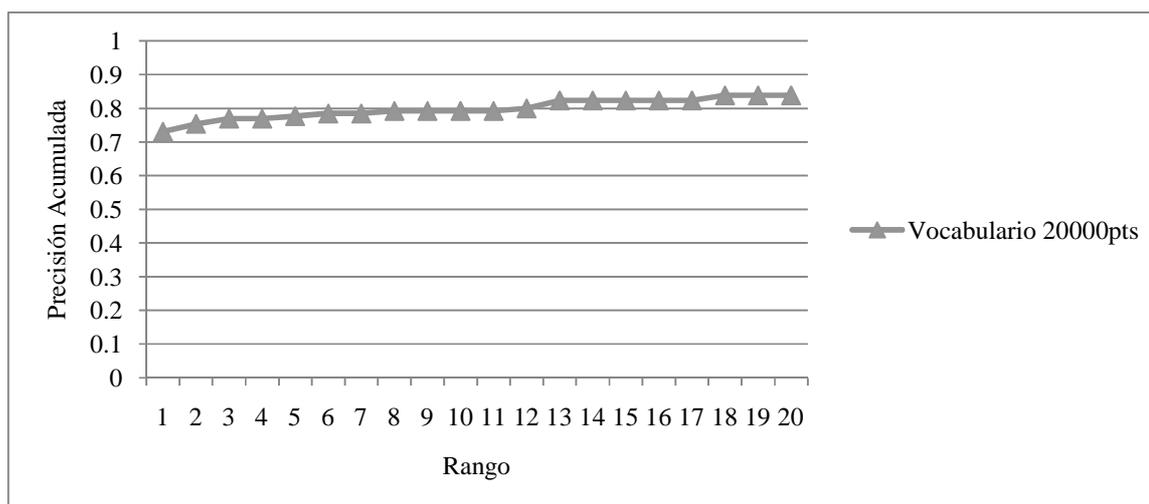


Figura 38. Curva CMC con 20000 centroides de árbol jerárquico.

Como se puede apreciar en las tablas y figuras anteriores, el utilizar representantes obtenidos a partir del árbol jerárquico incrementa la efectividad del sistema de recuperación, por lo que los siguientes experimentos se realizan utilizando centroides obtenidos de un árbol jerárquico. Ahora los experimentos se realizan con una mayor cantidad de centroides para de esta manera observar si aumenta la efectividad del sistema de recuperación.

#### IV.2.3.3 40000 puntos centroides de árbol jerárquico

En la Tabla XIII y la Figura 39 se presentan los resultados de realizar 130 consultas usando cuarenta mil centroides con k-medias jerárquico, de las cuales 14 no fueron encontradas en las primeras 20 posiciones.

Tabla XIII. Cantidad de imágenes y precisión acumulada en las posiciones 1-20 con 40000 centroides de árbol jerárquico usando bolsa de palabras y distancia L2.

<b>Rango</b>	<b>Cantidad</b>	<b>Precisión acumulada</b>	<b>Rango</b>	<b>Cantidad</b>	<b>Precisión acumulada</b>
<b>1</b>	103	0.792308	<b>11</b>	0	0.884615
<b>2</b>	1	0.800000	<b>12</b>	0	0.884615
<b>3</b>	4	0.830769	<b>13</b>	0	0.884615
<b>4</b>	1	0.838462	<b>14</b>	0	0.884615
<b>5</b>	4	0.869231	<b>15</b>	0	0.884615
<b>6</b>	0	0.869231	<b>16</b>	0	0.884615
<b>7</b>	1	0.876923	<b>17</b>	0	0.884615
<b>8</b>	1	0.884615	<b>18</b>	0	0.884615
<b>9</b>	0	0.884615	<b>19</b>	0	0.884615
<b>10</b>	0	0.884615	<b>20</b>	1	0.892308

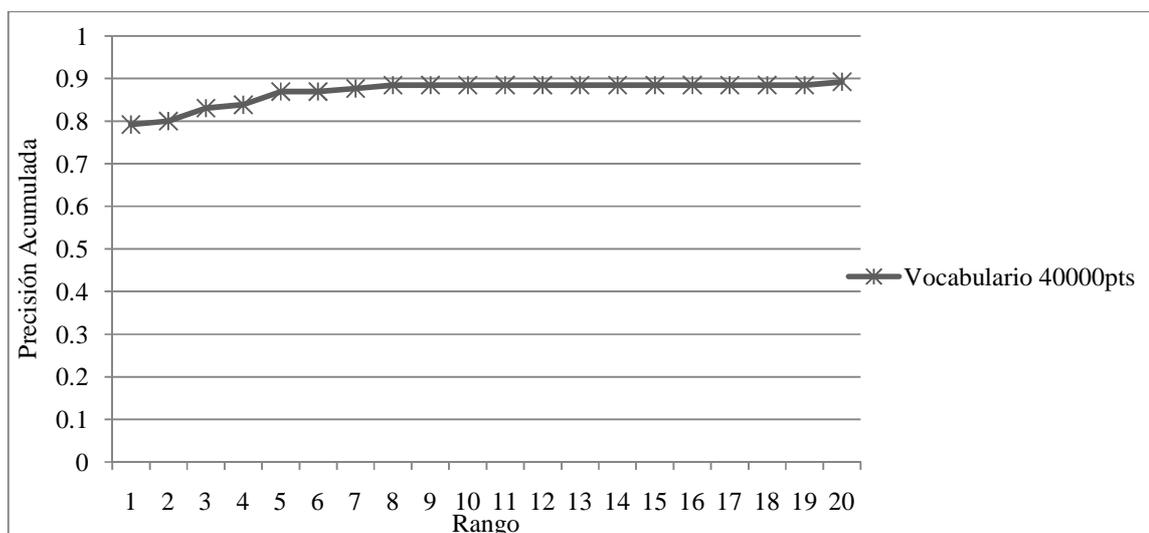


Figura 39. Curva CMC con 40000 centroides de árbol jerárquico

Como se muestra en la Tabla XIV, los resultados utilizando bolsas de palabras disminuyen bastante el tiempo de recuperación. Con los experimentos mostrados hasta el momento no se presenta incremento en la efectividad en cuanto a la recuperación. No obstante, en la siguiente sección se presentan otros experimentos mostrando que el uso de la bolsa de palabras resulta factible.

Tabla XIV. Comparación de tiempos de recuperación usando diferentes esquemas de recuperación.

	<b>1ro y 2do más cercano por par de imágenes. Sección IV.2.1</b>	<b>ANN 1ro y 2do en base de datos. Sección IV.2.2.1</b>	<b>PNN 1ro y 2do en base de datos. Sección IV.2.2.2</b>	<b>Bolsa de palabras. Sección IV.2.3</b>
Tiempo promedio de búsqueda total	476.5 seg.	54.7 seg.	441.4 seg.	12.5 seg.
Tiempo por par de imágenes	0.17 seg.	0.02045 seg.	0.165 seg.	0.00446 seg.

En los experimentos realizados hasta el momento se ha utilizado la distancia Euclidiana, sin embargo, se propone utilizar diferentes tipos de distancias para observar si se incrementa la efectividad del sistema de recuperación. En la siguiente sección se explica esta idea y se muestran los resultados de los experimentos.

#### IV.2.3.4 Uso de distintas distancias

Basándonos en el trabajo de Howart y Rüger (2005) se efectuaron experimentos con medidas de similitud utilizando distancias fraccionales. Se parte de la fórmula de la norma  $L_p$ ,

$$dist_d^p(x, y) = [\sum_{i=1}^d \|x^i - y^i\|^p]^{1/p}, \quad (17)$$

donde  $d$  es la dimensión del espacio y  $p$  un parámetro libre que generalmente es mayor o igual a uno. Una distancia fraccional se define por una distancia  $dist^p$  con  $p \in (0,1)$  la cual en un sentido matemático no es una distancia por que se viola la desigualdad del triangulo ya que el círculo de radio uno bajo  $dist^p$  deja de ser convexo para  $p < 1$ . Sin embargo, en el artículo citado se menciona que la utilización de distancias fraccionales con valores de  $p$  entre 0.25 y 0.75 incrementa la efectividad en un sistema de recuperación de imágenes donde se extrajeron características visuales de color, textura y estructura.

Para estos experimentos, se utilizó la base de datos con imágenes ecualizadas utilizando  $p=0.5, 0.75$  y 1. A continuación se presentan los resultados obtenidos.

##### IV.2.3.4.1 40000 representantes con $p=0.5$

En la Tabla XV y la Figura 40, se presentan los resultados de recuperar 130 consultas usando cuarenta mil representantes con distancia  $L_{0.5}$ , donde 7 no fueron encontradas en las primeras 20 posiciones.

Tabla XV. Cantidad de imágenes y Precisión acumulada en las posiciones 1-20 con 40000 centroides de árbol jerárquico usando bolsa de palabras y distancia L0.5.

Rango	Cantidad	Precisión acumulada	Rango	Cantidad	Precisión acumulada
1	105	0.807692	11	0	0.915385
2	7	0.861538	12	0	0.915385
3	3	0.884615	13	1	0.923077
4	1	0.892308	14	1	0.930769
5	0	0.892308	15	0	0.930769
6	2	0.907692	16	0	0.930769
7	0	0.907692	17	2	0.946154
8	0	0.907692	18	0	0.946154
9	0	0.907692	19	0	0.946154
10	1	0.915385	20	0	0.946154

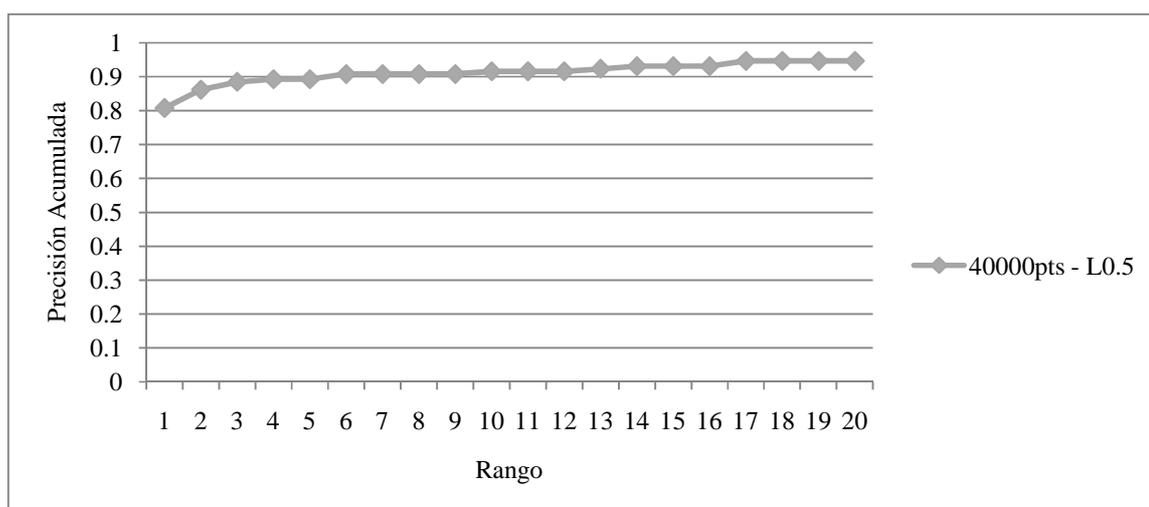


Figura 40. Curva CMC con 40000 centroides de árbol jerárquico y distancia L0.5

IV.2.3.4.2 40000 puntos con  $p=0.75$ 

En la Tabla XVI y la Figura 41, se presentan los resultados de recuperar 130 consultas usando cuarenta mil representantes con distancia L0.75, donde 6 no fueron encontradas en las primeras 20 posiciones.

Tabla XVI. Cantidad de imágenes y Precisión acumulada en las posiciones 1-20 con 40000 centroides de árbol jerárquico usando bolsa de palabras y distancia L0.75.

Rango	Cantidad	Precisión acumulada	Rango	Cantidad	Precisión acumulada
1	108	0.830769	11	0	0.953846
2	7	0.884615	12	0	0.953846
3	2	0.9	13	0	0.953846
4	2	0.915385	14	0	0.953846
5	0	0.915385	15	0	0.953846
6	1	0.923077	16	0	0.953846
7	0	0.923077	17	0	0.953846
8	1	0.930769	18	0	0.953846
9	1	0.938462	19	0	0.953846
10	2	0.953846	20	0	0.953846

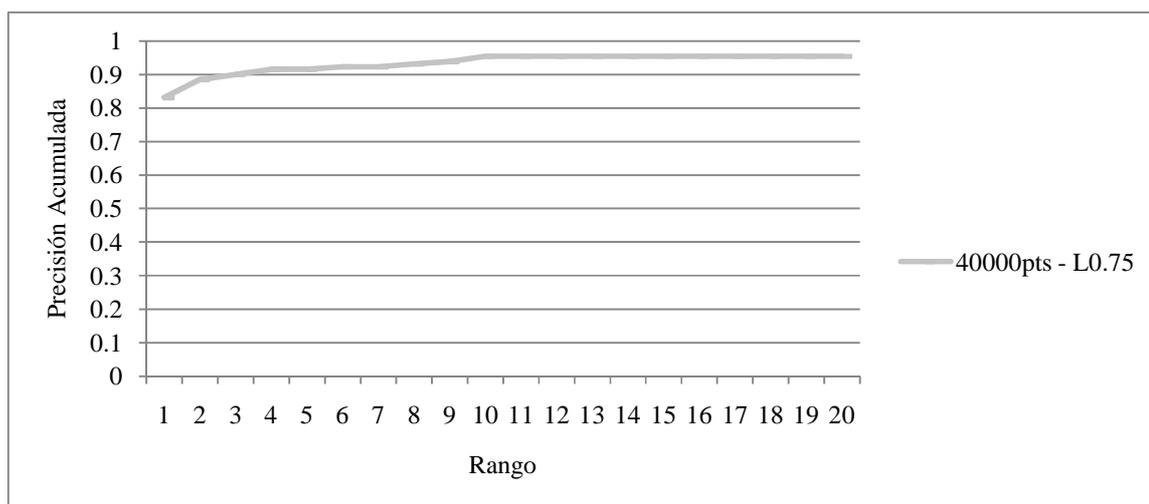


Figura 41. Curva CMC con 40000 centroides de árbol jerárquico y distancia L0.75.

IV.2.3.4.3 40000 puntos con  $p=1$ 

En la Tabla XVII y la Figura 42, se presentan los resultados de recuperar 130 consultas usando cuarenta mil representantes con distancia L1, donde 6 no fueron encontradas en las primeras 20 posiciones.

Tabla XVII. Cantidad de imágenes y Precisión acumulada en las posiciones 1-20 con 40000 centroides de árbol jerárquico usando bolsa de palabras y distancia L1.

Rango	Cantidad	Precisión acumulada	Rango	Cantidad	Precisión acumulada
1	108	0.830769	11	0	0.953846
2	5	0.869231	12	0	0.953846
3	6	0.915385	13	0	0.953846
4	1	0.923077	14	0	0.953846
5	0	0.923077	15	0	0.953846
6	2	0.938462	16	0	0.953846
7	0	0.938462	17	0	0.953846
8	2	0.953846	18	0	0.953846
9	0	0.953846	19	0	0.953846
10	0	0.953846	20	0	0.953846

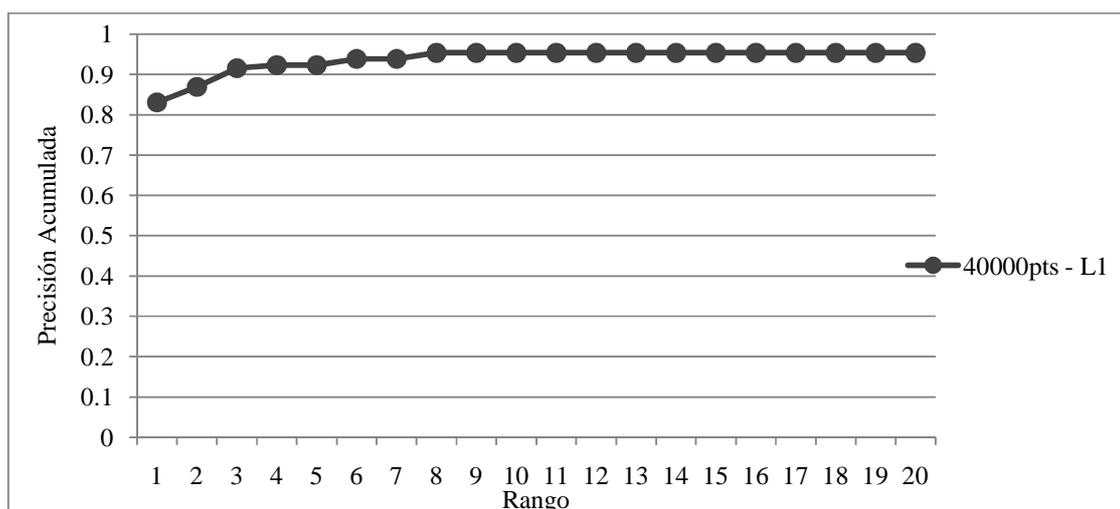


Figura 42. Curva CMC con 40000 centroides de árbol jerárquico y distancia L1.

Como se puede observar, los resultados presentados con diferentes distancias arrojan que el uso de las distancias L0.5, L0.75 y L1 logran superar los resultados obtenidos con la distancia L2, en donde L0.75 y L1 son las que logran destacar e incluso muestran resultados de recuperación similares a los obtenidos con los otros esquemas de recuperación. Se observa que la distancia L1 es la más conveniente de utilizar, ya que presenta los mejores resultados y es la más sencilla de computar pues solo representa suma de diferencias absolutas. Además no viola la desigualdad del triángulo.

A continuación, se presentan experimentos que fueron realizados sobre una modificación del método de bolsa de palabras.

#### *IV.2.3.5 Modificación del método de bolsa de palabras*

En la sección II.3.1 se mencionan experimentos efectuados donde se utilizaron búsquedas del vecino más cercano en las hojas del árbol de vocabulario. Otros experimentos consistieron en utilizar para cada descriptor de la imagen un número K de vecinos más cercanos en las hojas del árbol de vocabulario, en donde se asignaron pesos dependiendo de la posición de cercanía al descriptor de la consulta. En la Figura 43 se esquematiza la forma de representar una imagen mediante los K centroides vecinos más cercanos.

La forma de representar una imagen de consulta  $Q$  y una imagen de la base de datos  $I$ , es la siguiente:

$$Q_i = \left( \sum_{j=0}^{n_i} w_j \right) \quad (18)$$

$$I_i = \left( \sum_{j=0}^{m_i} w_j \right) \quad (19)$$

Donde  $n_i$  y  $m_i$  son el número de descriptores de las imágenes de consulta y de la base de datos respectivamente, que tienen al nodo  $i$  como parte de sus K vecinos más cercanos,  $w_j$

es un elemento del conjunto  $\mathbf{W} = \{w_1, w_2, \dots, w_K\}$  que consiste en los pesos asignados a las posiciones de la cercanía de los descriptores.

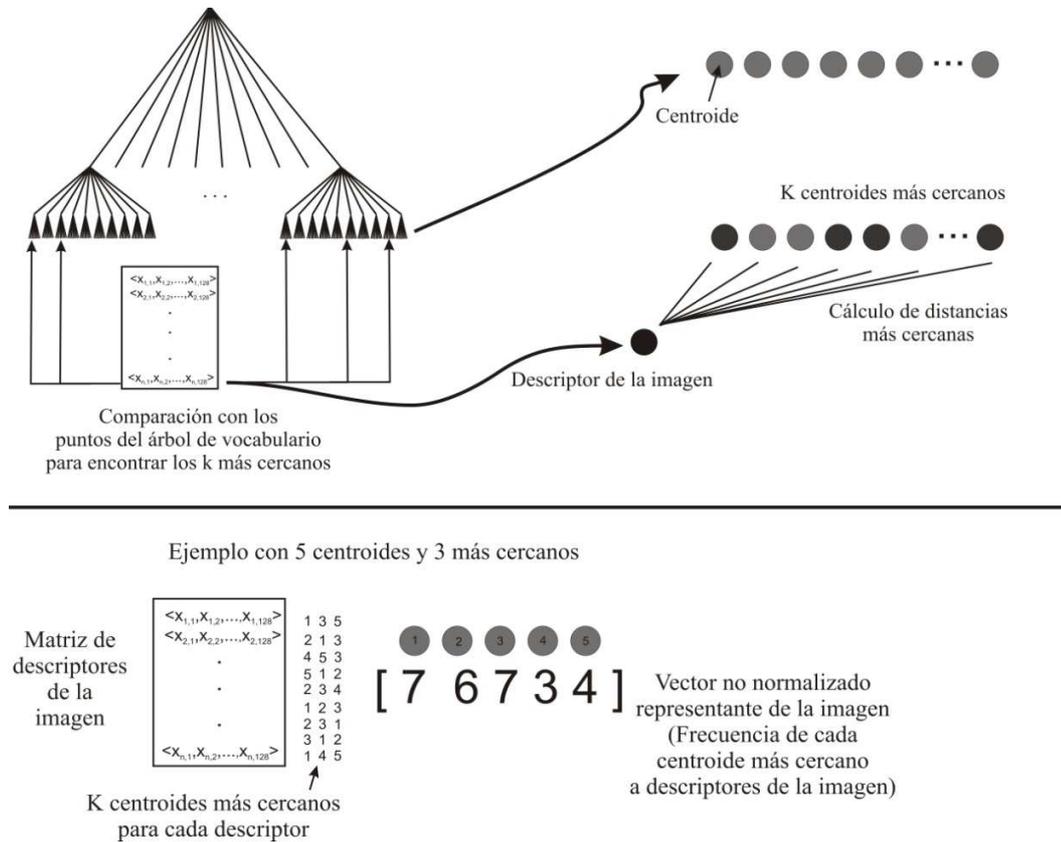


Figura 43. Representación de imágenes mediante modificación de esquema de bolsa de palabras.

Las Fórmulas 18 y 19 indican que cada uno de los K vectores más cercanos a un punto de la consulta es agregado a la representación en el elemento  $i$  con un peso que depende de su posición de cercanía.

Para los experimentos realizados se fijó  $w_i$  como constante = 1,  $K = 10$  y se emplearon  $\mathbf{W} = \{1,1,1,1,1,1,1,1,1,1\}$  y  $\mathbf{W} = \{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{128}, \frac{1}{256}, \frac{1}{512}\}$ . A continuación se muestran los resultados:

#### IV.2.3.5.1 Resultados utilizando $\mathbf{W} = \{1,1,1,1,1,1,1,1,1,1\}$

En la Tabla XVIII y la Figura 44, se presentan los resultados de recuperar 130 consultas usando cuarenta mil representantes con distancia L1 y  $\mathbf{W} = \{1,1,1,1,1,1,1,1,1,1\}$ , donde 124 no fueron encontradas en las primeras 20 posiciones.

Tabla XVIII. Cantidad de imágenes y precisión acumulada en las posiciones 1-20 con 40000 usando modificación de esquema de bolsa de palabras con  $\mathbf{W} = \{1,1,1,1,1,1,1,1,1,1\}$ .

Rango	Cantidad	Precisión acumulada	Rango	Cantidad	Precisión acumulada
1	0	0	11	0	0.023077
2	0	0	12	0	0.023077
3	1	0.007692	13	1	0.030769
4	0	0.007692	14	0	0.030769
5	0	0.007692	15	0	0.030769
6	0	0.007692	16	1	0.038462
7	1	0.015385	17	0	0.038462
8	1	0.023077	18	1	0.046154
9	0	0.023077	19	0	0.046154
10	0	0.023077	20	0	0.046154

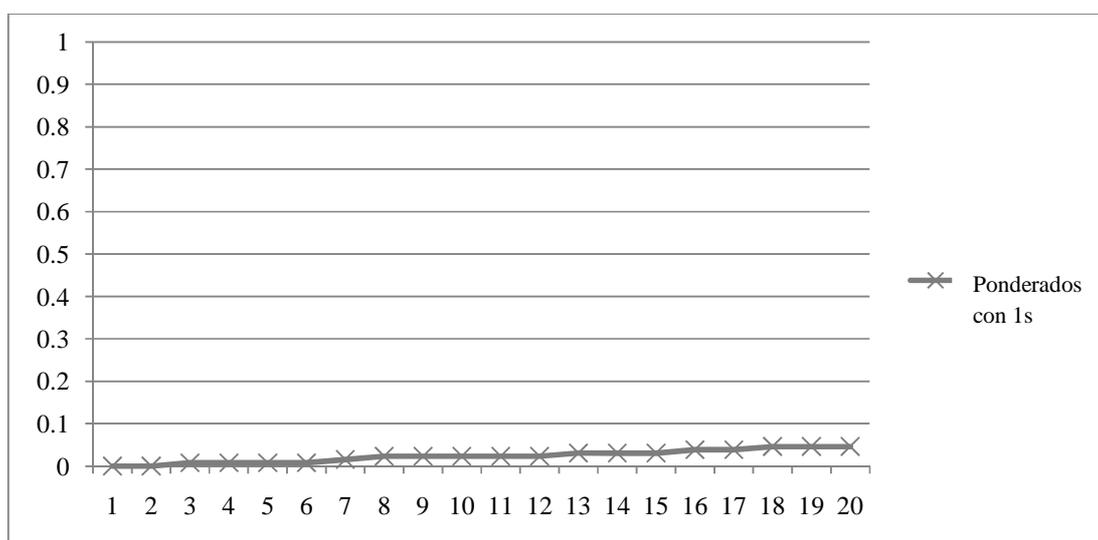


Figura 44. Curva CMC usando modificación de esquema de bolsa de palabras con  $\mathbf{W} = \{1,1,1,1,1,1,1,1,1,1\}$ .

IV.2.3.5.2 Resultados utilizando  $\mathbf{W} = \{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{128}, \frac{1}{256}, \frac{1}{512}\}$ .

En la Tabla XIX y la Figura 45, se presentan los resultados de recuperar 130 consultas usando cuarenta mil representantes con distancia L1 y  $\mathbf{W} = \{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{128}, \frac{1}{256}, \frac{1}{512}\}$ , donde 121 no fueron encontradas en las primeras 20 posiciones.

Tabla XIX. Cantidad de imágenes y precisión acumulada en las posiciones 1-20 con 40000 usando modificación de esquema de bolsa de palabras con  $\mathbf{W} = \{1, 1, 1, 1, 1, 1, 1, 1, 1, 1\}$ .

Rango	Cantidad	Precisión acumulada	Rango	Cantidad	Precisión acumulada
1	0	0	11	0	0.053846
2	1	0.007692	12	0	0.053846
3	1	0.015385	13	0	0.053846
4	0	0.015385	14	0	0.053846
5	0	0.015385	15	0	0.053846
6	1	0.023077	16	0	0.053846
7	2	0.038462	17	0	0.053846
8	0	0.038462	18	1	0.061538
9	1	0.046154	19	1	0.069231
10	1	0.053846	20	0	0.069231

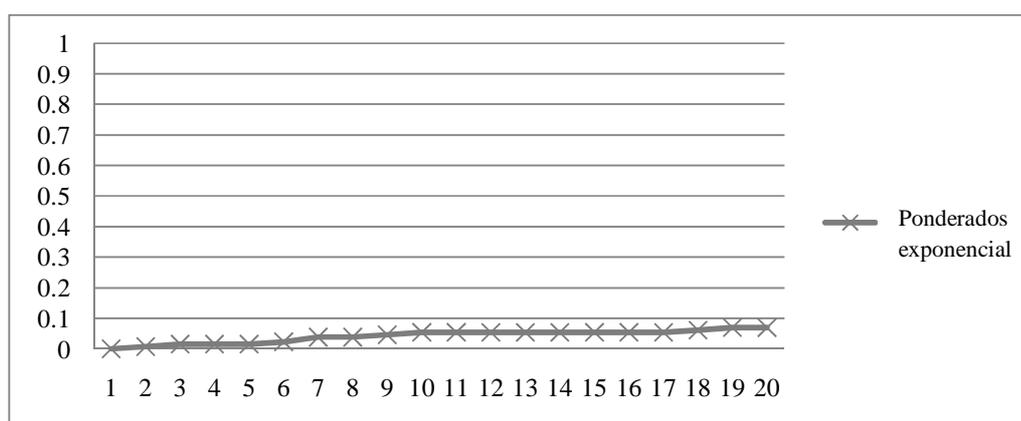


Figura 45. Curva CMC usando modificación de esquema de bolsa de palabras con

$$\mathbf{W} = \{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{32}, \frac{1}{64}, \frac{1}{128}, \frac{1}{256}, \frac{1}{512}\}.$$

La intención de representar las imágenes utilizando un número  $K$  de vecinos más cercanos es la de brindar una especie de representación difusa. Sin embargo se puede mostrar que los resultados obtenidos disminuyen bastante la efectividad del sistema de recuperación. Por tal motivo no se continuó utilizando este esquema de recuperación.

Otro tipo de experimentos fue sobre la base de datos de imágenes transformadas presentada en el capítulo III.

#### *IV.2.3.6 Experimentos con base de datos de imágenes transformadas*

Las 2100 imágenes transformadas con las que cuenta esta base de datos la cual fue presentada en la sección III.3, fueron utilizadas como consultas, en donde cada una tiene 20 imágenes relevantes dentro de la base de datos. El esquema de recuperación utilizado fue el de bolsa de palabras usando centroides jerárquicos y distancia L1. En la Tabla XX y la Figura 46 se presentan los resultados obtenidos.

Tabla XX. Exhaustividad y Precisión de consultas con base de datos de imágenes transformadas.

<b>Exhaustividad</b>	<b>Precisión</b>	<b>Exhaustividad</b>	<b>Precisión</b>
<b>0.05</b>	0.9998	<b>0.55</b>	0.9835
<b>0.1</b>	0.9997	<b>0.6</b>	0.9817
<b>0.15</b>	0.9995	<b>0.65</b>	0.9787
<b>0.2</b>	0.9981	<b>0.7</b>	0.9755
<b>0.25</b>	0.997	<b>0.75</b>	0.9706
<b>0.3</b>	0.9954	<b>0.8</b>	0.9622
<b>0.35</b>	0.9918	<b>0.85</b>	0.951
<b>0.4</b>	0.9901	<b>0.9</b>	0.9337
<b>0.45</b>	0.9885	<b>0.95</b>	0.9016
<b>0.5</b>	0.9859	<b>1</b>	0.8263

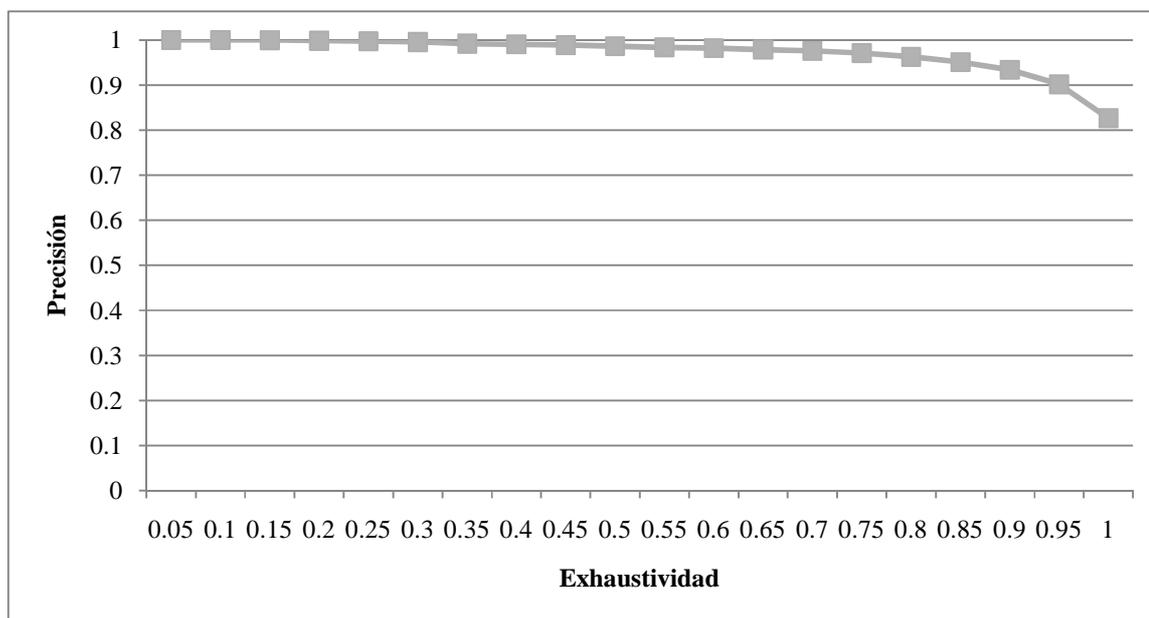


Figura 46. Curva Precisión-Exhaustividad con consultas de base de datos de imágenes transformadas.

Como ya se mencionó, no es posible comparar los resultados con los obtenidos en trabajos previos debido a que se utilizan diferentes bases de datos. Sin embargo, la figura 45 presenta condiciones más cercanas al caso ideal presentado en la figura 21 que los resultados mostrados por Jain *et al.*(2007).

En la figura 46 se presenta una gráfica que muestra la precisión en el rango 20, utilizando como consultas todas las imágenes de la base de datos de imágenes transformadas (2100), las imágenes transformadas de color (300), difuminado (200), ruido (100), perspectiva (400), rotación (500), escalamiento(500) e iluminación(100).

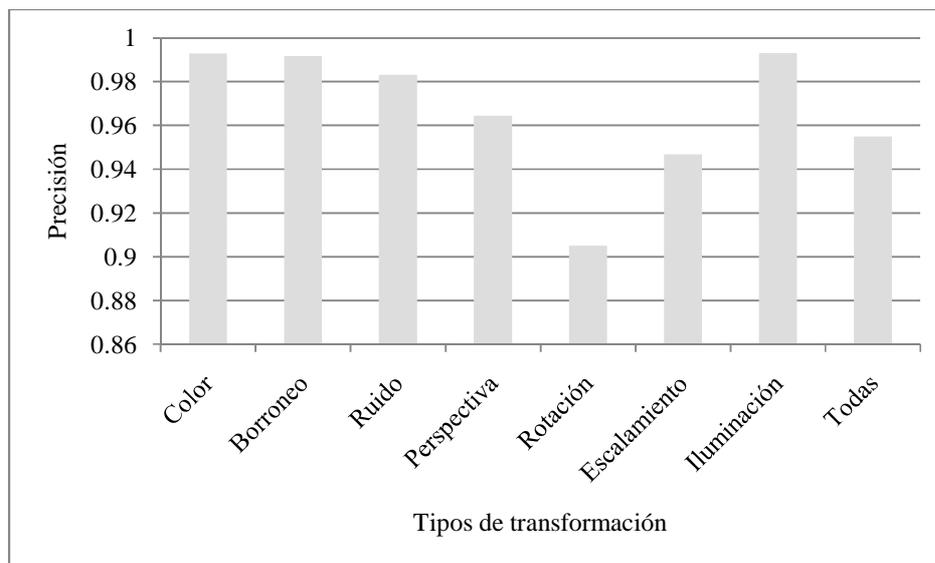


Figura 47. Gráfica de precisión de cada transformación en el rango 20.

En la gráfica anterior, se muestra que el sistema de recuperación presenta una menor efectividad con las imágenes que sufrieron las transformaciones de rotación, escalamiento y perspectiva presentadas en la sección III.3. Mientras que el sistema responde de mejor manera a las transformaciones de color, difuminado e iluminación. Hay que notar que la precisión nunca es menor de 0.9.

Los experimentos realizados hasta el momento, han sido con búsquedas por imagen ejemplo. En las figuras 48, 49, 50 y 51 se presentan resultados obtenidos al utilizar el esquema de bolsa de palabras con distancia L1 y agrupamiento jerárquico a partir de características extraídas mediante el algoritmo SIFT sobre las consultas para imágenes similares presentadas en la sección III.4.

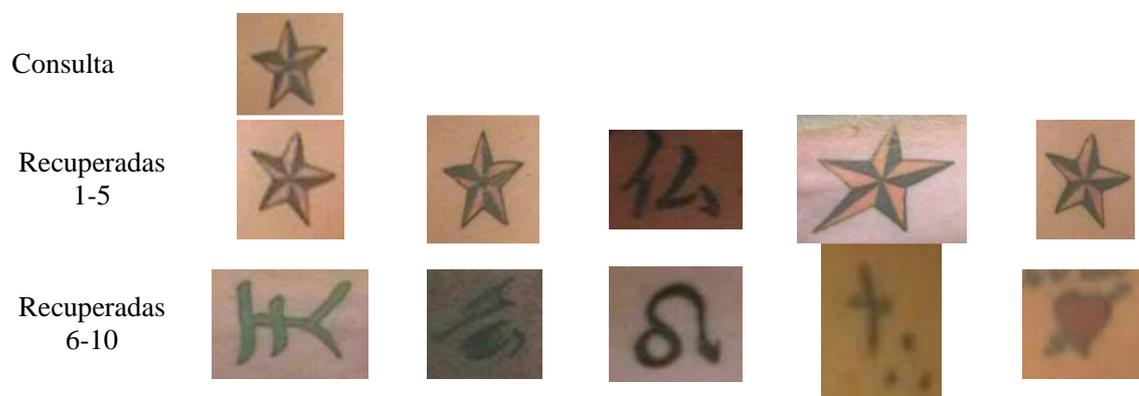


Figura 48. Recuperación de imágenes similares usando SIFT para consulta dada.



Figura 49. Recuperación de imágenes similares usando SIFT para consulta dada.



Figura 50. Recuperación de imágenes similares usando SIFT para consulta dada.

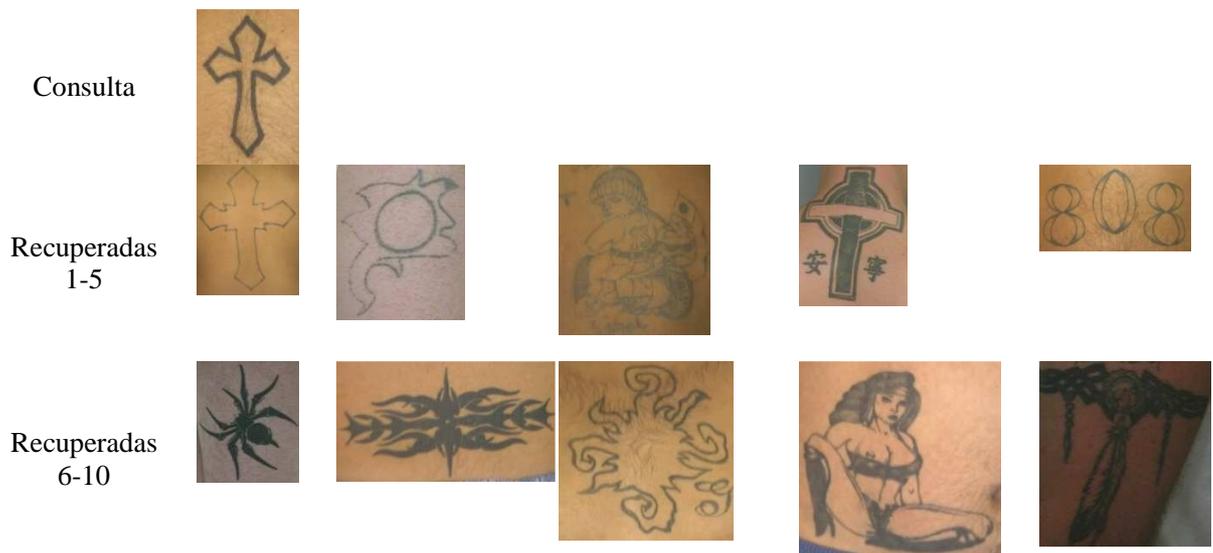


Figura 51. Recuperación de imágenes similares usando SIFT para consulta dada.

Como se aprecia en las figuras previas, el algoritmo SIFT no es efectivo para recuperar imágenes similares, ya que lo que compara son puntos de interés los cuales son representantes locales de las imágenes y no tienen contenido semántico que permita la recuperación de imágenes visualmente similares.

A continuación, se presentan los resultados obtenidos mediante el uso de la transformada wavelet bi-dimensional utilizando la aplicación “Imgseek”.

### IV.3 Experimentos utilizando Imgseek (características wavelets)

En esta sección se presentan los resultados obtenidos con el esquema de búsqueda presentado en la parte II.2.1 sobre la base de datos mostrada en la sección III.4.

En las figuras 52, 54, 56 y 58 se muestran los resultados de haber realizado la consulta correspondiente sobre la base de datos sin haber sido rotada a su orientación predominante mediante la transformada de Hough. Mientras que en las figuras 53, 55, 57 y 59 las consultas dadas fueron realizadas sobre la misma base de datos pero modificada, rotando hacia la orientación dominante mediante Hough.

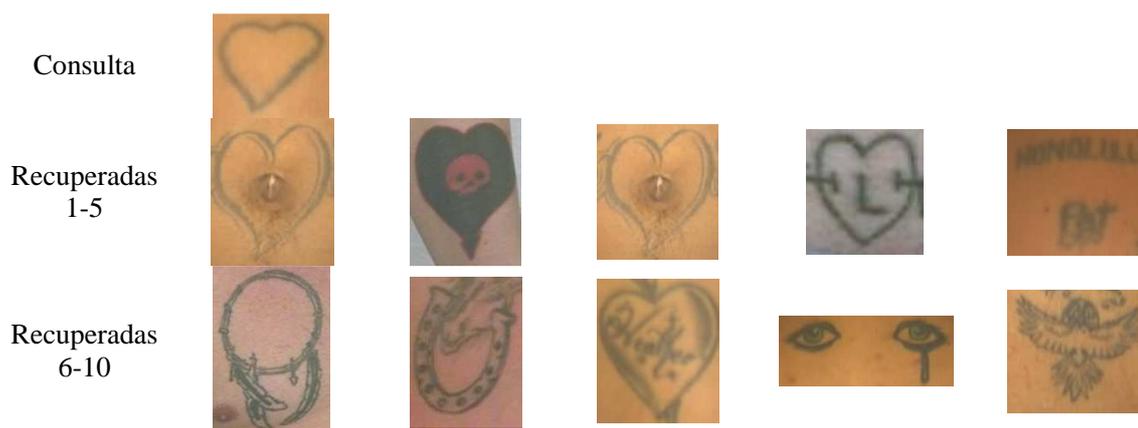


Figura 52. Diez primeras imágenes recuperadas por Imgseek sin realizar orientaciones para la consulta dada.

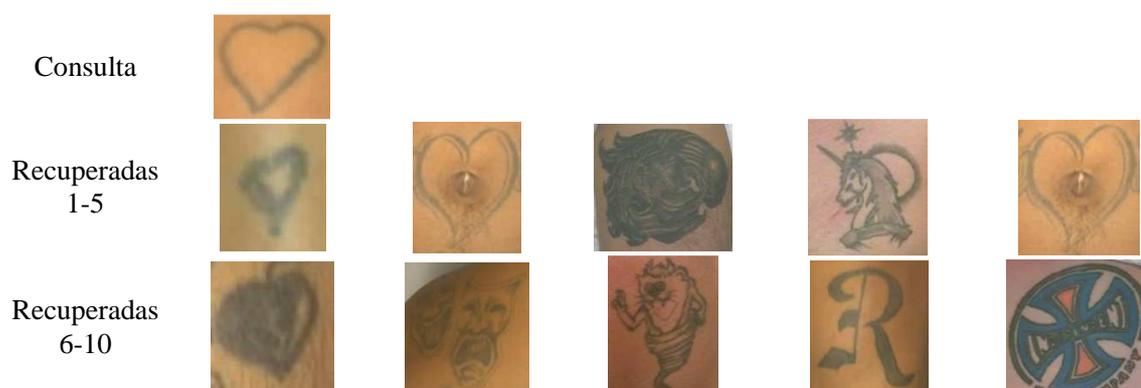


Figura 53. Diez primeras imágenes recuperadas por Imgseek realizando orientaciones para la consulta dada.



Figura 54. Diez primeras imágenes recuperadas por Imgseek sin realizar orientaciones para la consulta dada.



Figura 55. Diez primeras imágenes recuperadas por Imgseek realizando orientaciones para la consulta dada.



Figura 56. Diez primeras imágenes recuperadas por Imgseek sin realizar orientaciones para la consulta dada.

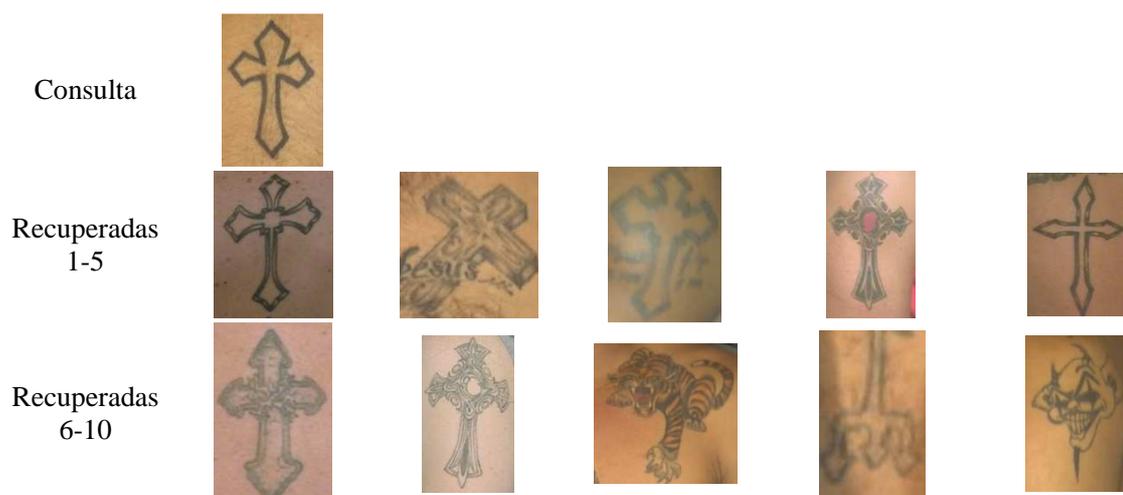


Figura 57. Diez primeras imágenes recuperadas por Imgseek realizando orientaciones para la consulta dada.

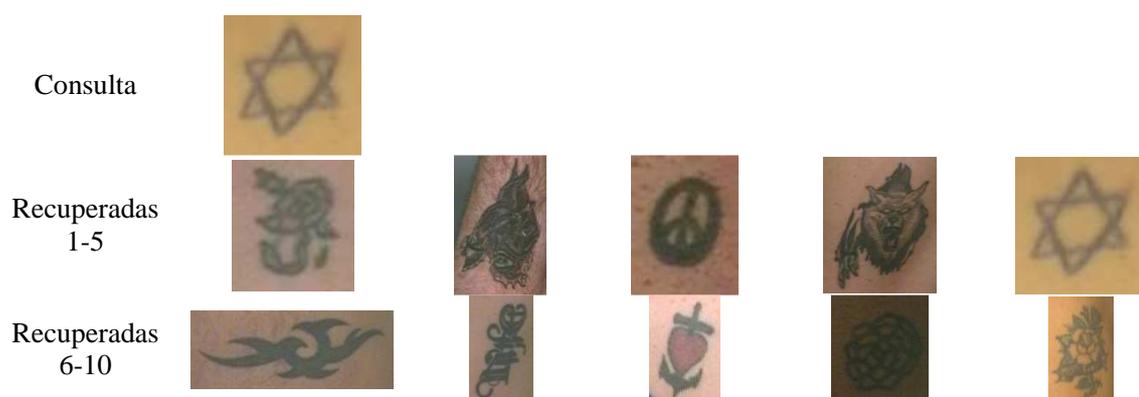


Figura 58. Diez primeras imágenes recuperadas por Imgseek sin realizar orientaciones para la consulta dada.

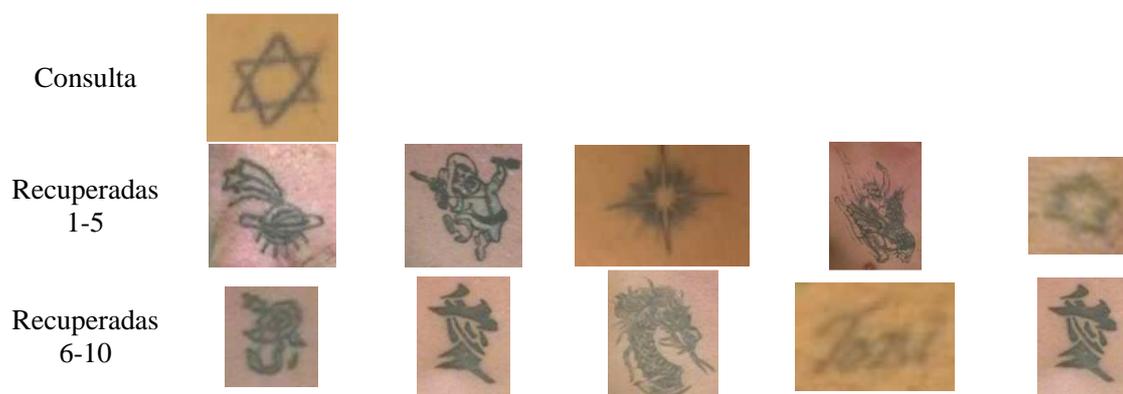


Figura 59. Diez primeras imágenes recuperadas por Imgseek realizando orientaciones para la consulta dada.

En algunos casos (búsqueda de cruces), resultó que el orientar las imágenes produjo mejores resultados, en otros, la orientación no presentó ventajas. Este tipo de consultas es complicado y requiere de un mayor estudio, así como de la creación de una base de datos más completa donde se puedan realizar experimentos y se tenga una relevancia previamente establecida.

## CAPITULO V

---

### V Conclusiones

---

En el desarrollo de esta tesis se estudió la aplicación de algunas técnicas de extracción de características en imágenes, con el objetivo de poder utilizarlas en la construcción de sistemas de recuperación de imágenes en base a su contenido visual. También se estudiaron formas de reducir la información obtenida de los extractores de características, así como mejorar las búsquedas para acelerar el tiempo de recuperación.

Específicamente, se recuperaron imágenes de tatuajes a partir de una base de datos construida como parte de este trabajo de investigación. Mediante la construcción de la base de datos se tuvo la oportunidad de conocer las características típicas de las imágenes de tatuajes en condiciones reales. El conocer dichas características permite establecer sugerencias de captura al momento de la toma de fotografía, por ejemplo, evitar tomar fotografías donde la ropa o joyas interfiera en el área del tatuaje, rasurar el área del tatuaje si existe cantidad de vello considerable que interfiera con el dibujo, definir la iluminación y los fondos correctos, establecer una forma de capturar aquellos tatuajes que no tienen límites definidos y se extienden por varias partes del cuerpo, si se da el caso que un individuo posea varios tatuajes, capturarlos de forma individual. Todas estas sugerencias se pueden estandarizar y reglamentar para que sean proporcionadas a las agencias policíacas y eviten tomar fotografías que no saquen el máximo provecho del sistema de recuperación establecido.

Fue posible detectar puntos débiles o no resueltos para la implementación de un sistema de recuperación automática de tatuajes, por ejemplo el recorte de las imágenes de tatuajes a partir de las imágenes originales y la recuperación de imágenes similares.

Se encontró que las características extraídas mediante SIFT, resultan efectivas para la recuperación de imágenes de tatuajes, sin embargo debido a la gran cantidad de información que proveen, es necesario el implementar una técnica para reducir la información y acelerar el tiempo de búsqueda.

Se probaron diferentes maneras de optimizar la búsqueda, usando librerías de vecinos más cercanos aproximados y permutaciones, encontrando que los mejores resultados se obtenían con búsquedas aproximadas en los descriptores de toda la base de datos.

Se implementó un sistema basado en bolsas de palabras, con el que se hicieron pruebas con diferentes maneras de seleccionar los representantes de los objetos. Se encontró que el mejor desempeño se obtuvo con 40000 representantes obtenidos mediante k-medias jerárquico. Además, se probaron diferentes funciones distancia para la comparación entre imágenes donde se obtuvieron los mejores resultados con la distancia L1.

También se hicieron pruebas de búsqueda por similitud usando representación por wavelets del sistema Imgseek, se encontró que se tienen problemas para recuperar imágenes con rotación lo cual se pudo mejorar en parte usando procesamiento mediante rotación hacia la dirección de mayor relevancia.

## **V.1 Trabajo futuro**

Para la creación de un sistema de recuperación automático de imágenes de tatuajes basada en contenido visual que tenga los requerimientos mencionados en los antecedentes, aún es necesaria investigación que permita las consultas de imágenes similares. Para tal aporte será necesario contar con una base de datos de mayor dimensión y encontrar la relevancia de aquellas consultas que se definan para los experimentos que se propongan realizar.

También sería recomendable buscar formas de reducir aún más la información y buscar acelerar el tiempo de búsqueda usando por ejemplo índices.

Otro trabajo necesario es buscar que el recorte de los tatuajes sobre las imágenes originales no sea de forma manual, sino automático o semi-automático.

Trabajo importante por realizar es el fusionar la información proporcionada por los tatuajes en un sistema multi-biométrico para de esta manera poder realizar vínculos entre delincuentes.

## Bibliografía

- Acton, S. T., y Rossi, A. 2008. Matching and Retrieval of Tattoo Images: Active Contour CBIR and Glocal Image Features. Image Analysis and Interpretation-IEEE Southwest Symposium on Image Analysis and Interpretation. Santa Fe, NM. 26-26 de marzo, 21-24p.
- Baeza-Yates, R., y Ribeiro-Neto, B. 2004. Modern Information Retrieval. Addison Wesley. 513pp.
- Bakken, T. 2007. An evaluation of the SIFT algorithm for CBIR. Telenor. Technical report Telenor ASA. 31pp.
- Beis, J. S., y Lowe, D. G. 1997. Shape Indexing Using Approximate Nearest-Neighbour Search in High-Dimensional Spaces. Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition San Juan, Puerto Rico, 17-19 de junio: IEEE. 1000-1006p.
- Belkin, M., y Niyogi, P. 2003. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. Neural Computation , 15 (6): 1373-1396p.
- Black, J., Fahmy, G., y Panchanathan, S. 2002. A method for evaluating the performance of content-based image retrieval systems. En: M. Lew, N. Sebe, y J. Eakins, Image and Video Retrieval. 2383: 45-92p.
- Chávez, E., Figueroa, K., y Navarro, G. 2008. Effective Proximity Retrieval by Ordering Permutations. IEEE Transactions on Pattern Analysis and Machine Intelligence , 30 (9): 1647-1658p .
- Datta, R., Joshi, D., Li, J., y Wang, J. Z. 2008. Image Retrieval: Ideas, Influences, and Trends of the New Age. ACM Computing Surveys, 40 (2):5-1p.
- del Bimbo, A. 2001. Visual Information Retrieval. San Francisco, California, USA: Morgan Kaufmann Publishers, Inc. 270pp.
- Duda, R. O., y Hart, P. E. 1972. Use of the Hough transformation to detect lines and curves in pictures. Communications of the ACM , 15 (1): 11 – 15p.

- Dunn, J. S., y Podio, F. 2005. Biometric Consortium., <http://www.biometrics.org/>, Recuperado el 2009.
- FBI, Federal Bureau of Investigation, Next Generation Identification, <http://www.fbi.gov/hq/cjisd/ngi.htm>. Recuperado el 2009.
- Fodor, I. K. 2002. A survey of dimension reduction techniques. Technical report Lawrence Livermore National Laboratory. 18pp.
- Gonzalez, R. C., Woods, R. E., y Eddins, S. L. 2004. Digital Image Processing Using MATLAB, Estados Unidos: Pearson, Prentice Hall. 609pp.
- Gunther, N. J., y Beretta, G. B. 2001. A Benchmark for Image Retrieval using Distributed Systems over the Internet: BIRDS-I. Proceedings of SPIE, the International Society for Optical Engineering, San Jose CA, USA, 24 de enero. 4311: 252-267p.
- Guyon, I., y Elisseeff, A. 2003. An introduction to variable and feature selection. The Journal of Machine Learning Research , 3:1157 – 1182p.
- He, X., y Niyogi, P. 2004. Locality Preserving Projections. En: Thrun Sebastian and Saul Lawrence and Schölkopf Bernhard, Advances in Neural Information Processing Systems 16. MIT press, Cambridge MA. 1577pp.
- Howart, P., y Rüger, S. 2005. Fractional Distance Measures for Content-Based Image Retrieval. En: Losada David E. y Fernández-Lun Juan M., Lecture Notes in Computer Science. Springer Berlin / Heidelberg. 3408: 447-456p.
- ITL, N. 2000. Information Technology Laboratory. (U.S. Department of Commerce), National Institute Of Standars and Technology: <http://www.itl.nist.gov/> Recuperado el Agosto de 2009.
- Jacobs, C. E., Finkelstein, A., y Salesin, D. H. 1995. Fast multiresolution image querying. International Conference on Computer Graphics and Interactive Techniques, Proceedings of the 22nd annual conference on Computer graphics and interactive techniques, New York, NY, USA 6-11 Agosto 1995: ACM. 277 – 286p.
- Jain, A. K., y Vailaya, A. 1996. Image retrieval using color and shape. Pattern recognition , 29 (8): 1233-1244p.

- Jain, A. K., y Vailaya, A. 1998. Shape-based retrieval: A case study with trademark image databases. *Pattern recognition* , 31 (9): 1369-1390p.
- Jain, A. K., Dass, S. C., y Nandakuma, K. 2004. Can soft biometric traits assist user recognition. *Proceedings of the SPIE, Orlando FL. USA, 12 de abril. 5404: 561-572p.*
- Jain, A. K., Ross, A., y Salil, P. 2004. An Introduction to Biometric Recognition. *IEEE Transactions on Circuits and Systems for Video Technology Special Issue on Image- and Video-Based Biometrics* , 14 (1): 4-20p.
- Jain, A., Lee, J.-E., y Jin, R. 2007. Tattoo-ID: Automatic tattoo image retrieval for suspect y victim identification. En *Advances in Multimedia Information Processing, Springer Berlin / Heidelberg. 4810:256-265p.*
- Jermyn, I. H., Shaffrey, C. W., y Kingsbury, N. G. 2002. Evaluation Methodologies For Image Retrieval Systems. *Proceedings ACIVS-Advanced Concepts for Intelligence Vision Systems. Ghent. Belgium 9-11 de septiembre. S00-1 – S00-8 p.*
- Korikana, S. K., y Chandrasekaran, V. 2008. Novel Robust and Invariant Feature Extraction by Spatio-temporal Decomposition of Images. En S. Korikana, y V. Chandrasekaran (Ed.), *Proceedings of the 2008 IEEE 8th International Conference on Computer and Information Technology Workshops, Sydney, QLD, 8-11 de julio: IEEE Computer Society. 401-405p.*
- Lee, J. E., Jain, A. K., y Jin, R. 2008. Scars, marks and tattoos (SMT): soft biometric for suspect and victim identification. *Biometrics Symposium Tampa, FL, 23-25 de septiembre : IEEE. 1-8p.*
- Lindeberg, T. 1994. Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics* , 21 (2): 224-270p.
- Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* , 60 (2): 91-110p.
- Ma, W., y Manjunath, B. 1996. Texture features and learning similarity. *Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA, 18-20 de junio: IEEE Computer Society 425-430p.*

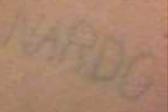
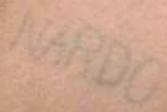
- Maio, D., Maltoni, D., Cappelli, R., Ferrara, M., Jain, A., Wayman, J., y otros. 2006. Fingerprint Verification Competition . Obtenido de Fingerprint Verification Competition 2006: <http://bias.csr.unibo.it/fvc2006>.
- Manjunath, B., y Ma, W. 1996. Texture Features for Browsing and Retrieval of Image Data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* , 18 (8): 837 – 842p.
- Manning, C. D., Raghavan, P., y Schütze, H. 2008. *Introduction to Information Retrieval*. Cambridge University Press. 544p.
- Marchand-Maillet, S., y Beretta, G. 2005. Home of CBIR Benchmarking., The Benchathlon Network: <http://www.benchathlon.net/>. Recuperado el 2009.
- Mikolajczyk, K., y Schmid, C. 2005. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* , 27 (10): 1615 – 1630p .
- Moon, H., y Phillips, P. J. 2001 . Computational and performance aspects of PCA-based face-recognition algorithms. *Perception* , 30 (3): 303 – 321p.
- Mount, D. M., y Arya, S. 2006. ANN: A Library for Approximate Nearest Neighbor Searching, 1.1.1. Department Of Computer Science, Universtiy of Maryland: <http://www.cs.umd.edu/~mount/ANN/> Recuperado el 2009.
- Newsam, S., y Kamath, C. 2005. Comparing Shape and Texture Features for Pattern. ISyT/SPIE's Annual Symposium on Electronic Imaging. San Jose, CA, United States, 17 de enero: Society of Photo-Optical Instrumentation Engineers, Bellingham. 5672: 106-117p.
- Niederberger C., R. 2007. ImgSeek, 0.8.5., <http://www.imgseek.net/> Recuperado el 2009.
- Nistér, D., y Stewenius, H. 2006. Scalable recognition with a vocabulary tree. *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. New York, USA, 17-22 de junio: IEEE Computer Society. 2, 2161- 2168p.
- Saracevic, T. 2007. Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance. *Journal of the American Society for Information Science and*

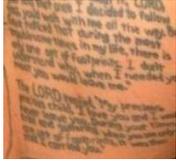
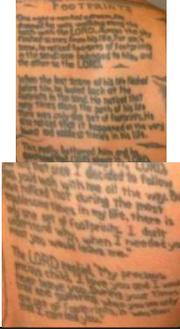
Technology. *Journal of the American Society for Information Science and Technology* , 58 (13): 1915-1933p.

- Smeulders, A. W., Worring, M., Santini, S., y Jain, R. 2000) Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence* , 12 (12): 1349 – 1380p.
- Valentine, B., y Schober, R. 2000. *Gangs and their Tattoos, Identifying Gangbangers on the Street and in Prison*. United States Of America: Paladin Press. 160pp.
- Vasconcelos, N. 2007. From pixels to semantic spaces: Advances in content-based image retrieval. *Computer* , 40 (7): 20-26p.
- Vedaldi, A. 2006. Research code. (University of California), Andrea Vedaldi: <http://www.vlfeat.org/~vedaldi/index.html> Recuperado el 2009.
- Wayman, J., Jain, A., Maltoni, D., y Maio, D. 2004. *Biometric Systems*. Springer-Verlag New York, Inc. Secaucus, NJ, USA. 370pp.
- White, R. 2008. MSU News May 27th. Obtenido de Michigan State University: <http://news.msu.edu/story/899/> Recuperado el 2008.
- Xiong, G. 2006. Matlab Central. (Mathworks). Local Adaptive Thresholding: <http://www.mathworks.com/matlabcentral/fileexchange/8647> Recuperado el 2009.

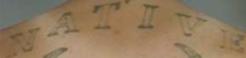
## Anexos

Imágenes de consulta con sus respectivas imágenes relevantes:

Nombre Imagen	Imagen	Tamaño	Nombre Imagen	Repeticiones	Tamaño
1004 SYMBOL		80x75	1005 SYMBOL		74x75
1022 TEXT		137x98	1023 TEXT		144x97
1081.3 TEXT		146x82	1082.1 TEXT		134x100
1085 OBJECT		437x521	1084.1 OBJECT		281x319
1093.2 OTHER		79x111	1094 OTHER		128x121
1099 ANIMAL		237x197	1100 ANIMAL		255x208
1110 ABSTRACT		318x517	1109 ABSTRACT		232x518
1112.1 ABSTRACT		125x240	1111 ABSTRACT		317x342

Nombre Imagen	Imagen	Tamaño	Nombre Imagen	Repeticiones	Tamaño
1112.2 OTHER		196x125	1114 OTHER 1113 OTHER 1111.2 OTHER		319x102 326x99 191x150
1130 OTHER		215x248	1131 OTHER		228x244
1141 TEXT		480x448	1142 TEXT 1140 TEXT		480x530 480x443
1157.1 SYMBOL		94x210	1158 SYMBOL		144x202
1161.1 SYMBOL		89x143	1162 SYMBOL		146x169
117.1 OTHER		148x153	115.7 OTHER		156x170
1184 OTHER		314x181	1183.1 OTHER		315x175

Nombre Imagen	Imagen	Tamaño	Nombre Imagen	Repeticiones	Tamaño
1193 SYMBOL		204x201	1194 SYMBOL		205x208
1212.1 HUMAN		218x426	1215.2 HUMAN		88x160
1214 ANIMAL		405x184	1215.3 ANIMAL		204x94
1232 TEXT		197x42	1229 TEXT		389x98
1233.1 OBJECT		70x41	1228.1 OBJECT		142x76
1262 TEXT		118x157	1261 TEXT		120x181
1265.1 OBJECT		110x120	1267.2 OBJECT 1268.2 OBJECT 1268.1 OBJECT		63x61 62x70 102x125
1265.2 SYMBOL		78x83	1266.2 SYMBOL 1268.4 SYMBOL		82x89 49x47
1265.3 OBJECT		185x396	1265.3 OBJECT 1266.1 OBJECT		295x389 176x216
1267.1 SYMBOL		82x94	1266.3 SYMBOL		72x84
1274 TEXT		285x245	1275 TEXT		289x201

Nombre Imagen	Imagen	Tamaño	Nombre Imagen	Repeticiones	Tamaño
1296 ABSTRACT		288x555	1297 ABSTRACT 1298 ABSTRACT		446x592 400x572 479x432
1305.2 TEXT		444x112	1306.1 TEXT		442x116
1305.3 ANIMAL		360x269	1307.2 ANIMAL 1306.2 ANIMAL		356x317 261x320
1306.3 ANIMAL		60x58	1305.4 ANIMAL 1307.1 ANIMAL		54x58 53x58
1306.4 TEXT		382x172	1307.3 TEXT 1307.1 TEXT		379x191 378x193
1313.1 OBJECT		151x349	1313.2 OBJECT		154x347
1313.1 TEXT		231x140	1313.3 ABSTRACT		232x134
1322 ABSTRACT		153x300	A323 ABSTRACT		169x307
1334 OTHER		233x114	1333 OTHER		223x119

Nombre Imagen	Imagen	Tamaño	Nombre Imagen	Repeticiones	Tamaño
1339 TEXT		398x75	1340 TEXT		397x72
1345 TEXT		120x49	1346 TEXT		124x53
1357 ABSTRACT		325x504	1356 ABSTRACT		268x515
1358 OTHER		276x442	1359 OTHER		279x418
1360.2 OTHER		127x104	1362.2 OTHER		122x116
1361.1 ANIMAL		145x176	1360.1 ANIMAL		140x209
1374.1 ANIMAL		110x140	1375.1 ANIMAL		114x143
1374.2 OBJECT		126x147	1375.2 OBJECT		117x160
1384 OTHER		88x107	1383 OTHER		80x104
1396 OBJECT		144x108	1397 OBJECT		143x112
1401 TEXT		164x106	1402 TEXT		166x106

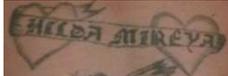
Nombre Imagen	Imagen	Tamaño	Nombre Imagen	Repeticiones	Tamaño
1429.1 ABSTRACT		298x333	1430.1 ABSTRACT		407x299
1429.2 OTHER		145x98	1430.2 OTHER		84x146
1431.2 TEXT		187x138	1431.1 OBJECT 1431.3 TEXT		255x238 205X113
1432.2 TEXT		187X144	1432.1 OBJECT 1432.3 TEXT		253X252 209X119
1434 OTHER		249x212	1435 OTHER		249x209
1448 HUMAN		147x90	1449 HUMAN		151x109

Nombre Imagen	Imagen	Tamaño	Nombre Imagen	Repeticiones	Tamaño
1456 ABSTRACT		273x570	1457 ABSTRACT 1455 ABSTRACT 1458 ABSTRACT		341x576 319x559 273x449
154.3 HUMAN		261x347	153.2 HUMAN		356x335
1541 ABSTRACT		480x600	1539 ABSTRACT 1542 ABSTRACT		480x511 480x479

Nombre Imagen	Imagen	Tamaño	Nombre Imagen	Repeticiones	Tamaño
1545 ABSTRACT		480x415	1544 ABSTRACT 1543 ABSTRACT		480x442 476x398
1549 ABSTRACT		401x366	1548 ABSTRACT		419x600
1559 ABSTRACT		480x600	1554 ABSTRACT		480x600
1568 ABSTRACT		480x600	1567 ABSTRACT		480x600
1571 ABSTRACT		480x516	1572 ABSTRACT 1570 ABSTRACT		432x418 480x600

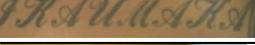
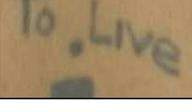
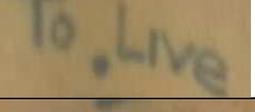
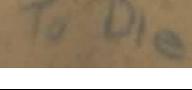
Nombre Imagen	Imagen	Tamaño	Nombre Imagen	Repeticiones	Tamaño
1648 TEXT		100x109	1649 TEXT		121x124
1657 TEXT		100x484	1658 TEXT		119x483
1673 OTHER		165x158	1674 OTHER		149x73
1695 TEXT		406x183	1694 TEXT		401x165
1718 OTHER		279x121	1719 OTHER		282x110
1753 ABSTRACT		162x511	1752 ABSTRACT		219x427
1758 OTHER		204x323	1757 OTHER		185x404
1763 TEXT		99x170	1764 TEXT		108x161

Nombre Imagen	Imagen	Tamaño	Nombre Imagen	Repeticiones	Tamaño
1785 TEXT		141x381	1783 TEXT 1784 TEXT		157x542 200x521
1788 TEXT		220x368	1787 TEXT 1786 TEXT		158x407 156x405
1815 ANIMAL		128x238	1816 ANIMAL		117x238
1817 OBJECT		69x69	1818 OBJECT		79x79
1824 OBJECT		67x59	1823 OBJECT 1822 OBJECT		54x60 70x54
187.1 TEXT		114x76	188.2 TEXT		175x57
1919.1 OBJECT		198x148	1937.1 OBJECT		215x223
1919.4 TEXT		78x61	1937.4 TEXT		74x68 84x87
1920.1 TEXT		159x70	1919.2 TEXT		159x72
1938 TEXT		200x102	1921 TEXT		175x98

Nombre Imagen	Imagen	Tamaño	Nombre Imagen	Repeticiones	Tamaño
1984.2 ANIMAL		224x271	1984 OBJECT PLANT ANIMAL SYMBOL		224x451
1985.2 ABSTRACT		293x320	1985 OTHER ABSTRACT		314x437
2008.1 ABSTRACT		110x112	2008.2 ABSTRACT		111x111
2016.1 ANIMAL		87x125	1966 ANIMAL		165x111
2049 ABSTRACT		128x118	2050 ABSTRACT		131x107
2056.2 HUMAN		108x87	2057.2 HUMAN		114x85
2057.1 OTHER		178x162	2056.1 OTHER		167x198
2091.2 TEXT		60x31	1090 TEXT		65x31
2094.1 OBJECT TEXT		355x123	2095.2 OBJECT TEXT		417x148
2094.2 ANIMAL		183x197	2095.2 ANIMAL		212x230

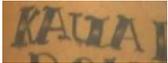
Nombre Imagen	Imagen	Tamaño	Nombre Imagen	Repeticiones	Tamaño
2097.1 HUMAN		88x157	2096 HUMAN		132x142
210 TEXT		768x544	211 TEXT		800x468
221 TEXT		380x249	222 TEXT		450x266
230.2 SYMBOL		102x101	231.2 SYMBOL		96x99
230.3 HUMAN		483x183	231.1 HUMAN		333x151
238.1 HUMAN		258x239	237.2 HUMAN		146x298
446.2 ANIMAL		225x222	445.2 ANIMAL		242x216
493.5 OTHER		258x393	495.3 OTHER		215x376
505.2 OTHER		288x191	504.1 OTHER		284x179

Nombre Imagen	Imagen	Tamaño	Nombre Imagen	Repeticiones	Tamaño
507.2 TEXT		346x106	508 TEXT		432x125
509 OBJECT		497x216	513.3 OBJECT 510.3 OBJECT		326x186 280x210
52.3 TEXT		140x47	52.2 OBJECT		144x153
57.1 TEXT		550x339	57.2 ABSTRACT		452x405 384x342
58.2 ABSTRACT		241x656	57.3 ABSTRACT		284x630

Nombre Imagen	Imagen	Tamaño	Nombre Imagen	Repeticiones	Tamaño
610 OTHER		645x925	611 OTHER		689x863
614.3 TEXT		723x593	615.2 TEXT		699x680
620.4 OBJECT		234x324	621.2 OBJECT		222x380
622 TEXT		774x149	623 TEXT		731x135
63 TEXT		708x288	861 TEXT		800x281
632.2 TEXT		209x134	633.2 TEXT		225x112
632.3 TEXT		229x88	633.3 TEXT		212x99
633.1 SYMBOL		156x272	632.3 SYMBOL		201x259

<b>Nombre Imagen</b>	<b>Imagen</b>	<b>Tamaño</b>	<b>Nombre Imagen</b>	<b>Repeticiones</b>	<b>Tamaño</b>
650.2 OBJECT		208x271	651.1 OBJECT		198x269
653.2 OTHER		263x595	654 OTHER		294x634
666 ANIMAL		504x450	671 ANIMAL		540x444
667.3 OTHER		299x142	668.2 OTHER		387x223
677 ABSTRACT		309x872	678 ABSTRACT		260x690
696.2 HUMAN		404x539	697.1 HUMAN		349x625
721.3 HUMAN		120x316	716.4 HUMAN		189x353
739 OTHER		345x132	738 OTHER		336x130

Nombre Imagen	Imagen	Tamaño	Nombre Imagen	Repeticiones	Tamaño
758 ANIMAL		797x858	759.1 ANIMAL		563x824
857.2 OBJECT		177x198	858.2 OBJECT		166x202
857.3 TEXT		238x147	858.4 TEXT		237x148
857.4 TEXT		247x165	858.3 TEXT		237x160
861 TEXT		800x303	862 TEXT		785x307
89.1 ANIMAL		356x345	88.2 ANIMAL		376x287
963.1 HUMAN		188x227	962.3 HUMAN		186x276
963.3 HUMAN		246x204	964.2 HUMAN		200x191

Nombre Imagen	Imagen	Tamaño	Nombre Imagen	Repeticiones	Tamaño
966.2 TEXT		333x127	968.1 TEXT 969.1 TEXT 967.1 TEXT		341x187 360x188 341x182
968.2 TEXT		311x139	966.1 TEXT		241x161
973.3 OBJECT		147x191	975.1 OBJECT		142x177
999.1 TEXT		112x74	1001.1 TEXT		109x65