



CENATAV

Centro de Aplicaciones de
Tecnologías de Avanzada
MINISTERIO DE LA INDUSTRIA BÁSICA

RNPS No. 2142
ISSN 2072-6287
Versión Digital

REPORTE TÉCNICO
**Reconocimiento
de Patrones**

SERIE AZUL

**Estado del arte sobre los métodos de
detección de características locales**

Lic. Reynel Remón Martínez,
Dr. C. Edel García Reyes

RT_019

febrero 2010





CENATAV

Centro de Aplicaciones de
Tecnologías de Avanzada
MINISTERIO DE LA INDUSTRIA BÁSICA

RNPS No. 2142
ISSN 2072-6287
Versión Digital

SERIE AZUL

REPORTE TÉCNICO
**Reconocimiento
de Patrones**

**Estado del arte sobre los métodos de
detección de características locales**

Lic. Reynel Remón Martínez,
Dr. C. Edel García Reyes

RT_019

febrero 2010



Estado del arte sobre los métodos de detección de características locales

Lic. Reynel Remón Martínez, Dr. C. Edel García Reyes

Centro de Aplicaciones de Tecnología de Avanzada, 7a #21812 e/ 218 y 222, Siboney, Playa, Ciudad de La Habana, Cuba

remon@cenatav.co.cu

RT_019 CENATAV

Fecha del camera ready: 30 de octubre de 2009

Resumen: Las características locales se refieren a un conjunto de rasgos relevantes que se encuentran presentes en las imágenes en un ámbito local. Este término agrupa tanto puntos de interés, entre los que se encuentran diversos tipos de uniones, como pequeñas regiones de la imagen, los cuales deben cumplir un conjunto de propiedades deseables como son la distintividad de la característica, así como que debe presentar cierto grado de invarianza ante diversas transformaciones. En este trabajo, se realiza un estudio de los principales algoritmos de detección de características locales presentes en la literatura, así como de las ideas en las que se basan para realizar la detección. Considerando esto último, junto con el tipo de imagen sobre la que trabajan, presentamos una taxonomía que los agrupa para una mejor comprensión.

Palabras claves: características locales, puntos de interés, detección de esquinas, uniones, regiones invariantes, detectores de blob, descriptores de regiones

Abstract: Local Features are a group of relevant regions present in images which have a local scope. Local features are a general concept that involves different kind of corners, blobs or salient regions. These features present a set of properties like distinctiveness, invariance and robustness under different transformations. In this work we perform a study about existing methods of local features detection. We focus on main assumptions in which they based their detection procedure. Taking this into consideration we propose a taxonomy for grouping existing approaches.

Keywords: Local Features, Corner Detectors, Invariant Regions, Blob Detectors, Region Description

1. Introducción

La extracción de características locales en imágenes es un tema que ha sido estudiado por más de 50 años. Sus orígenes se pueden rastrear hasta mediados de los años 50's en la observación realizada por [1] en la cual planteaba que las esquinas contienen una información importante sobre la forma de un objeto. Pero no fue hasta principio de la década del 70, cuando empezaron a surgir los primeros trabajos en este tema. Inicialmente, los autores se concentraron en extraer esquinas definidas como puntos donde los contornos presentan una alta curvatura, pero a medida que fueron evolucionando las investigaciones en estos temas se comenzaron a introducir operadores capaces de extraer otro tipo de información local relevante y distintiva en imágenes más complejas.

1.1 Características locales

No existe una definición precisa de lo que son las características locales, pueden ser diversos tipos de esquinas (uniones tipo T, tipo L, tipo Y), pueden ser regiones de intensidad uniforme (conocidos como blob), regiones con un alto contenido de información. Lo que si queda claro es que tienen un carácter local. De esta forma, una característica local puede ser definida como un patrón presente en la imagen que se diferencia de su vecindario inmediato. Normalmente están asociadas a una variación en las propiedades de la imagen (intensidad, color, textura), aunque no tienen que estar necesariamente localizadas en el preciso lugar del cambio. Idealmente, estos rasgos debieran corresponder a partes semánticamente significativas de los objetos presentes en la escena. En la práctica esto no sucede, pues para ello sería necesaria una interpretación del contenido global de la imagen que no es deseable en etapas tempranas de procesamiento.

Una característica local ideal debe corresponder a un punto geométrico: una localización definida en el espacio pero sin área. En la práctica, debido a la naturaleza discreta de las imágenes, es necesario analizar un vecindario para su extracción, dándoles de forma implícita una magnitud espacial. Para algunas aplicaciones como las de calibración de cámaras o reconstrucción 3D, esta información es completamente ignorada en procesamientos posteriores y solo se tiene en consideración la posición. En estos casos, se refieren a estos rasgos como puntos de interés.

Sin embargo, en la mayoría de las aplicaciones estas características necesitan ser descritas de forma tal que puedan ser identificadas y comparadas, y esto fuerza a tener en cuenta el vecindario que las rodea. Dicho vecindario es analizado, convirtiendo sus propiedades en descriptores. En este contexto, nos referimos a regiones en vez de puntos de interés. Este tipo de extractores deben ser capaces de detectar no solo la localización de la característica, sino que también deben estimar el tamaño y la forma de la región distintiva asociado a este.

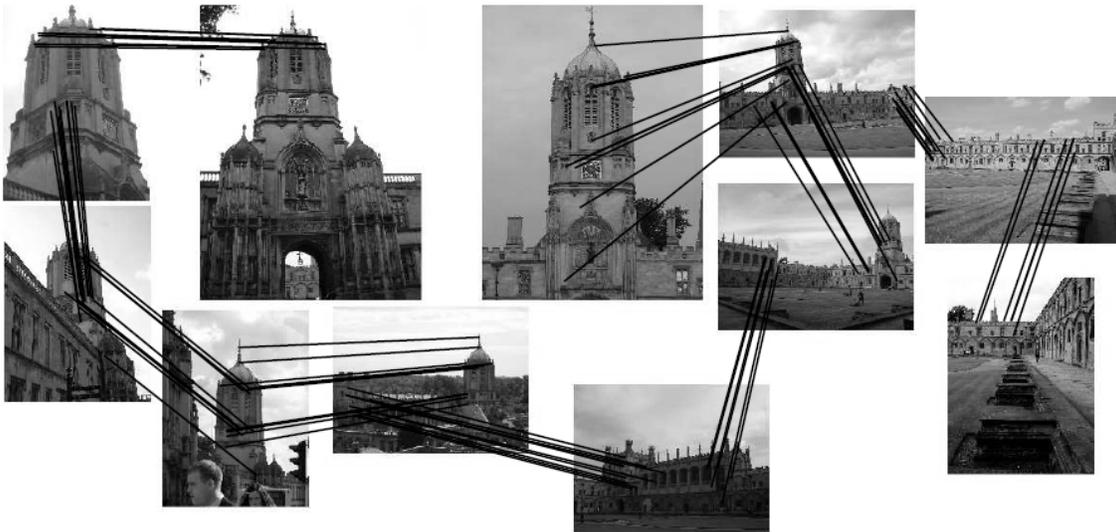


Fig. 1. Agrupamiento de imágenes basado en regiones de interés [2]

Las características locales invariantes permiten no solo encontrar correspondencias entre imágenes sujetas a cambios significativos de las condiciones de observación y oclusiones, sino

que también deben mantener una descripción del contenido de la imagen útil en la recuperación y categorización de imágenes [2-3] (Fig. 1), reconstrucción de escenas, en el reconocimiento [4-5] (Fig. 2) y seguimiento de objetos [6-8]. En [9] se analiza la superioridad de las características locales con respecto a las globales en este tipo de aplicaciones.

Teniendo en cuenta su posible uso, los extractores de rasgos pueden ser agrupados en tres categorías:

- Características con interpretación semántica en un contexto limitado por el dominio de una aplicación específica. Este tipo de rasgos son utilizados en aplicaciones de teledetección donde las líneas extraídas pueden corresponder a carreteras, o regiones de textura homogénea pueden indicar zonas de agua o vegetación.
- Conjunto de puntos independientes que sean identificables, pero su interpretación semántica no es relevante. Esta es la situación en la mayoría de las aplicaciones de calibración de cámara, reconstrucción 3D, estimación de pose, alineamiento de imágenes.
- Conjunto de características relacionadas entre sí, como una representación robusta de la imagen. Las características por separado no tienen una interpretación semántica, y no tienen que tener una localización exacta, puesto que el objetivo no es compararlas individualmente, sino más bien analizar sus estadísticas. Estas características son utilizadas por aplicaciones de reconocimiento de objetos, recuperación de imágenes, minado de videos.

Claramente cada categoría impone sus propias restricciones, por lo que un tipo de rasgo puede presentar buenos resultados en una determinada aplicación y resultar inútil en el contexto de un problema diferente.

1.2 Principales propiedades de las características locales

Las características locales tienen un conjunto de propiedades que determinan su comportamiento en una aplicación específica. El extractor ideal es aquel que tenga un perfecto desempeño en todas estas propiedades. En la práctica esto no ocurre, debido principalmente a que muchas de ellas determinan funcionamientos contrarios, por lo que el aumento de una conlleva a que otras se vean afectadas. En la elección de un extractor se hace necesario llegar a un consenso sobre cuáles deben ser priorizadas, esto a su vez depende de las características necesarias en el dominio de aplicación en el que será utilizado [9].

Entre las propiedades de los extractores se encuentran:

- Repetitividad: Dadas dos imágenes pertenecientes a la misma escena tomadas bajo diferentes condiciones de observación, es deseable que la mayor parte de las características detectadas en una sean redetectadas en la otra.
- Distintividad: Los patrones de las características deben ser lo suficientemente representativos como para ser correctamente identificados.
- Localidad: Las características deben representar un área local de la imagen.
- Estabilidad: La posición de las características detectadas no deben variar al analizar varias imágenes de la misma escena.
- Cantidad: El número de características detectadas debe ser razonablemente alto que permita la extracción de rasgos en pequeños objetos presentes en la imagen.
- Exactitud: Las características detectadas deben ser localizadas de forma precisa, con respecto a la escala y posiblemente a la forma.
- Eficiencia: El proceso de extracción debe ser realizado de manera que permita su utilización por aplicaciones que funcionen en tiempo real.

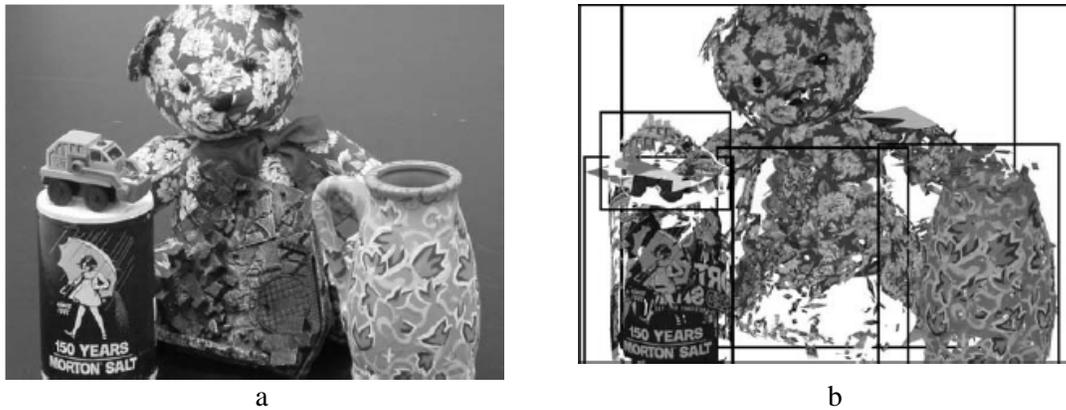


Fig. 2. Características locales en el reconocimiento de objetos [5]

La propiedad más importante es la Repetitividad y puede ser alcanzada por dos vías diferentes, a través de la Invarianza (cuando grandes deformaciones son esperadas, se intentan modelar matemáticamente y se desarrollan métodos que se comporten invariantes ante estas transformaciones) y la Robustez (en el caso de transformaciones relativamente pequeñas, a veces es suficiente hacer el detector menos sensible a estas deformaciones). El principal problema de las características con alto grado de invarianza consiste en que su poder discriminativo se ve afectado y el proceso de detección se hace más complejo. Teniendo en cuenta esto, a veces es preferible la utilización de detectores robustos, combinados con descriptores de regiones, en vez de utilizar aquellos invariantes ante transformaciones más complejas.

En este trabajo nos vamos a centrar en el estudio de los algoritmos de detección presentes en la literatura (Capítulo 2), así como en las principales investigaciones que se están llevando a cabo en esta área (Capítulo 3) Para un análisis más detallado de los descriptores se pueden consultar [10-11].

2 Revisión de los principales métodos de extracción de características locales

En esta sección realizaremos una breve revisión de los principales métodos de extracción de puntos/regiones de interés presentes en la literatura. Debido a la gran cantidad de trabajos que han sido analizados, y para facilitar su comprensión, hemos decidido agrupar estos algoritmos en una taxonomía (Fig. 3). Hay que destacar que estas categorías no son mutuamente excluyentes, ya que pueden existir operadores que por su naturaleza puedan ser ubicados en más de una de ellas.

A grandes rasgos, los algoritmos de extracción de características locales pueden ser separados en dos categorías principales: aquellos que realizan la extracción de esquinas sobre imágenes de contornos y los que extraen las características a partir de un análisis de la información presente en la imagen.

Los algoritmos pertenecientes a la primera familia, realizan un análisis de la curvatura de los contornos presentes en la imagen: algunos de estos métodos lo realizan teniendo en cuenta la información almacenada en el código de cadenas (forma de codificación de curvas digitales) o analizando cómo se comporta el gradiente en los puntos de bordes (edgels). Las esquinas son

seleccionadas en puntos donde los contornos tienen una curvatura significativa, la cual es calculada teniendo en cuenta una región que define la relevancia de la esquina. Existe un conjunto de métodos que se centran en la selección de la región de soporte óptima que permita una buena detección de esquinas con diferentes escalas presentes en la imagen. Otro grupo de algoritmos realizan el análisis multiescala basándose en el cálculo de la transformada wavelet para la extracción de bordes, esquinas y arcos. El principal tipo de características detectados por estos operadores son los puntos que corresponden a esquinas en la imagen.

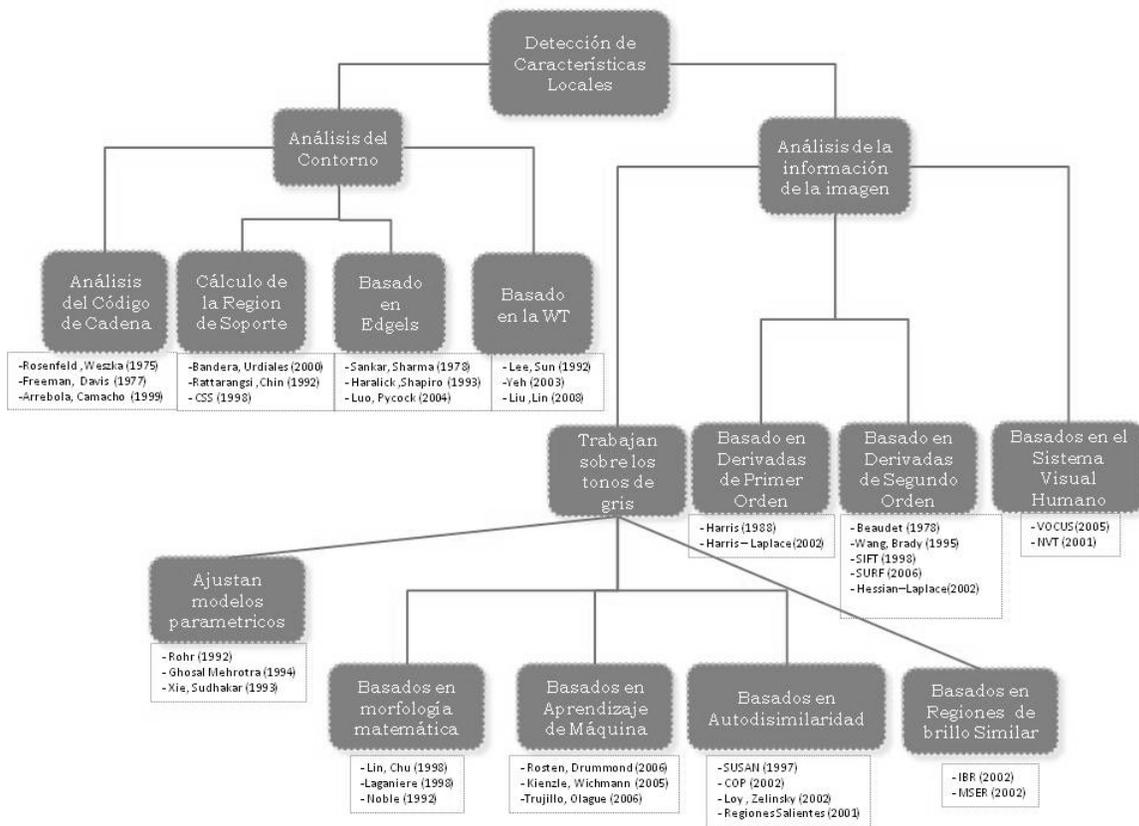


Fig. 3. Taxonomía de las diferentes familias de algoritmos de extracción de características locales

Por su parte, los algoritmos que trabajan con todo el contenido de la imagen son capaces de detectar además de estas otro tipo de características como son regiones de intensidad uniforme, regiones con alta entropía y esquinas formadas por diversos tipos de uniones. Pueden dividirse principalmente en los que utilizan geometría diferencial, basándose en las derivadas de primer y segundo orden. Otro grupo de métodos detecta las esquinas en base a cómo se comportan las distribuciones de los niveles de gris en regiones de la imagen. Estas vecindades pueden ser analizadas siguiendo la heurística de que las esquinas se encuentran definidas en regiones de alta disimilaridad, puesto que en ellas se produce un cambio significativo de intensidad en al menos dos direcciones (Métodos basados en la autodisimilaridad). Por el contrario, otro grupo de métodos se centran en encontrar regiones con muy baja autodisimilaridad y detectan regiones de intensidad homogénea (blob). Las vecindades de las esquinas pueden ser utilizadas para entrenar

clasificadores que sean capaces de reconocer esquinas en imágenes posteriores o pueden ser ajustadas a modelos de esquinas preestablecidos. Otro conjunto de detectores usan la morfología matemática en este sentido. Por último, se encuentran los sistemas basados en el sistema visual humano para detectar regiones altamente distintivas en las imágenes. Son de forma general sistemas más complejos, puesto que realizan la combinación de distintos tipos de rasgos.

A continuación realizaremos un análisis con un poco más de detalle de los principales métodos pertenecientes a cada una de estas familias.

2.1 Detectores basados en el análisis de la curvatura del contorno

Un borde en una imagen es una variación abrupta de la función de intensidad, la cual define una frontera entre dos regiones. Las esquinas se encuentran localizadas en la intersección entre dos o más bordes, reflejándose a menudo como un cambio bidireccional en la señal. Esta idea es utilizada por una familia de detectores para detectar puntos con alta curvatura utilizando los mapas de bordes de las imágenes. Estos métodos se centran en la detección precisa de la localización de los puntos. Originalmente estos métodos fueron aplicados a dibujos de línea en vez de a escenas naturales.

En los finales de la década de los 70's y durante los 80's y 90's fueron muy estudiados, sin embargo en la actualidad ha habido menos actividad en esta área, mientras que los métodos basados directamente en el análisis de los niveles de gris han ganado más fuerza. Esto se debe principalmente a que los objetivos de estos detectores son diferentes a los que típicamente tenemos hoy en día, ya que consideraban solo como relevantes a aquellos puntos que correspondían a esquinas verdaderas en 3D. Sin embargo, en la mayoría de las aplicaciones que utilizan los puntos de interés en la actualidad están interesados en detectar características robustas, estables y distintivas, sin tomar en cuenta si corresponden a esquinas verdaderas o no.

2.1.1 Detectores basados en el encadenamiento de bordes

Dada una imagen digital los primeros métodos geométricos para la detección de esquina en una imagen digital, realizaban primeramente una segmentación de la forma (ver Fig. 4.), extrayendo sus fronteras y codificándolas de acuerdo al *código de cadena*. A partir de aquí realizaban la búsqueda de los puntos principales a lo largo del borde. Este tipo de detectores presentan como principal desventaja una alta complejidad algorítmica, al verse involucrados muchos pasos intermedios en el procedimiento general. Además presentan una alta dependencia de los algoritmos de segmentación y codificación utilizados, siendo sensible a los errores introducidos en cualquiera de estos pasos.

El código de cadena es una forma de codificación de curvas digitales que representa un contorno como un punto de inicio y una secuencia de movimientos a través de los píxeles de bordes basados en 4 o 8-conectividad, la dirección de cada movimiento es codificado por $\{i | 0 \leq i \leq 7\}$ que representa las direcciones 45i. Este código puede ser visto como un conjunto de segmentos de líneas rectas encadenadas con una longitud y una dirección específica. En los primeros intentos de atacar el problema utilizando esta variante se concentraron en encontrar métodos para de forma eficiente y efectiva estimar la curvatura de los contornos.



Fig. 4. Imágenes de contornos sobre las que estos métodos son aplicados

Un enfoque común consiste en la utilización de cuerdas (línea recta entre dos puntos de una curva) para estimar tanto la pendiente como el ángulo que forma la curva en un punto. Estos métodos de forma general calculan la curvatura en cada punto, seleccionando como candidatos aquellos que se encuentran por encima de un umbral, y después realizan un proceso de eliminación de los no máximos para descartar aquellos puntos que no son máximos locales en un segmento lo suficientemente largo.

Para estimar la curvatura [12-13] trazan dos cuerdas en cada punto: una hacia adelante en la secuencia y otra hacia atrás y seleccionan como esquinas las posiciones donde el ángulo entre dos cuerdas alcanza un mínimo local. Este algoritmo puede dar resultados incorrectos cuando los bordes están muy cerca entre ellos [14]. Una versión modificada de este método es presentada por [15] donde utilizan el mismo criterio pero realizan primero el cálculo de una media local de las magnitudes de los ángulos. Alternativamente las esquinas son definidas como discontinuidades aisladas en la pendiente media de cuerdas entre segmentos sucesivos de la cadena [16-17]. Para esto desplazan una cuerda conectando las terminaciones de una secuencia de S enlaces a través de la cadena y toman las diferencias angulares entre segmentos sucesivos como medida de curvatura local, [18] siguen esta misma idea y realizan la selección tomando en cuenta el promedio de las pendientes en un subconjunto de puntos de la secuencia. [19] propone un método que es un caso particular de la diferencia de pendientes calculada entre cuerdas, y demuestra su efectividad para distinguir entre características de curvas y de esquinas, así como para determinar con precisión la localización de estos últimos, las transiciones entre líneas rectas y curvas y los centros de estas. Por su parte [20], reconoce como relevantes a aquellas localizaciones donde se pueden inscribir en la curva triángulos de un tamaño y un ángulo de apertura específico. En [21] se utiliza una variante diferente a las anteriores. Definen la deflexión tangencial y la curvatura de los contornos discretos basados en las propiedades geométricas y estadísticas asociadas a la estructura de valores y vectores propios de matrices de muestra extraídas a partir de los códigos de cadena.

Partiendo de que los números que codifican al contorno en el código de cadena corresponden aproximadamente a la pendiente de la curva, la curvatura aproximada puede ser encontrada utilizando diferencias finitas y las esquinas son detectadas identificando patrones específicos[22]. Los histogramas de los valores del código de cadena alrededor de los puntos candidatos son comparados utilizando la correlación cruzada normalizada y las esquinas son detectadas en los puntos de mínimo local[23]. Además, un estimado del valor de la pendiente puede ser encontrado utilizando histogramas circulares suavizados de códigos de cadena [24].

2.1.2 Métodos que se basan en el cálculo de la región de soporte

La mayoría de los algoritmos descritos anteriormente, reciben como entrada un conjunto de parámetros que van a definir la región de soporte (longitud del segmento de curva que se va a tener en cuenta para la selección de los puntos dominantes) sobre la que serán estimadas las propiedades locales de los puntos. Estos parámetros son seleccionados a priori de acuerdo al nivel de detalle presente en la curva, de forma general resulta complicado elegir un conjunto de parámetros capaces de lidiar con características de diferentes tamaños. Mientras mayor sea la región de soporte perderán importancia los rasgos más pequeños, mientras que por el contrario si se disminuye aumentará el número de características redundantes. De aquí se obtiene que para la detección de los puntos de interés no basta solo con disponer una buena medida de relevancia, sino que se debe estimar de forma precisa la región de soporte.

Existe un grupo de métodos que se basan en determinar de forma precisa esta región (la correspondiente escala de la característica) en vez de estimar la curvatura discreta de los contornos. Esto es realizado de forma independiente basado en las propiedades locales de la curva. Por ejemplo, el máximo local de las longitudes de las cuerdas puede ser utilizado para definir esta región, en la cual cada esquina dentro de esta debe tener una curvatura máxima[25]. Esta idea fue adoptada teniendo en cuenta el planteamiento de [26] de que cada punto dominante tiene su propia visión de la curva. Las esquinas pueden ser definidas también como el centro de una región de soporte con alto valor de curvatura media donde la región es grande y simétrica con respecto a su centro [27]. [28] presentó un método para determinar dicha región en blob convexos y sugirió que los no convexos pueden ser analizados descomponiéndolos en convexos. Las regiones libres de discontinuidades significativas en su curvatura (tomando en cuenta la diferencia entre la longitud de la curva entre dos puntos pertenecientes a esta y la distancia euclidiana entre ellos) alrededor de los puntos candidatos pueden ser utilizados junto con la curvatura para estimar los cambios de pendiente en la región [29] o el ángulo de la cuerda que une a las terminaciones de la región [30].

Un problema presente en la utilización del cálculo de una función de curvatura en los puntos del contorno es que dicha función es sensible al ruido presente en la señal, por lo que una variante común consiste en filtrarla para eliminar el ruido. El filtro gaussiano es el más comúnmente utilizado, sin embargo los métodos que lo utilizan deben lidiar con el problema de la selección de un factor de escala que garantice una región de soporte óptima. [31] presentan un método para la determinación de una función de suavizado en el cálculo de la curvatura óptima como la solución de un problema de optimización donde se busca minimizar el contorno óptimo suavizado sujeto a restricciones en la naturaleza del ruido.

Con vistas a solucionar los problemas existentes al tener una escala fija de suavizado, se han propuesto métodos que analizan una representación de la curva en el espacio de escalas. [32] extrae aquellos puntos de curvatura máxima que se mantienen estables en un rango de escalas. [33] analizan primeramente el comportamiento de los puntos de esquinas aislados y las esquinas dobles. El análisis realizado muestra que el espacio de escala contiene patrones de líneas que continúan, se mezclan o terminan. Esta información puede ser vista en forma de árbol que representa de manera simple y concisa la información de las esquinas en múltiples escalas. Las ramas del árbol que son mayores (en escala) que las ramas padres son consideradas como puntos de esquina estables. La escala natural de las curvas fue también estudiada por [34] para evitar la representación exhaustiva sobre todo el rango de escalas. Un marco teórico para un mecanismo automático de selección de escala fue propuesto por [35].

En vez de realizar el suavizado de la curva en múltiples escalas e integrar la información a través de estas [36] proponen un método para realizar el suavizado de una curva de forma adaptativa basado en el nivel de detalle del contorno, para lo cual utilizan la varianza local de la curvatura. De esta forma no se requiere la construcción del espacio de escala completo y su representación por medio de un árbol. El análisis de la curvatura en el espacio de escala (CSS por sus siglas en inglés) [37] realiza un proceso de detección de candidatos a esquinas en las escalas más altas (son seleccionados los puntos de curvatura máximas locales, donde el punto elegido debe ser significativamente mayor que sus vecinos) y les hace un seguimiento a través de las escalas menores para mejorar la localización. De esta forma, en una escala la función de curvatura solo debe ser calculada en los pequeños vecindarios de las esquinas en vez de cada punto del contorno. [38] también utilizan el CSS junto con un procedimiento para la selección de la región de soporte de forma dinámica (basado en la curvatura local en vez del valor global, utilizado por los métodos tradicionales que realizan el análisis en el CSS) y extraen aquellos puntos donde la curvatura disminuye rápidamente hacia el mínimo más cercano y el valor de curvatura máxima del vecindario es pequeño.

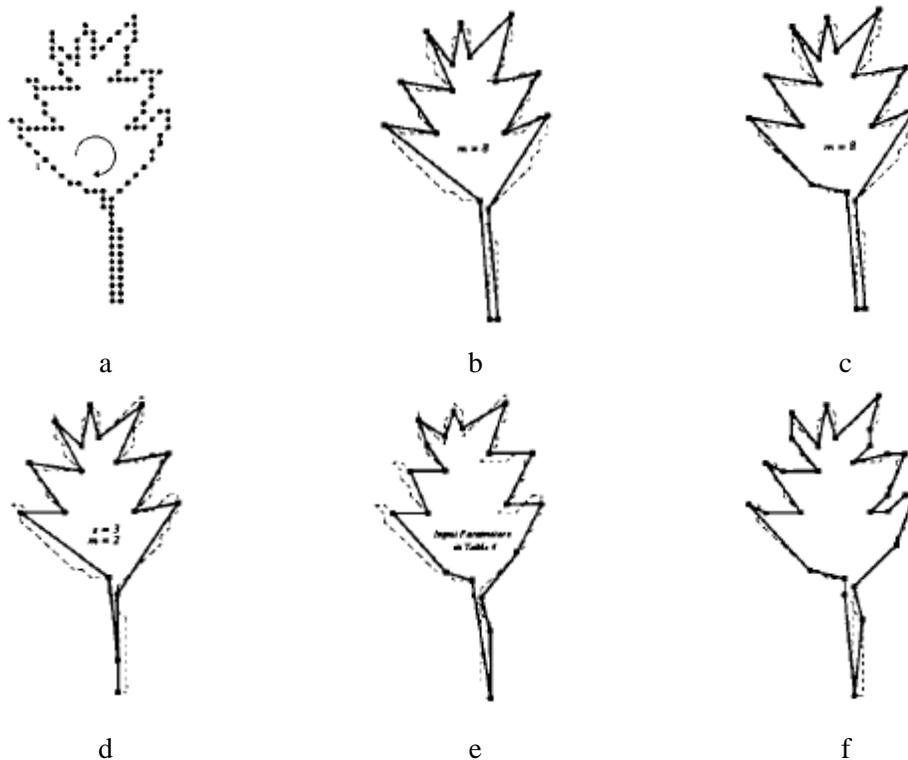


Fig. 5. Reconstrucción de imágenes utilizando los puntos de interés detectados por los siguientes métodos sobre imágenes de contorno. (a) Imagen original. (b) [12]. (c) [15]. (d) [17]. (e) [21]. (f) [25]

Otro conjunto de métodos realizan la extracción de puntos característicos mediante una parametrización de los contornos utilizando B-splines y las esquinas son detectadas en los puntos con un valor alto de la segunda derivada donde el spline tiene una alta desviación del punto de control [26, 39]. Otros autores realizan una aproximación por partes de la curva sujeto a ciertas restricciones [40-41], y detectan los puntos que corresponden aproximadamente a las

intersecciones reales o extrapoladas de los segmentos de línea adyacentes en el polígono. [42] estiman los parámetros de dos líneas ajustadas a los segmentos vecinos del punto candidato a esquina, y es seleccionado si estos son significativamente diferentes entre sí desde el punto de vista estadístico.

2.1.3 Detectores basados en Edgels

Los detectores pertenecientes a este grupo hacen uso principalmente de operadores de la geometría diferencial, tomando como punto de partida a los algoritmos de detección de bordes. La familia de detectores analizada anteriormente basaba su funcionamiento en el código de cadena como una forma de representación de las curvas digitales y buscaban los puntos de mayor curvatura para ser seleccionados como relevantes. En esto se centran las medidas de importancia de un punto, que es expresada como una función de la fortaleza del contorno y/o de la dirección del gradiente, la cual es calculada usualmente en una vecindad local de los puntos de bordes (denominados como edgels en la literatura).

[43] fueron los pioneros en este tipo de algoritmos al introducir un método que no necesitaba parámetros y tenía la ventaja de poder lidiar con características de múltiples tamaños. En vez de intentar calcular la curvatura en cada punto del contorno, asignaban un valor de los posibles 1,-1 o 0, si la curvatura era positiva, negativa o 0 respectivamente. Esto fue realizado de manera local en una vecindad de 3x3. El algoritmo iterativamente calcula un promedio de las curvaturas de los puntos con el de sus vecinos inmediatos, seleccionando en los valores máximos las esquinas. Teniendo en cuenta de manera simultánea tanto los bordes como su dirección, la vecindad de los edgels es comparada con regiones a ambos lados de este en la dirección del contorno para encontrar puntos con baja autosimilaridad [44]. Otra variante consiste en sustituir cada punto de borde por un segmento de línea y definir una medida de fortaleza de la presencia de esquinas basado en el cambio de la dirección del gradiente en estos segmentos [45].

[46] encuentran un vector potencial a partir de establecer una analogía en el cual el mapa de bordes de Canny es visto como una densidad de flujo elemental residente en el plano de la imagen. Mediante la rotación de los vectores gradientes de los bordes obtienen los vectores tangentes a estos, los cuales son suavizados teniendo en cuenta el inverso de su distancia al punto de la imagen en cuestión. Se refieren a este campo como vector potencial. De acuerdo a esta representación obtienen que el máximo local de la magnitud de este vector que exhibe consistencia direccional corresponde a los bordes, por otra parte los mínimos locales que mantienen a su vez consistencia direccional corresponden a las líneas simétricas. En su trabajo extienden esta representación topográfica para detectar esquinas que son definidas como los puntos de ensilladura donde estas estructuras se interceptan. Para detectar estos puntos utilizan una variante de comparación de modelos en la cual buscan las estructuras de valles y crestas consistentes en la dirección del vector potencial y en la ortogonal a esta respectivamente. Este mismo autor en un trabajo posterior [47] definen una representación universal para los distintos tipos de esquinas basados en los momentos del gradiente en el espacio de escala.

Otra familia de detectores utiliza la transformada de Hough en la detección de los contornos de la imagen que corresponden a líneas rectas y encuentran donde éstas se interceptan para seleccionar las esquinas. [48] utiliza la transformada generalizada de Hough definida por [49] y sustituye cada edgel por un segmento de línea, seleccionando aquellas posiciones donde estos segmentos se interceptan. Esta transformada es utilizada para la detección de ocurrencias de una forma arbitraria en una imagen mediante la generalización de este procedimiento. Se utiliza la

información de los bordes para definir una correspondencia entre la orientación de un edgel a un punto de referencia de la forma. Este punto de referencia puede ser visto como el origen de un sistema de coordenadas locales y establecer una forma de calcular una medida que describa cuán bien un punto de la imagen puede funcionar como origen de la forma en cuestión. [50] detectan las esquinas como las líneas rectas que pasan a través de los orígenes de estos sistemas de coordenadas locales.

Otros detectores de bordes presentan algunos problemas en el marcado de los bordes alrededor de las esquinas, por esa razón los autores plantean la utilización de la magnitud del gradiente y un análisis del nivel de gris en un contexto local para la detección de los bordes, en vez de utilizar operadores para la detección global. Este proceso lo llevan a cabo de manera dinámica y de esta manera se obtienen puntos de bordes correctos para poder aplicarles la transformada de Hough. Una idea similar es seguida por [51] que usan la transformada en la detección de líneas rectas en el mapa de bordes, a la vez que utilizan la transformación inversa para encontrar las intersecciones.

2.1.4 Detectores basados en la Transformada Wavelet (WT)

Además del filtrado gaussiano, los puntos de esquinas pueden ser determinados de forma invariante ante las diferencias de escala utilizando la transformada wavelet (WT) sobre la orientación del contorno. En este sentido han sido utilizadas tanto la información de los extremos locales [52] como el módulo del resultado de la WT [53-54] para detectar arcos y esquinas de manera efectiva, utilizando la magnitud de la pendiente para diferenciar entre ambos. Este método resulta más eficiente que los métodos tradicionales de detección de esquinas multiescalas porque solo extraen la WT para una o dos escalas.

[55] proponen un detector de bordes basados en la razón entre el módulo de la WT interescalas para detectar de manera confiable los bordes que se encuentran cercanos a las esquinas. A partir de aquí y utilizando la varianza de la orientación son eliminados los puntos que no satisfacen las condiciones de esquina e introducen una nueva medida basada en las propiedades de invarianza de la orientación ante los cambios de escala. [56] por su parte, realiza un análisis de la descomposición de valores singulares para la selección de la escala global natural en la WT discreta, definiendo esta como el nivel asociado con el valor propio más prominente (dominante), y su vector propio correspondiente es considerado como la escala óptima. El valor máximo del módulo de la WT sobre diferentes escalas es utilizado por [57] para definir la escala natural local de los candidatos a esquinas y es tomada como medida de relevancia. [58] analiza los vectores propios de las matrices de covarianza de las coordenadas de los puntos de borde sobre un segmento pequeño junto con la WT para eliminar la respuesta ante los arcos. Los contornos 2-D de los objetos son representados inicialmente por la tangente 1-D obtenida a partir de estos vectores propios. Basándose en la suposición de que en las esquinas se produce una fuerte variación de la tangente, se utiliza la WT para descomponer los ángulos de la tangente y capturar las variaciones irregulares. Estos métodos no pueden ser aplicados directamente sobre la imagen del gradiente debido principalmente a que se encuentran basados en la WT 1D. [59] introducen un método novedoso para la extracción de esquinas que utiliza la WT 2D y un conjunto de umbrales predeterminados.

2.2 Detectores basados en la información de la imagen

La suposición de que las esquinas se encuentran localizadas sobre los bordes resulta un modelo inadecuado para el análisis de las regiones de texturas en imágenes y para la extracción de puntos aislados, además que presenta problemas en la detección de esquinas tipo uniones. Teniendo en cuenta esto, un gran número de detectores operan directamente sobre la imagen en tonos de gris, sin tener que detectar previamente los bordes. Estos métodos se basan en condiciones débiles por lo que pueden ser utilizadas sobre un amplio rango de imágenes. La mayoría de los métodos pertenecientes a esta familia utilizan las derivadas de primer y segundo orden, mientras que otros utilizan heurísticas para encontrar regiones de alta variación.

2.2.1 Detectores basados en la información de los tonos de gris

Este tipo de métodos trabaja directamente sobre los niveles de intensidad de gris de las imágenes. Utiliza heurísticas para estimar cómo deben comportarse las vecindades de las esquinas. Estos métodos son los más eficientes desde el punto de vista computacional, ya que no involucran el cálculo de derivadas. Como principal desventaja presentan la sensibilidad al ruido (incluso mayor que el presentado por los métodos diferenciales), debido principalmente a que sus operaciones no involucran la utilización del desenfoque gaussiano.

2.2.1.1 Morfología Matemática en la detección de bordes

Una familia importante de métodos utiliza la morfología matemática en la detección de los puntos de alta curvatura. Los primeros trabajos en esta área [60-61] definen operaciones morfológicas para la extracción de picos (esquinas convexas como resultado de la diferencia entre la imagen y la apertura de la misma) y valles (esquinas cóncavas detectadas utilizando la diferencia entre la cerradura de la imagen y la imagen), utilizando discos como unidades estructurales. [62] utiliza también estos operadores para mejorar la localización de las esquinas detectadas, combinando la morfología con el encadenamiento de bordes. Esta variante requiere que la imagen de entrada sea binaria. [63] introduce la utilización de unidades estructurales de diversos tipos (Fig. 6) y define un operador llamado cerrado asimétrico que consiste en la dilatación de la imagen utilizando un elemento estructural dado, seguido por la erosión para lo cual utilizan un elemento distinto. Establecen una medida de fortaleza de la esquina basada en la diferencia entre las imágenes resultantes de aplicar este operador con diferentes unidades estructurales. Aunque la combinación de operadores propuestos se comporta bien en presencia de rotación, este método no es rotacionalmente invariante puesto que solo considera las direcciones vertical, horizontal y las diagonales. Otra desventaja presente en esta variante es la respuesta múltiple ante una esquina que en muchas aplicaciones no es deseable.

[64] investigaron las propiedades del filtro de apertura morfológica en lugar del filtrado gaussiano para la construcción del espacio de escala. En su trabajo demuestran que este filtro no introduce nuevos cruces por cero al pasar hacia una escala más gruesa, y presentan como ventajas con respecto al suavizado gaussiano que este operador evita que se mezclen cruces por cero, vecinos, además de ser más eficiente computacionalmente.

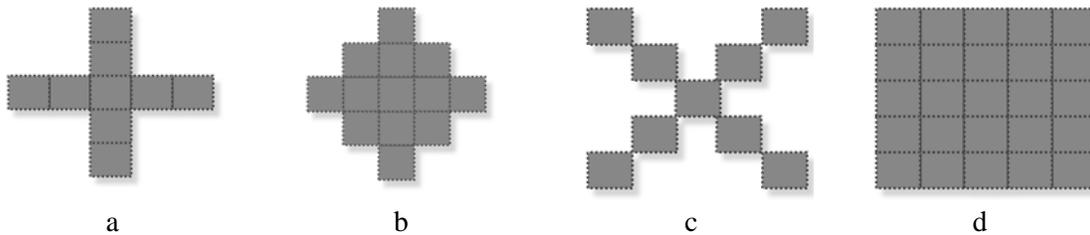


Fig. 6. Elementos estructurales utilizados en la detección de bordes morfológicas. (a) Elemento estructural +. (b) Elemento estructural rombo. (c) Elemento estructural X. (d) Elemento estructural cuadrado.

2.2.1.2 Métodos basados en la autodisimilaridad

Este tipo de métodos se basan en la heurística de que la vecindad de una esquina es una región no-homogénea, puesto que en ella se produce una variación de los tonos de gris en al menos dos direcciones. Esta idea puede ser generalizada para seleccionar puntos que cumplan con esta condición. Uno de los primeros trabajos en esta área es el presentado por [65] quienes indican que las esquinas cumplen la condición de ser significativamente diferentes al valor de la mediana en su vecindario. Sin embargo, este método no puede lidiar con uniones más complejas o con bordes suavizados. [66] seleccionan como esquinas aquellos puntos donde arcos concéntricos contiguos de píxeles son significativamente diferentes con respecto al píxel central.

[67] miden esta disimilitud teniendo en cuenta la proporción de píxeles en una vecindad circular que es similar al núcleo de esta (píxel central). Este constituye el principio USAN (univalue segment assimilating nucleus). Las esquinas son definidas como puntos SUSAN (smallest USAN, es decir mínimos locales). Los píxeles pertenecientes a cada vecindad son particionados en dos grupos: aquellos que tienen un valor similar al núcleo y los que no. Del análisis realizado por los autores se pueden extraer las siguientes reglas para clasificar cada región (Fig. 7):

- **Regiones homogéneas:** El área local de similar brillo cubre casi la totalidad de la ventana circular.
- **Región de bordes:** La proporción entre la cantidad de píxeles con valor similar al núcleo de los que no lo tienen se encuentra alrededor de un 50%.
- **Región de esquina:** La proporción anterior cae hasta un 25%.

Para hacer este método más robusto se le da un mayor peso a los píxeles más similares al núcleo. Adicionalmente utilizan un conjunto de reglas para eliminar las características “malas”. Este detector presenta buenos resultados de detección en imágenes rotadas, pero no es invariante al cambio de escala. Otro problema de este operador es que detecta puntos sobre los bordes que no son esquinas los cuales son sensibles a ruido y por tanto son menos discriminativos.

[68] definen un operador llamado COP (Crosses as Oriented Pair), el cual utiliza el promedio local de los USAN calculados sobre un par de cruces orientados, para estimar las direcciones probabilísticas dominantes y definen diferentes tipos de esquinas con más de una dirección dominante. Este método es capaz de lidiar con uniones múltiples tipo **X**, **Y**, y **T** con un costo computacional bajo. Una idea similar es explotada por [69] donde definen una medida de la fortaleza de la esquina en función de las diferencias entre el píxel central y las terminaciones de los diámetros del círculo trazados en varias direcciones. Si este valor es pequeño para cualquier orientación posible entonces el punto no es una esquina. Esto es computacionalmente eficiente porque dado que el proceso puede ser detenido en el momento en que un pequeño valor es

encontrado. Una prueba similar es realizada por [7, 70] para seleccionar los candidatos a esquinas, y estos son clasificados definitivamente como esquinas si existe un conjunto de n píxeles contiguos en una vecindad circular que son más claros o más oscuros que el valor del centro. [71] introducen ligeras modificaciones para obtener una mayor eficiencia computacional. Partiendo de la vecindad circular de un píxel, eliminan los cercanos a este que tienen un valor de gris similar y son diametralmente opuestos, utilizando los sobrevivientes para estimar el Laplaciano como la diferencia entre la posición central y los píxeles del círculo. Esta operación es realizada sobre los primeros octavos de la pirámide generada para hacer este método robusto a los cambios de escala.

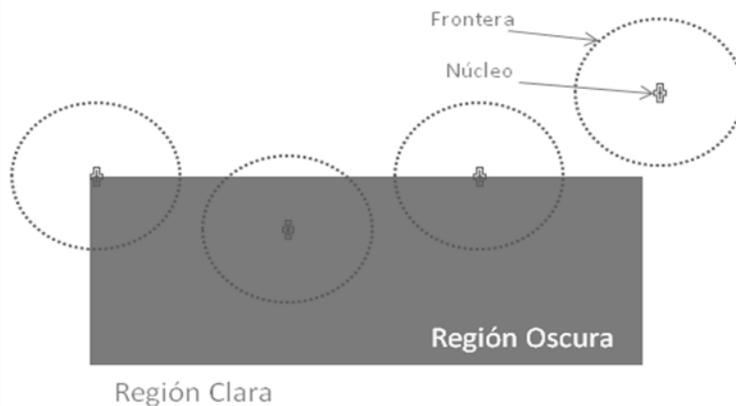


Fig. 7. Principio de funcionamiento del operador SUSAN. En las esquinas la proporción de píxeles similares al núcleo es alrededor del 25% del tamaño del disco.

Una transformada rápida de simetría radial es realizada en [72] para detectar los puntos de mayores valores donde se cumple que el gradiente es fuerte y radialmente simétrico y tiene signo uniforme alrededor del radio. A partir de cada imagen se generan las proyecciones de orientación O_n y magnitud M_n mediante el análisis del gradiente g en cada punto p a partir de los cuales se extraen los píxeles afectados positivamente P_{+ve} (píxeles que son apuntados por la dirección de g y se encuentran a una distancia n de p) y los afectados negativamente P_{-ve} (píxeles se encuentran opuestos a la dirección de g y a distancia n). Para cada par de píxeles afectados, los correspondientes P_{+ve} en O_n y M_n son incrementados en 1 y en $\|g(p)\|$ respectivamente, mientras que para P_{-ve} se realiza de manera similar aunque decrementando. Este operador, además de las esquinas detecta puntos que se encuentran sobre los bordes, dado que analiza las diferencias en la dirección del gradiente. La escala de las características puede ser ajustada seleccionando un valor adecuado de n , permitiendo la detección de características específicas cuando a priori se conoce su escala.

Otra variante se basa en la propuesta de [73], el cual define la relevancia en términos de la complejidad o la incertidumbre de la señal, medida utilizando la entropía de Shanon. La idea consiste en seleccionar las regiones con un valor de entropía alto, la cual es calculada sobre el histograma del vecindario. En [74], incorporan un criterio adicional: la auto-disimilaridad

medida en el espacio de escala para obtener un método robusto ante esta transformación. Esta medida es calculada para los candidatos a características utilizando (1).

$$W = \frac{s^2}{2s-1} \sum_I \left| \frac{\partial p(I; s)}{\partial s} \right| \quad (1)$$

donde $p(I)$ es estimada empíricamente basado en la distribución de los niveles de gris en el histograma de una vecindad de radio s . Una extensión de este detector invariante hacia las transformaciones afines fue propuesto en [75], en el cual el máximo sobre el espacio de escala y los parámetros de la forma elíptica que delimita a la región afín, son buscados simultáneamente.

2.2.1.3 Detectores basados en modelos paramétricos

Varias técnicas asumen que las esquinas tienen la apariencia general de una o más regiones de intensidad uniforme en un fondo de una intensidad diferente y uniforme a su vez. Basado en esta suposición, muchos métodos definen modelos paramétricos para la detección de las esquinas. Una de las ventajas principales de este tipo de variantes consiste en que de esta forma es posible extraer la información completa de la esquina (parámetros del modelo) lo cual es deseable en muchas aplicaciones.

Una esquina puede ser modelada como un borde [76] o una familia de bordes [77] rectos difusos analizados en una vecindad de puntos con alta variación en los tonos de gris. Dependiendo del número de estas líneas se ajusta a la imagen un modelo paramétrico utilizando Mínimos Cuadrados para identificar el tipo de unión. El modelo puede incluir el ángulo, la orientación, el contraste y la curvatura, entre otros [34]. [78] desarrollan un modelo paramétrico para caracterizar la función de intensidad local en una imagen y realizan una proyección sobre un conjunto de momentos ortogonales de Zernike para generar los polinomios que serán utilizados en el estimado de los parámetros.

[79] proponen un método de detección de esquinas basado en el principio de conservación del nivel de gris. A partir de la imagen binaria y una ventana circular, los puntos de esquinas son estimados a partir de la relación geométrica entre el centroide y los tres vértices de la cuña formada en la ventana. Estos puntos son tomados como aproximaciones iniciales en la solución de un sistema de ecuaciones no lineales que se construye a partir de la conservación del área y los tres momentos de masa de la cuña parcial delimitada por la ventana circular. De esta forma se puede obtener un estimado más preciso de la localización de la esquina, del cual también se puede extraer la información sobre la orientación y el ángulo que forma. De manera similar en [80] máscaras de convolución pueden ser derivadas para varias cuñas las cuales optimizan la razón de ruido en la señal y el error de localización, bajo la hipótesis de la presencia de ruido gaussiano [81]. La esquina es descrita por el producto entre el seno en la dirección x por la exponencial en la dirección y y en una porción de la máscara, y por el producto del seno en las dos direcciones de la porción restante. [82] se plantean la detección de esquinas como un problema de optimización, definiendo la función de costo de acuerdo a las características deseables de las esquinas como la curvatura y la disimilaridad entre regiones, utilizando calentamiento simulado (simulated annealing) para minimizarla. La estructura geométrica de las esquinas, así como la variación del nivel de gris son tenidas en cuenta para evaluar la calidad de las configuraciones de esquina.

2.2.1.4 Detectores basados en el aprendizaje de máquina

El principal problema de los métodos descritos anteriormente consiste en extraer a partir del valor de gris una medida que represente la similaridad del píxel a una esquina. Una alternativa es considerar la detección de esquinas como un problema de reconocimiento de patrones, entrenar un clasificador en el modelo y entonces aplicar el clasificador sobre las imágenes de entrada. Por ejemplo, la habilidad de las redes neuronales de adaptarse y aprender con su paralelismo inherente puede ser aprovechada para resolver estos problemas. Un perceptrón multicapa puede ser entrenado con ejemplos de esquinas y aplicado sobre la imagen después de cierto procesamiento [83-84]. El clasificador resultante obtiene un estimado de las probabilidades a posteriori, en contraste con los métodos tradicionales que producen etiquetas binarias sujetas a un conjunto de umbrales ajustados empíricamente.

[85] se basan en el detector presentado en sus trabajos anteriores [6-7] para entrenar un árbol de decisión en el reconocimiento de las configuraciones de esquina en términos de las proporciones de píxeles más claros, más oscuros o similares al píxel central en una vecindad circular. De esta manera obtienen una generalización del detector que funciona a una mayor eficiencia que la mayoría de los detectores existentes, alcanzando altos niveles de repetitividad bajo diferentes condiciones de observación y diferentes tipos de rasgos. Aunque presenta como desventajas la sensibilidad al ruido y puede manifestar respuestas a pequeñas líneas.

La percepción humana puede ser utilizada en lugar de un modelo [86]. Las imágenes son mostradas a un número de sujetos de prueba y aquellas localizaciones consistentes (como una medida del seguimiento de ojo) son seleccionadas como relevantes y se entrena una máquina de soporte vectorial para reconocer estos puntos.

Otro conjunto de autores se plantean la detección de puntos como un problema de optimización y presentan una solución basada en el aprendizaje de máquina utilizando programación genética (PG). [87] intenta evolucionar el operador de [88] a través de la PG y reportan una mejora de un 15% con respecto al original. En un trabajo posterior [89] presenta un operador evolutivo optimizado para computar el flujo óptico en una secuencia de imágenes particular, pero a pesar de los resultados prometedores alcanzados, el criterio de optimización utilizado no garantiza la generalización para otras aplicaciones donde se requiera la detección de puntos.

[90-91] proponen una solución utilizando la PG basada en maximizar un conjunto de propiedades que los puntos de interés deben cumplir. Como resultado obtienen dos operadores, el primero de los cuales IPGP1 presenta una estructura simple y consiste en extraer las altas frecuencias de la imagen mediante la sustracción de la versión suavizada de la original, mientras que un segundo paso suaviza las altas frecuencias extraídas. Éste es prácticamente el mismo principio seguido por el operador de Diferencias de Gaussianas (DoG). El segundo operador IPGP2 representa una versión modificada del operador de [92]. En un trabajo posterior, presentan la generalización de IPGP1 para lidiar con los cambios de escala [93].

2.2.1.5 Algoritmos que extraen regiones de intensidad homogénea

La mayoría de los métodos de detección de características locales que trabajan directamente sobre los niveles de intensidad de gris de la imagen, extraen, en su mayoría puntos que corresponden a la localización de las esquinas. La mayoría de estos métodos se comportan de forma robusta ante la rotación, pero no presentan la misma invarianza ante los cambios de escala

y transformaciones afines. Esto se debe principalmente a que estos operadores son sensibles al desenfoque (se basan en el análisis de las distribuciones de los tonos de los píxeles en vecindades circulares, o en el ajuste a modelos de esquina ideales), y la mayoría de los métodos tradicionales invariantes a escala construyen un espacio de escalas utilizando el desenfoque gaussiano aumentando progresivamente la varianza. Además, el análisis multiescala presenta más sentido en características tipo blob, ya que las esquinas (intersecciones de bordes) se comportan relativamente estables en un amplio rango de escalas. Es por ello que los principales métodos invariantes a escala y transformaciones afines que trabajan directamente con los niveles de gris se concentran en determinar características tipo blob, utilizando técnicas basadas principalmente en la segmentación de regiones estables con valor de intensidad homogénea.

Precisamente esta es la idea que se encuentra tras el algoritmo de Regiones Extremales Maximalmente Estables (MSER) propuesto por [94], que detecta regiones conexas en la imagen que presentan un nivel de gris similar. La palabra maximal se refiere a que la intensidad de la región es mayor o menor que la de los píxeles del borde exterior de esta. Estas regiones se comportan estables con respecto a los cambios monotónicos de la intensidad en la imagen, así como las transformaciones geométricas contiguas preservan la topología (las componentes conexas no se separan). Este operador detecta áreas de la imagen donde la binarización local se mantiene estable sobre un rango de umbrales, y las regiones detectadas de esta manera se comportan estables ante diferentes transformaciones afines. Alternativamente, regiones afines pueden ser definidas a partir de los puntos que se mantienen estables en el contorno de las MSER [95]. Este detector es actualmente el más eficiente entre los existentes en la literatura y presenta un desempeño similar e incluso superior a sus contrapartes basados en la estimación de la forma a partir de la matriz de segundo momento, [96], aunque presenta limitaciones en presencia de imágenes desenfocadas. En un trabajo posterior [97] se atacan estas limitantes, mediante el análisis de la estabilidad de los píxeles pertenecientes a la frontera de las regiones en vez de a la región completa.

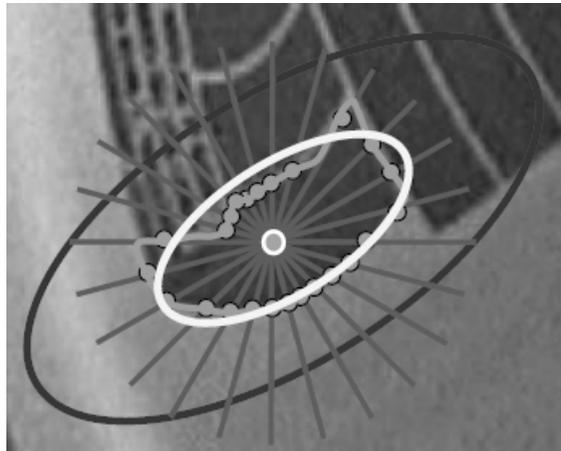


Fig. 8. Principio de funcionamiento del método IBR

[98-99] definen un operador que detecta los valores extremos de intensidad en un espacio de escala y a partir de éstos, traza rayos en diferentes direcciones sobre una vecindad circular, evaluando una función sobre cada rayo (2).

$$f(t) = \frac{abs(I(t) - I_0)}{\max \left(\frac{\int_0^t abs(I(t) - I_0)}{t}, d \right)} \quad (2)$$

donde t es un parámetro arbitrario que se encuentra sobre el rayo, $I(t)$ el valor de la imagen en t , I_0 el valor de intensidad en el extremo y d es un número pequeño para evitar las divisiones por 0. La posición donde esta función alcanza un extremo es invariante a las transformaciones afines. De manera general este máximo es alcanzado en lugares donde la intensidad cambia bruscamente. Finalmente, los puntos invariantes detectados sobre cada rayo son enlazados para delimitar una región invariante. Este método produce regiones similares al MSER con la desventaja de ser más costoso computacionalmente (Fig. 8).

2.2.2 Detectores basados en las derivadas de primer orden

La detección de características locales basadas en las derivadas de primer orden se encuentran entre los más utilizados en diversas aplicaciones. Un detector de esquinas que selecciona como puntos relevantes aquellos donde se produce un máximo local de una medida de varianza direccional fue introducido por [88] en el contexto del cálculo de trayectorias de robots móviles. Esta variación de la intensidad es estimada colocando una pequeña ventana cuadrada sobre la imagen y desplazándola en las 8 direcciones principales, tomando como medida la suma de las diferencias cuadradas entre los píxeles correspondientes en las dos ventanas (Fig. 9).

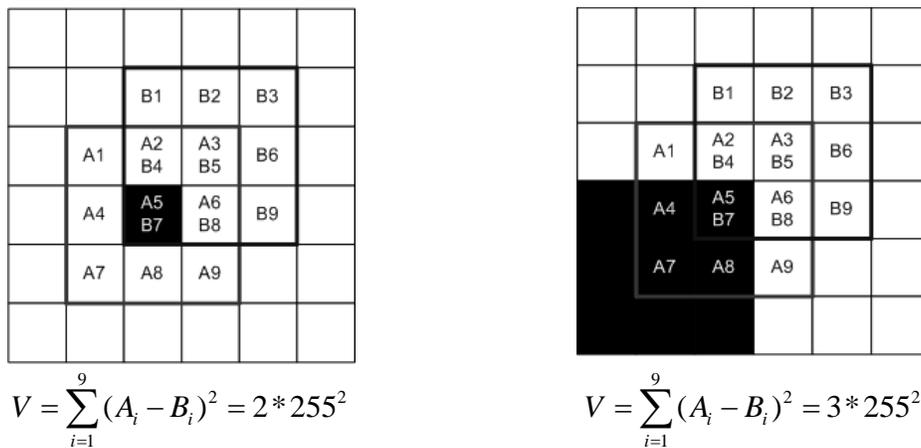


Fig. 9. Ejemplo del funcionamiento del el detector de [88]

Si el píxel central se encuentra interior a un objeto (o en el fondo) la intensidad se debe comportar relativamente constante en la ventana, por tanto la variación producto del desplazamiento en cualquier dirección debe ser pequeña; por otra parte, si la ventana se encuentra sobre un borde y es desplazada en dirección perpendicular a este habrá una variación significativa, por el contrario de si se desliza en la dirección del borde; mientras que si se

encuentra sobre una esquina o sobre un píxel aislado, se produce una alta variación independientemente de la dirección en que se desplace la ventana.

Este operador presenta una respuesta anisotrópica, ya que la variación de intensidad solo es calculada teniendo en cuenta un conjunto discreto de desplazamientos. Para superar esta dificultad [100] definen una función para medir las variaciones de la intensidad en todas las direcciones, esto lo logran mediante una expansión analítica del operador de [88]. De esta manera obtienen la *matriz de segundo momento* o *matriz de autocorrelación* (3), basada en las derivadas de primer orden calculadas con máscaras de Sobel en una ventana gaussiana, para explorar las propiedades estadísticas de las variaciones de intensidad.

$$M = \begin{bmatrix} I_x^2(x, \sigma_D) & I_x(x, \sigma_D)I_y(x, \sigma_D) \\ I_x(x, \sigma_D)I_y(x, \sigma_D) & I_y^2(x, \sigma_D) \end{bmatrix} \quad (3)$$

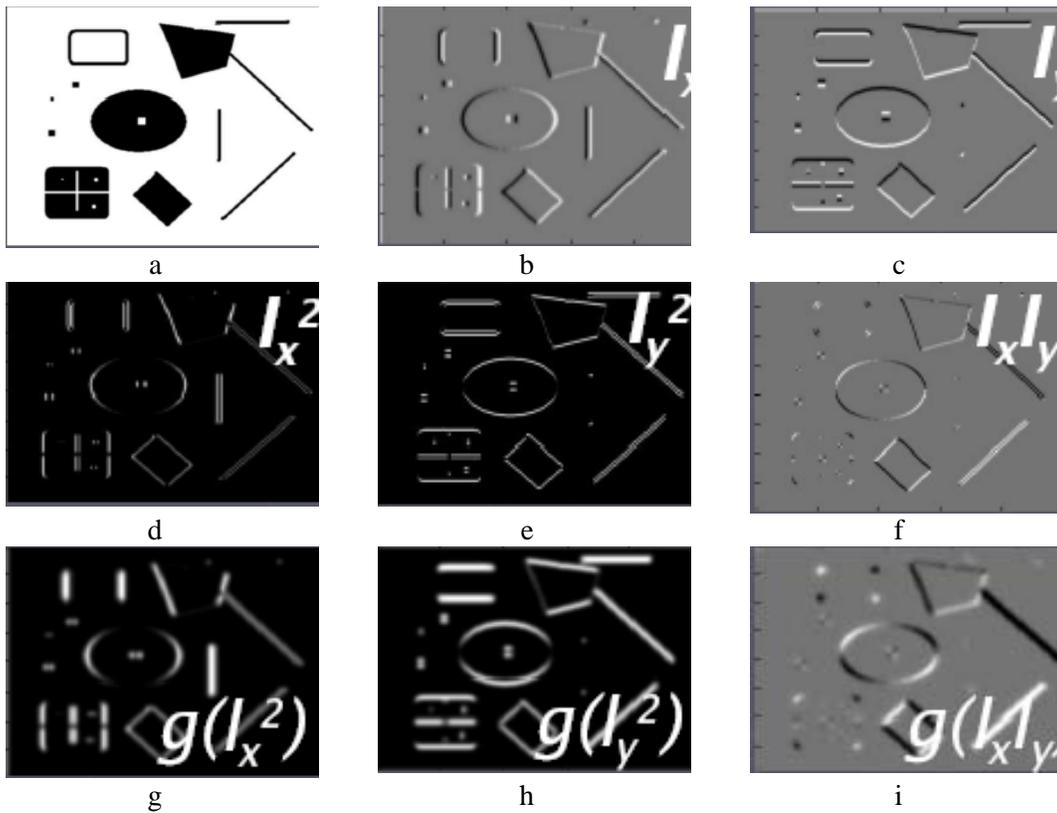




Fig. 10. Método de Harris paso a paso. (a) Figura original. (b) y (c) derivadas de la imagen en las direcciones x y y respectivamente. (d) - (f) componentes de la matriz de autocorrelación. (g) - (i) desenfoco gaussiano sobre los componentes de la matriz de autocorrelación. (j) Imagen resultante con las esquinas resaltadas

Los valores propios de M representan los dos cambios de direcciones principales en la vecindad de un punto, lo cual es proporcional a la curvatura principal de la superficie de la imagen y forman un detector rotacionalmente invariante (Fig. 10). Establecen una medida para estimar la fortaleza de las esquinas que viene dado por (4) :

$$C(x, y) = d \operatorname{tr}(M) - k(\operatorname{traza}(M))^2 \quad (4)$$

Basado en esto, las esquinas pueden ser detectadas en posiciones donde M tiene ambos valores propios grandes, en otras palabras, en posiciones donde la señal de la imagen varía significativamente en dos direcciones ortogonales. Estudios comparativos realizados [101-102] demuestran que este es uno de los detectores con mayor tasa de repetitividad y mas informativos. Este detector ha sido estudiado por [103] y demostrado que es óptimo para uniones tipo L.

Otro conjunto de medidas han sido propuestas en la literatura basadas en la matriz de segundo orden: [104] utiliza la razón entre el determinante y la traza; por su parte [105] utiliza solo el determinante de M . Suponiendo que la imagen está sujeta a deformaciones afines, del análisis realizado por [106] se desprende que es más conveniente utilizar el menor valor propio de M como medida de fortaleza de la esquina. Este conjunto de métodos han sido utilizados con éxito, pero tienen la desventaja de que se encuentran basados en heurísticas, recientemente [107] evita la utilización de heurísticas y realiza la detección de esquinas basado en el condicionamiento de puntos con respecto a una ventana de comparación bajo diferentes transformaciones como translación, rotación y cambios de escala. Un análisis comparativo realizado en [108-109], muestra que estos cuatro operadores son equivalentes entre ellos y al de [100] con respecto a varias normas de matrices. Otras modificaciones han sido introducidas, como por ejemplo la utilización de derivadas gaussianas [110] y combinaciones de derivadas de primer y segundo orden [111].

2.2.3 Detectores basados en las derivadas de segundo orden

La matriz Hessiana (5) es obtenida a partir de la expansión de Taylor de la función de intensidad de la imagen $I(x)$, y los elementos que la forman son las derivadas gaussianas de segundo orden. Esta matriz describe importantes propiedades de la estructura local de la imagen.

$$H = \begin{bmatrix} I_{xx}(x, \sigma_D) & I_{xy}(x, \sigma_D) \\ I_{xy}(x, \sigma_D) & I_{yy}(x, \sigma_D) \end{bmatrix} \quad (5)$$

Uno de los primeros detectores [92] utiliza el determinante (6) y el Laplaciano (traza) (7) de esta matriz y define a las esquinas como los extremos locales en ellas (Fig. 11).

$$Det_H = |I_{xx}I_{yy} - (I_{xy})^2| \quad (6)$$

$$Lap_H = I_{xx}(x, \sigma_D) + I_{yy}(x, \sigma_D) \quad (7)$$

Los filtros basados en el *Laplaciano* tienen una desventaja significativa en el contexto de la extracción de blob: los máximos locales se encuentran frecuentemente en la cercanía de bordes rectos donde el cambio en la intensidad de la señal se produce en una sola dirección [112]. Estos rasgos son menos estables porque su localización es más sensible al ruido o a pequeños cambios en el vecindario de la textura. Una variante más sofisticada consiste en seleccionar los puntos donde Det_H y Lap_H alcanzan máximos de manera simultánea. De esta forma son resaltados aquellos puntos donde se produce un cambio significativo de la señal en dos direcciones ortogonales. Una idea similar es explotada en el detector de Harris [100], aunque usando solo las primeras derivadas.

[113] proponen una medida basada en el cambio de la dirección del gradiente a lo largo de un contorno, multiplicado por la magnitud local del gradiente para hacerla más robusta a ruidos. (8). [114] mostró que este operador es equivalente al de [115] si la supresión de los no-máximos se aplica al gradiente antes de multiplicar por su magnitud.

$$C_K = \frac{I_{xx}I_y^2 + I_{yy}I_x^2 - 2I_{xy}I_xI_y}{I_x^2 + I_y^2} \quad (8)$$

A su vez, C_K puede ser multiplicado por el cambio de la dirección del borde [116].

[117] se basan en el facet y modelan la función de intensidad de la imagen $I(x,y)$ a través de una superficie descrita por un polinomio bicúbico (9) y establecen una medida para estimar la fortaleza del borde basado en el cálculo de la razón de cambio en el ángulo del gradiente a través de la expresión (10). [118] mostraron que la única diferencia entre las expresiones (8) y (10) es

el factor $\sqrt{(I_x^2 + I_y^2)}$ que corresponde a la magnitud del gradiente.

$$I(x, y) = k_1 + k_2x + k_3y + k_4x^2 + k_5xy + k_6y^2 + k_7x^3 + k_8x^2y + k_9xy^2 + k_{10}y^3 \quad (9)$$

$$k = -2 \frac{(k_2^2 k_6 - k_2 k_3 k_5 + k_3^2 k_4)}{(k_2^2 + k_3^2)^{3/2}} \quad (10)$$

Un estudio teórico del modelo de vértices y un análisis de los principales métodos existentes hasta ese momento fue realizado por [119]. En él perciben que la posición exacta de las esquinas puede ser detectada como los cruces por cero estables en el espacio de escalas, así como que el máximo local de las medidas de [92] y [114] se mueven en este espacio sobre la línea bisectora que pasa a través de la posición exacta de la esquina. A partir de estas dos importantes conclusiones diseñan un operador que las combina. Primero hallan la imagen Laplaciana, después calculan la medida de [92] en dos escalas y realizan la detección de sus extremos en todas las direcciones. Alrededor de cada extremo local detectado en la primera escala se busca el máximo local en la segunda y seleccionan como esquina el punto perteneciente a la línea que los une donde se produce un cruce por cero en la imagen Laplaciana.

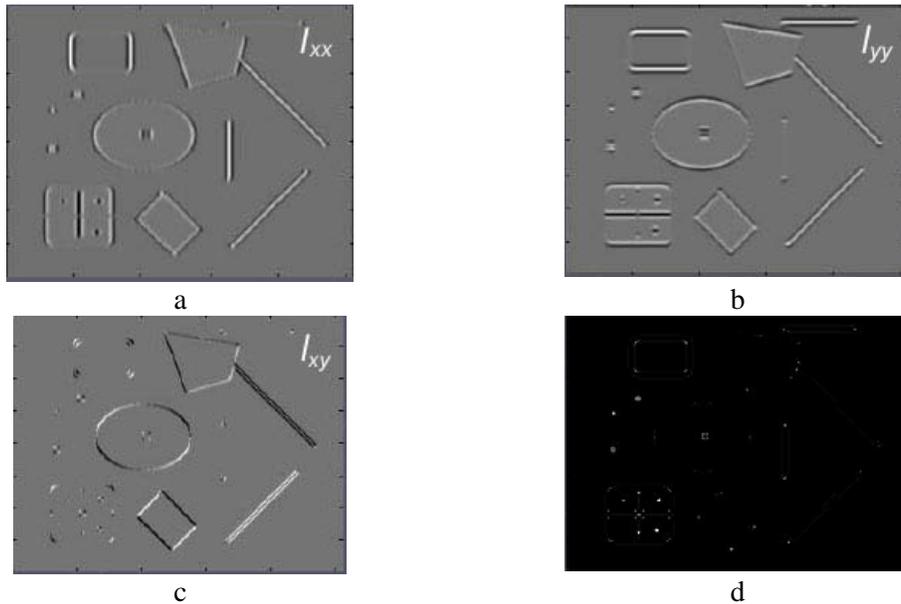


Fig. 11. Esquinas detectados utilizando la matriz hessiana. (a) – (c) Componentes de la matriz hessiana. (d) Esquinas seleccionadas

El Det_H se encuentra relacionado con el producto de las curvaturas principales $k_{\min} * k_{\max}$ conocido como Curvatura Gaussiana [120] (11).

$$(k_{\min} k_{\max}) = \frac{Det_H}{(1 + I_x^2 + I_y^2)^2} \quad (11)$$

En términos de geometría diferencial se puede decir que para un píxel $I(x,y)$ se cumple que:

- Si $k_{\min}k_{\max} > 0$ si y solo si el píxel es un punto elíptico.
- Si $k_{\min}k_{\max} < 0$ si y solo si el píxel es un punto hiperbólico.
- Si $k_{\min}k_{\max} = 0$ si y solo si el píxel es un punto parabólico.

De (11) se desprende que Det_H y Curvatura Gaussiana tienen el mismo signo. Esto significa que en la cercanía de una esquina Det_H tiene una parte elíptica y una hiperbólica en ambas caras del borde. En esta localización Det_H alcanza un máximo elíptico y este siempre se encuentra en el interior de la esquina independientemente del contraste de la imagen [119].

Nagel en sus trabajos [114-115, 121] propone un operador basado en el cálculo de la Curvatura Gaussiana, para seleccionar los puntos de máximo elíptico, e hiperbólico, así como el punto donde la curvatura se hace 0 entre ellos, correspondiendo al punto de pendiente máxima de la curva entre estos puntos en la superficie de la imagen.

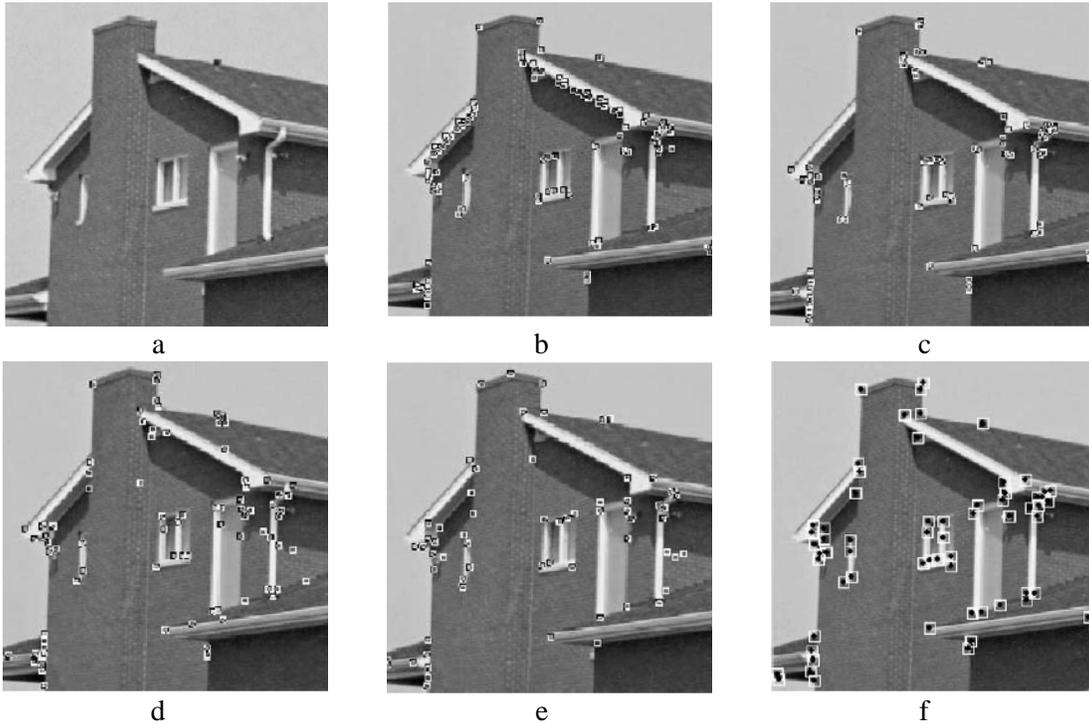


Fig. 12. Resultados de diversos algoritmos de detección de esquinas. (a) Imagen original. (b) Operador de Harris. (c) Operador de K&R. (d) Operador SUSAN. (e) Operador del CSS. (f) Operador COP

Por su parte [122] proponen un algoritmo de detección de bordes basado en la curvatura de la superficie. La curvatura total k de la imagen es definida por ecuación (12):

$$k \approx \frac{\partial^2 I}{\partial t^2} / \nabla I \quad (12)$$

donde $\partial^2 I \gg 1$ y t es la tangencial al borde y ∇I es el gradiente de la imagen.

Esta ecuación significa que la curvatura total es proporcional a la segunda derivada sobre la tangencial al borde e inversamente proporcional a la fortaleza del borde. Teniendo en cuenta esto, definen la medida de fortaleza de las esquinas como:

$$\Delta = \left(\frac{\partial^2 I}{\partial t^2} \right) - S |\nabla I|^2 \quad (13)$$

y suponiendo que las esquinas son a su vez puntos de borde, la medida queda simplificada como se muestra en la ecuación (14), permitiendo así que este algoritmo pueda ser utilizado en tiempo real :

$$\Delta = (\nabla^2 I)^2 - S |\nabla I|^2 \quad (14)$$

Hacia la invarianza en escalas y transformaciones afines.

a) Invariantes ante los cambios de escalas: Un problema presente en este tipo de operadores es que estos puntos de interés se encuentran optimizados solo para alcanzar una buena precisión translacional, pero la mayoría no ofrece buenos resultados ante los cambios de escalas, perspectiva e iluminación. En particular, muchas aplicaciones basadas en características locales utilizan un vector de rasgos que describe la localización donde se encuentra el punto de interés en vistas a obtener una buena redetección en otras imágenes. Esto provoca que los detectores que no se comporten invariantes ante estas distintas transformaciones producen descriptores poco robustos.

Entre los primeros intentos de abordar el problema de la detección robusta ante los cambios de escala tenemos a [123-124] que utilizan una representación piramidal obtenida mediante el filtro de diferencias gaussianas (DoG). Un punto es seleccionado como relevante si resulta ser un máximo en el espacio 3D y su valor absoluto es mayor que un umbral. Este operador es utilizado por [4] en el contexto del reconocimiento de objetos. Para la construcción del espacio de escala la imagen de entrada es sucesivamente desenfocada con filtros gaussianos y muestreada para obtener las imágenes de los niveles superiores (Fig. 13). La representación DoG es obtenida mediante la sustracción de dos imágenes sucesivas, utilizando los extremos locales en esta pirámide, determinando la localización y la escala de la característica. Los extremos locales del Laplaciano del Gaussiano (LoG) han sido utilizados para la selección de la escala característica de las estructuras presentes en la imagen [35].

Para la determinación de la escala característica de una estructura se hallan las respuestas del operador de selección de escala en un punto para un cierto rango de escalas y esta corresponde al extremo local de las respuestas obtenidas (Fig. 14). Pueden existir numerosos máximos o mínimos en una misma posición correspondientes a diferentes estructuras locales centrados en dicho punto. El núcleo del LoG es simétrico, por lo que esta operación puede ser interpretada como la búsqueda de características que tienen la misma forma del LoG. Cuando el núcleo del LoG corresponde con el tamaño de una estructura tipo blob (la varianza del filtro Gaussiano determina el tamaño de las características) la respuesta alcanza un extremo. Los máximos detectados del operador LoG sobre diferentes escalas son particularmente estables.

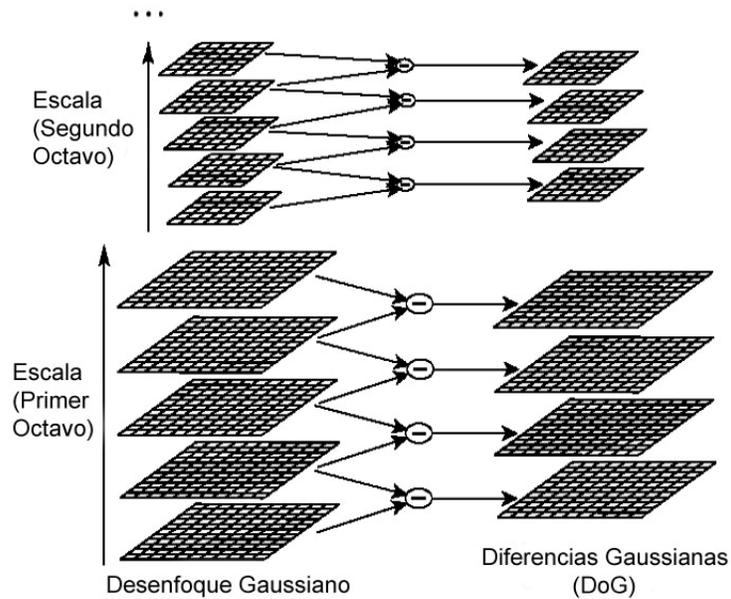


Fig. 13. Operador de las diferencias Gaussianas

En realidad la función DoG resulta ser una aproximación del LoG, aunque mucho más eficiente. Las limitaciones comunes de estos operadores están dadas porque los máximos locales pueden ser detectados en las cercanías de los contornos o líneas rectas donde el cambio se produce en una sola dirección. Las estructuras detectadas de esta manera resultan menos estables porque su localización es más sensible al ruido. Una variante más sofisticada consiste en seleccionar la escala para la cual tanto la traza como el determinante de la matriz Hessiana asumen un extremo local de manera simultánea [112]. La traza de esta matriz es igual al LoG pero teniendo en cuenta además el máximo del determinante se penalizan los puntos para los cuales la segunda derivada detecta cambios de señal en una sola dirección. [125] presentan una versión optimizada del operador Hessian-Laplace. El cálculo de la matriz hessiana en el espacio de escala, es aproximado por la convolución de la imagen con filtros tipo caja con un tamaño de ventana variable y la evaluación de esta matriz es realizada utilizando imágenes integrales. El estimado de la posición y la escala es realizado con el determinante del Hessiano en vez de la utilización de la combinación del determinante con LoG como el Hessian-Laplace. Los autores reportan que este operador es hasta 5 veces más rápido que el de diferencias de gaussianas, mientras que su desempeño es comparable con los que utilizan las derivadas gaussianas.

[126] proponen una generalización del operador de Forstner, que es invariante ante los cambios de escala, a la vez que es resistente ante los cambios de iluminación, mediante la sustitución de la comparación mediante mínimos cuadrados por un método basado en modelos.

La matriz de segundo momento ha sido extendida para lidiar con la detección de características robustas ante los cambios de escalas mediante la parametrización de los filtros gaussianos [127]:

$$M = \begin{bmatrix} I_x^2(x, \sigma_D) & I_x(x, \sigma_D)I_y(x, \sigma_D) \\ I_x(x, \sigma_D)I_y(x, \sigma_D) & I_y^2(x, \sigma_D) \end{bmatrix} * g(x, \sigma_I) \quad (15)$$

donde $g(x, \sigma_I)$ es un núcleo gaussiano con escala σ_I llamada escala de integración.

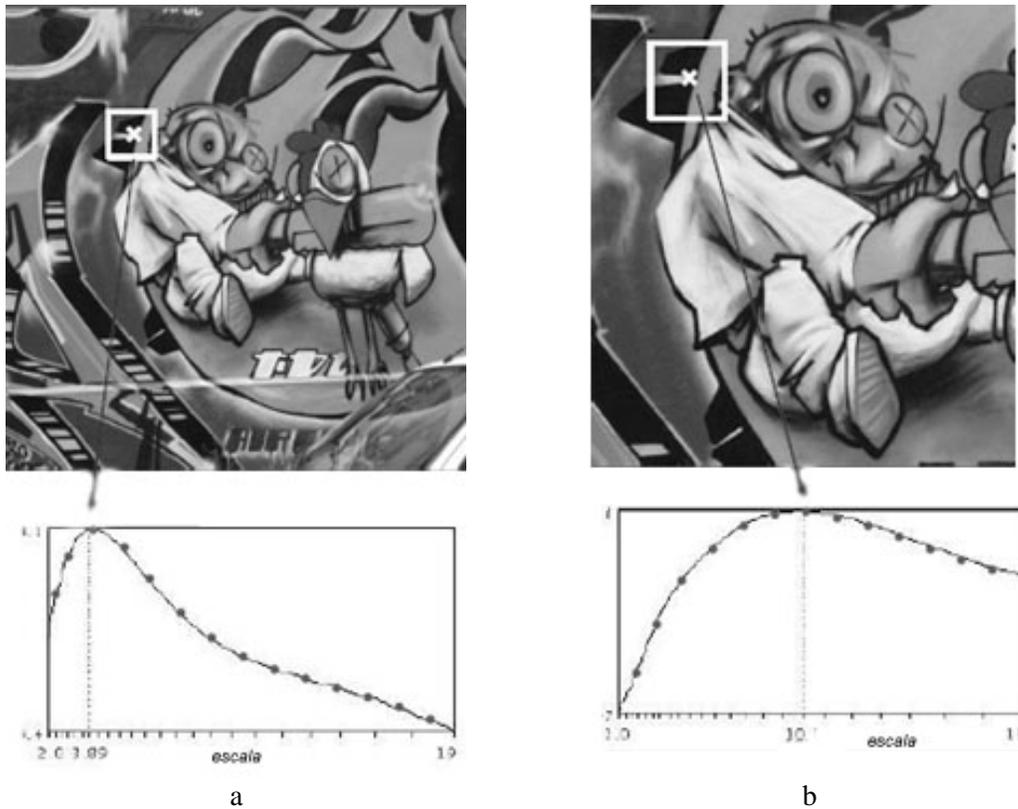


Fig. 14. Selección automática de escala. Se selecciona una región de la imagen donde alcanza un máximo local una función de la varianza del filtro gaussiano

[128] utilizan el operador de Harris para extraer puntos de interés sobre la representación del espacio de escala de una imagen obtenida mediante la convolución de filtros gaussianos con σ_D variable. Este método parte del supuesto de que la diferencia de escala entre las imágenes a tener en cuenta es conocida y obtiene una manera de representación donde se caracteriza la imagen mediante los puntos detectados en las diferentes escalas. El principal problema de esta representación multiscala es que de manera general una estructura se encuentra presente en un cierto rango de escala y los puntos son detectados en cada escala dentro de este rango, existiendo varios puntos que representan la misma estructura, con localización ligeramente diferente. Esto provoca un aumento de la complejidad de los algoritmos de comparación y la probabilidad de error. [129] combinan el detector de Harris multiscala con el LoG para la selección de la escala característica de manera automática.

b) Detectores invariantes ante las transformaciones afines: Este tipo de detectores constituyen una generalización de los métodos anteriores. En el caso de la transformación afin, el cambio de escala es diferente en cada dirección, y esto tiene influencia sobre la localización, la escala y la forma de la estructura local; por esta razón los detectores que tienen solo en cuenta la invarianza

ante las escalas no presentan buenos resultados ante transformaciones afines significativas. Teniendo en cuenta esto, la mayoría de este tipo de operadores utilizan un conjunto de puntos iniciales invariantes a escala y realizan la estimación de la forma elíptica que delimita la región.

[129-130] utilizan los puntos extraídos con el método de Harris sobre un rango de escalas y los ordenan de acuerdo a una medida normalizada de la fortaleza de la característica. En el caso de [130] las estructuras son comparadas entre si a través de las escalas y las primeras n son seleccionadas. Para llegar a la invarianza con respecto al punto de vista se modifica la forma de la ventana circular basado en la información local de la imagen. Su algoritmo desarrolla la idea de los descriptores de textura basados en formas propuesto por [131], el cual en su trabajo extiende la noción de espacio de escala al espacio de escalas afín en el que la ventana gaussiana rotacionalmente simétrica es sustituida por elípticas con matrices de covarianza asociadas a ella (matrices de forma)(16).

$$\mu_L(:, \Sigma_t, \Sigma_s) = g(:, \Sigma_s) \otimes ((\nabla L)(:, \Sigma_t)(\nabla L)(:, \Sigma_t))^T \quad (16)$$

donde $L(:, \Sigma)$ es la representación del espacio de escalas gaussiano afín de la imagen $I(\cdot)$, Σ_t es la matriz de covarianza asociada a la escala local y Σ_s es la asociada a la escala de integración. Utilizando estas matrices de covarianza y los núcleos gaussianos no uniformes asociados se generan $L(:, \Sigma)$ de acuerdo a (17).

$$g(x, \Sigma) = \frac{1}{2\pi\sqrt{\det \Sigma}} \exp(-x^T \Sigma^{-1} x / 2) \quad (17)$$

El esquema de adaptación de forma propuesto por Linderberg es utilizado para estimar la región normalizada de la imagen de la siguiente forma:

Sea la imagen 2D $I_R(x)$ y la imagen linealmente transformada $I_L(x) = I_R(Bx)$. Supóngase que la matriz de segundo orden de adaptación de forma ha sido calculada para ambas y son q_L y $q_R = Bq_L$ respectivamente. Para ambas imágenes se puede transformar la información presente en ellas a unos marcos normalizados utilizando la raíz cuadrada de la matriz de segundo momento. La imagen transformada es definida por (18) haciéndose de manera similar para I_R .

$$I_L(M_L^{-\frac{1}{2}}x) = I_L(x) \quad (18)$$

Linderberg deriva la siguiente propiedad que cumplen las matrices de segundo momento en el espacio de escalas afín (19):

$$\mu_L(q; \Sigma_t, \Sigma_s) = B^T \mu_R(Bq; B\Sigma_t B^T, B\Sigma_s B^T) B \quad (19)$$

y utilizando esta transformación se obtiene que en el marco normalizado se cumplen las relaciones (Fig. 16):



a



b



c

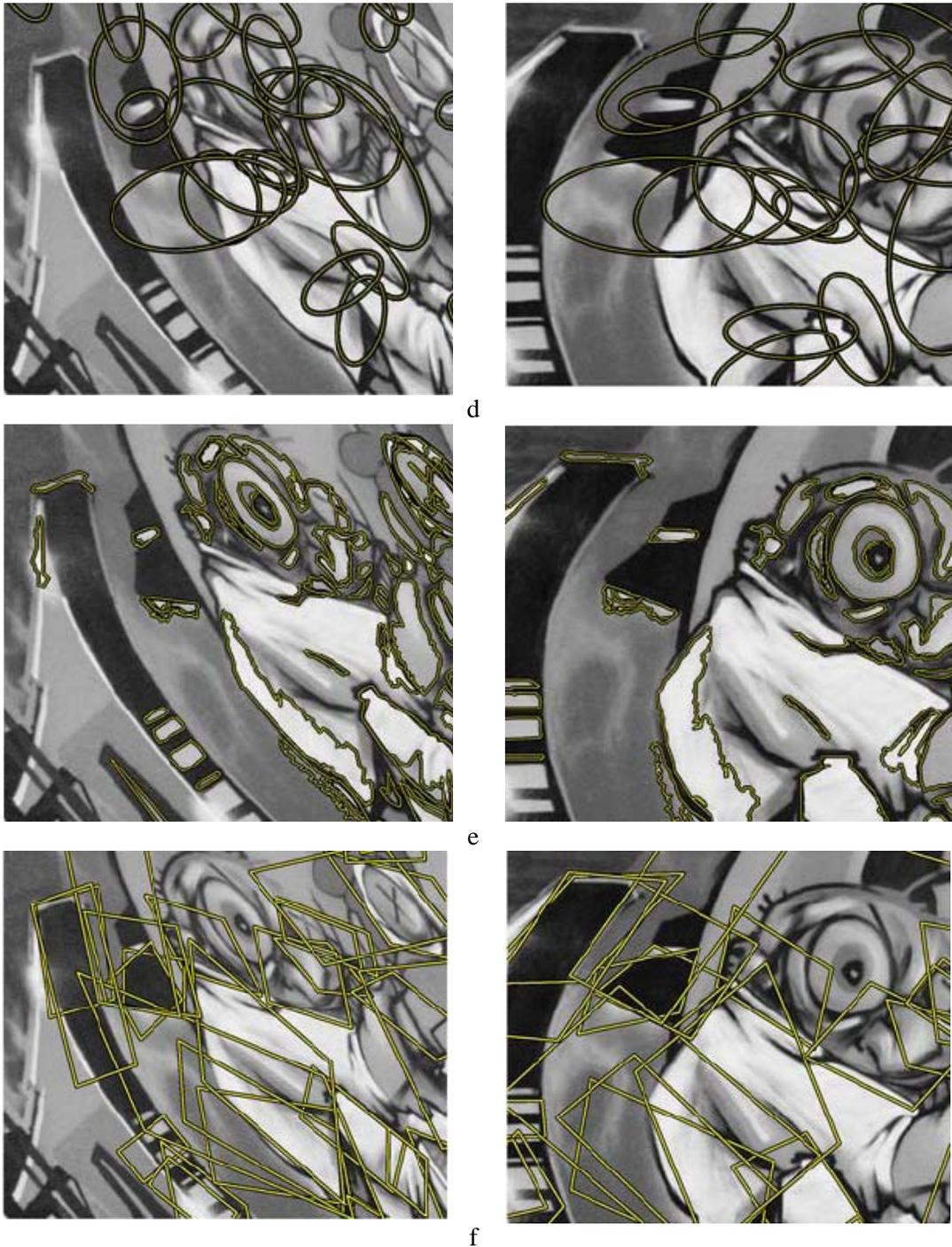


Fig. 15. Resultado de aplicar varios detectores invariantes ante transformaciones afines, tomado de la evaluación de los detectores de regiones afines realizada por (Mikolajczyk , Tuytelaars et al. 2005). (a) Harris-Afin. (b) Hessiano-Afin. (c) IBR. (d) Regiones Salientes.(e) MSER (f) EBR

$$\mu_L(q'; tI, sI) = I \tag{20}$$

$$I = \mu_L = B^T u_R B' = B'^T B' \tag{21}$$

donde I es la matriz identidad de 2x2 y B' es una rotación.

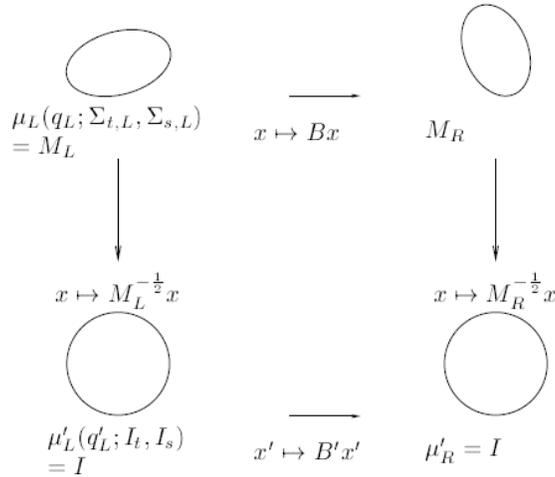


Fig. 16. Diagrama que ilustra el proceso de normalización de las regiones afines en las imágenes

En [129] se utiliza además el operador Hessian-Laplace para detectar las estructuras iniciales complementarias a las detectadas con su contraparte de Harris, y luego se le aplica este mismo procedimiento para la estimación de la forma elíptica de la ventana definiendo así el operador Hessian-Afín.

En lugar de estimar la forma de la región a través de la matriz afín de segundo momento, pueden ser utilizadas otras heurísticas. [99, 132] definen un método que se basa en la información de los bordes cercanos a las esquinas para el cálculo de la región. Parten de la idea de que este tipo de estructuras son generalmente más estables que las esquinas ya que pueden ser detectadas sobre un rango más amplio de condiciones de observación. Comienzan a partir de los puntos detectados con el operador de Harris y la información de los bordes cercanos extraídos con el operador de Canny [80] en un rango de escalas. Centrado en una esquina p se definen dos puntos p_1 y p_2 que se mueven sobre los bordes en dirección ortogonal alejándose de p . Su velocidad relativa es calculada utilizando los parámetros afines l_1 y l_2 :

$$l_i = \int abs(| p_i^{(1)}(s_i) p - p_i(s_i) |) ds_i \tag{22}$$

donde s_i es un parámetro de curva arbitrario $p_i^{(1)}(s_i)$ es la primera derivada con respecto a s_i .

Esta condición establece que las áreas entre $\langle p, p_1 \rangle$ y $\langle p, p_2 \rangle$ es la misma. A partir de estos puntos se forma un paralelogramo que va a definir la región a ser considerada.

2.2.4 Métodos basados en el Sistema Visual Humano

Este conjunto de métodos se encuentran basados en modelos fisiológicos que intentan simular el sistema de percepción humana. Este tipo de sistemas resultan los más complejos de entre los analizados en este trabajo, puesto que la selección de las regiones salientes es realizada mediante la combinación de diversos tipos de características (intensidad, orientación y color se encuentran entre las más utilizadas). En la literatura existen propuestas de modelos teóricos e implementaciones de estos. El objetivo de este tipo de métodos es detectar aquellas regiones altamente distintivas de las imágenes que son capaces de captar el foco de la atención visual humana (por ejemplo una mancha oscura en una pared blanca).

La mayoría de estos sistemas se encuentran basados en la Teoría para la Integración de Características (FIT), presentada por [133]. Este modelo establece que diferentes características son detectadas paralelamente en el campo visual, mientras que los objetos son detectados de forma separada en una etapa posterior que requiere de la atención visual. La información de los mapas de características resultantes (mapas que resaltan la relevancia de una región de acuerdo a la característica analizada). Uno de los principales planteamientos de esta teoría es que las características deben ser detectadas fácilmente, de manera eficiente y en forma paralela si difieren en al menos una característica del entorno homogéneo que lo rodea. El Modelo de Búsqueda Dirigida, propuesto por [134] se encuentra entre otro de los trabajos más significativos dentro de esta área. Originalmente surgió para dar respuesta a deficiencias encontradas en las primeras versiones del modelo FIT, sin embargo sus principales logros fueron explicar y predecir el resultado de experimentos de búsquedas visuales. Este modelo ha continuado su desarrollo con el paso de los años: Búsqueda Guiada 1.0 [135], Búsqueda Guiada 2.0 [134], Búsqueda Guiada 3.0 [136] y Búsqueda Guiada 4.0 [137].

[138] presentan una descripción detallada de la implementación de un modelo basado en FIT. En su trabajo utilizan una red neuronal para determinar las regiones más sobresalientes en mapas topográficos, lo cual modela como debe ser el funcionamiento del cerebro. Sin embargo desde el punto de vista de sistemas prácticos, la utilización de una red neuronal representa una sobrecarga puesto que pueden ser utilizadas vías alternativas más eficientes. [139-140] se basan en este modelo para realizar su implementación. Proponen la utilización del filtrado en las imágenes para el cálculo de los mapas de relevancia, permitiendo su utilización en aplicaciones reales. Como rasgos, consideran dos colores opuestos (rojo-verde y azul-amarillo), 16 orientaciones diferentes y curvatura local. Para el cálculo de la relevancia de una característica en particular, proponen un operador de conspicuidad el cual compara el valor de una característica con su entorno. Esta es una técnica común para la detección de contraste en imágenes y es conocido como mecanismo/diferencia centro-alrededor. El resultado de esto es almacenado en mapas de conspicuidad que reflejan la relevancia de acuerdo a una característica. Estos mapas son integrados en un único mapa de relevancia, el cual identifica un pequeño número de regiones de interés convexas. La principal desventaja de este sistema es la alta complejidad computacional, resultante de aplicar varias veces el filtrado sobre diferentes escalas.

Laurent Itti [141-143] presentaron un conjunto de Herramientas de Visión Neuromórfica (NVT), mediante la combinación de los mapas de saliencia y las redes neuronales propuestas por [138], con la utilización de filtros lineales, el cálculo de características utilizando las diferencias centro-alrededor y los mapas de conspicuidad de [140]. Los mapas de orientación, intensidad y color son calculados utilizando pirámides de imágenes. Proponen además una función de pesos para la combinación de los diferentes mapas de rasgos de forma tal que se beneficien aquellos con pocos picos. Esta técnica resulta mucho más eficiente que el proceso de relajación propuesto

por [140] y mantiene buenos resultados. Este sistema presenta falta de robustez ante transformaciones 2D como translación, rotación y reflexión [144]. Este mismo autor propone una versión mejorada (sistema SAFE) que se comporta más estable con respecto a transformaciones geométricas.

El sistema VOCUS (Detección Visual de Objetos utilizando un Sistema de Atención Computacional)[145], se encuentra basado en los trabajos de Itti y comparte a grueso modo la arquitectura propuesta por [138]. Al igual que este extrae a partir de la imagen tres tipos de rasgos principales: intensidad, orientación y color calculados sobre diferentes escalas. Para cada característica se calcula la pirámide de imágenes correspondiente, a partir de las cuales se obtienen los mapas de escala. Estos representan la relevancia de las distintas características sobre las escalas. Estos mapas son combinados sucesivamente en los mapas de rasgos, los mapas de conspicuidad y el mapa de relevancia en el cual la intensidad de gris es proporcional al grado de importancia. A partir de este mapa se extrae la región más saliente (MSR) y el foco de atención (FOA) se traslada a ella. Este sistema permite la detección de varios focos de atención, permitiendo seleccionar hasta 20 de ellos, a diferencia del sistema de [146] que realiza la detección en dos etapas: la primera de las cuales selecciona múltiples focos de atención (usualmente 4) y después realiza una selección sobre estos y mantiene un simple foco de atención. En la figura Fig. 17 se muestra un diagrama que ilustra el funcionamiento de este sistema. Este sistema ha mostrado buenos resultados en el contexto del SLAM visual [147], seguimiento de personas [148].

El análisis multiescala es realizado también por [149] los cuales se basan en una selección jerárquica de objetos para extraer la región de interés. Este proceso lo realizan de manera selectiva: primeramente son analizados las regiones sobre escalas más gruesas y si resultan interesantes se investigan en escalas más finas.

Este conjunto de métodos selecciona regiones de interés altamente distintivas en las imágenes y que presentan una alta repetitividad [150], pero entre sus principales desventajas se encuentran una gran complejidad computacional al tener que extraer distintos tipos de características y combinarlos entre sí. [151] presenta una implementación paralela de un sistema de atención visual capaz de funcionar en tiempo real.

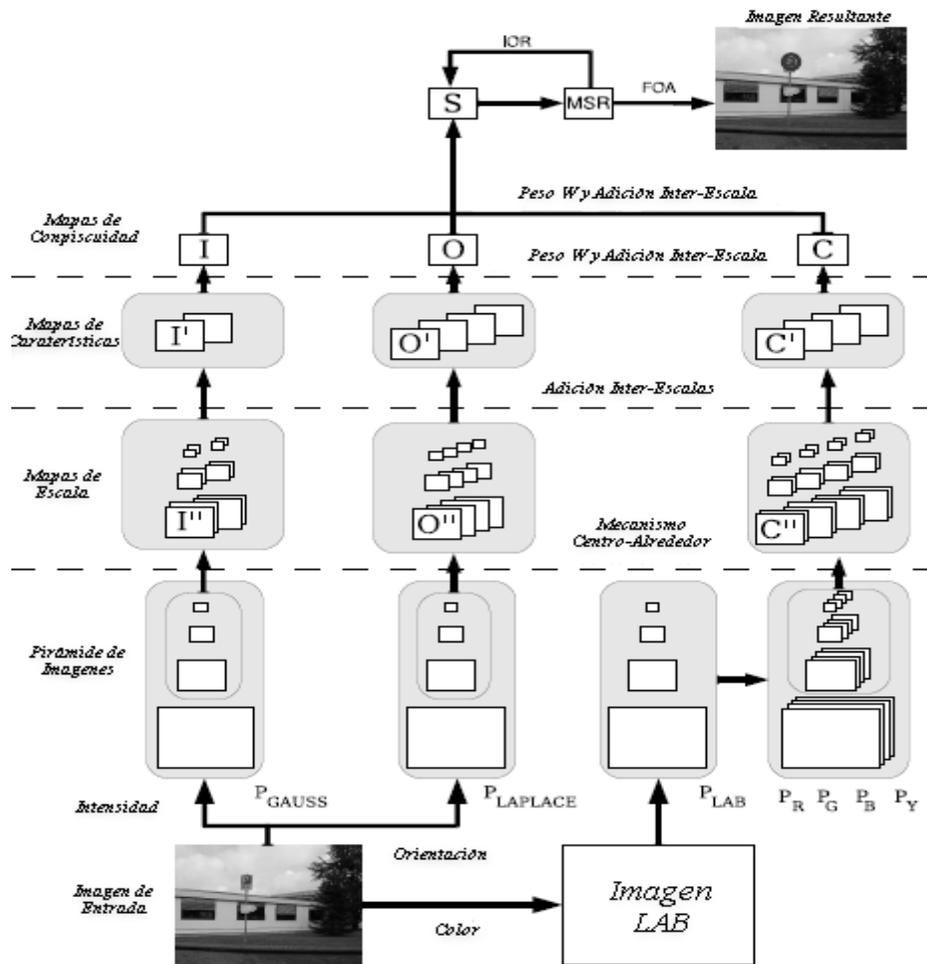


Fig. 17. Diagrama general que ilustra el funcionamiento del sistema VOCUS

2.3 Descripción de características

Con el fin de poder comparar los puntos y/o regiones extraídas, se almacena un vector que describe el patrón de la distribución de los tonos de gris presente en la misma. El descriptor más simple es el formado por un vector de intensidades y técnicas de correlación cruzada pueden ser utilizadas para obtener un valor de similitud. Sin embargo, este descriptor no es robusto ante cambios significativos de las condiciones de observación.

Existen en la literatura un gran número de descriptores de características locales, los cuales, pueden ser agrupados principalmente en tres grupos [10]:

Descriptores basados en Distribución: Utilizan histogramas para representar diferentes características o formas. Entre estas aproximaciones se encuentran los histogramas de las distribuciones de los tonos de gris; histogramas de las posiciones relativas en el vecindario de un

punto de interés [152], histogramas que representan las relaciones binarias (orden y recíproco) entre las intensidades de los píxeles del vecindario [153].

Dentro de este grupo de métodos se encuentra el descriptor SIFT (Característica Invariante ante las Transformaciones de Escala), el cual es posiblemente el descriptor más utilizado hoy en día. Consiste en un histograma 3D de la posición y orientación del gradiente de los puntos en el interior de la región. La contribución de cada elemento es regulado utilizando la magnitud del gradiente. Numerosas variantes de este descriptor han sido propuestas en la literatura, entre los que encontramos: PCA-SIFT [154], los cuales realizan un análisis de componentes principales sobre los rasgos extraídos con SIFT para lograr una mayor distintividad, Histograma de la Localización y Orientación del Gradiente (GLOH) [10] extraen las características SIFT sobre un sistema de coordenadas polares y realizan PCA sobre este para la reducción de la dimensionalidad, Contexto de Forma [155] siguen una idea similar al SIFT, pero teniendo en cuenta solo los píxeles de borde obtenidos a partir del operador de Canny.

Descriptores basados en técnicas de Frecuencia Espacial: Este conjunto de métodos se basan en el análisis de las componentes de frecuencia de la imagen. En este contexto han sido utilizados los filtros de Gabor [156], aunque se requiere un número elevado de estos filtros para capturar pequeños cambios en frecuencia y orientación. Las respuesta ante los wavelet de Haar, es utilizada por [125] como descriptor en el operador SURF y muestra en su trabajo como este obtiene resultados comparables con el SIFT.

Descriptores Diferenciales: Estos métodos utilizan el cálculo de las derivadas en la vecindad de los puntos como una forma de describir su vecindario. [157] deriva invariantes diferenciales a partir del análisis de las derivadas locales. [158] desarrollan un grupo de filtros direccionales, a partir de las derivadas sobre direcciones particulares, obteniendo descriptores invariantes a las rotaciones. También han sido utilizados filtros complejos derivados de la familia $K(x, y, \theta) = f(x, y)^{i\theta}$ donde θ es la orientación. Para la función $f(x, y)$ pueden ser tomadas derivadas gaussianas [130] o polinomios [159].

Estos algoritmos resultan buenos complementos de los detectores descritos anteriormente. Varias evaluaciones han sido realizadas en la literatura [10-11, 125], y la mayoría coincide con que el descriptor que mejor se comporta es el SIFT, aunque el Contexto de Forma y el SURF también presentan buenos resultados. Las variantes del SIFT que involucran al análisis de componentes principales, obtienen descriptores con un mayor poder de discriminación y más eficientes en la etapa de comparación, ya que usualmente los vectores de rasgos obtenidos de esta forma presentan un menor número de componentes. La principal desventaja de estos métodos resulta en el alto costo computacional que tiene su construcción, siendo más factible para un gran número de aplicaciones la utilización del SIFT.

2.4 Discusión

En el contexto de extracción de puntos de interés invariantes solo a la traslación y rotación, diversos autores han realizado estudios comparativos donde han tenido en cuenta operadores representativos de cada familia, en los cuales han establecido criterios para cuantificar una o más propiedades deseables en los detectores. Por ejemplo [111], analizan el comportamiento de los operadores de Harris [100], K&R [113] y Beaudet [92], con respecto a las propiedades exactitud, estabilidad y eficiencia (de la manera en que fueron definidas en la primera sección), en presencia de ruido. Sus resultados muestran que el operador de Harris sobrepasa a los otros, esto se debe principalmente a que los detectores basados en el Hessiano detectan características tipo

blob que no son bien localizadas. Una versión de este mismo operador presenta también los mejores resultados comparados con los métodos de Forstner [104], Horaud [160] y Heitger[161], de acuerdo a los criterios de repetitividad (el porcentaje total de los puntos observados que se repiten en dos imágenes de la misma escena) y distintividad, sujeto a cambios de escala, rotación e iluminación [102]. Para el cómputo de la repetitividad los autores utilizan una matriz de homografía que relaciona dos imágenes I e I_t y establecen que para los puntos detectados en la primera debe ser detectado el correspondiente a aplicar en el punto la transformación. Sin embargo[162], plantean que su detector basado en la transformada wavelet, sobrepasa al operador de Harris tanto en distintividad como en repetitividad, en experimentos realizados en el contexto de la recuperación de imágenes.

La evaluación de la repetitividad de un algoritmo presenta como principal desventaja que si un algoritmo marca todos los píxeles de la imagen como esquina, entonces tendrá un valor de 100%, esto quiere decir que este valor beneficia a los algoritmos que detectan un mayor número de esquinas (independientemente de si son falsas o no)[163-164]. Estos autores introducen dos nuevos criterios: consistencia (que constituye una variante del criterio de repetitividad) y exactitud. Para medir este último utilizaron el criterio de observadores expertos que seleccionaron manualmente las localizaciones de las esquinas en los conjuntos de imágenes de pruebas. Teniendo esto en cuenta realizaron la comparación de los operadores K&R, Harris, SUSAN [67] y CSS [37], sujetos a rotación, cambios de escala y transformaciones afines, presentando mejores resultados este último. Esto se debe principalmente a que no fueron tenidos en cuenta las extensiones invariantes a escalas de estos métodos.

[165] analizan el comportamiento de una serie de operadores de detección de puntos en el contexto de detección y seguimiento de marcas en una secuencia de video. Este análisis es realizado sobre la base de una localización precisa y una buena estabilidad temporal (las esquinas deben aparecer en todos los cuadros de la secuencia) en presencia de ruido. Otra vez el detector de Harris, junto con el KLT [8], fueron los que obtuvieron un mejor desempeño, dejando rezagados al K&R y SUSAN, presentando este último resultados imprevisibles. Esto se debe principalmente al que el operador de K&R involucra segundas derivadas, lo cual produce un efecto de amplificación del ruido. Por su parte, SUSAN pertenece a una familia de operadores que no utilizan las derivadas, pero a su vez la ausencia del desenfoque gaussiano lo hace sensible al ruido.

De manera general, teniendo en cuenta los estudios realizados, el detector de Harris es el que mejores resultados presenta en cuanto a repetitividad y es el más utilizado entre los detectores invariantes a rotación.

En cuanto a los detectores con mayor grado de invarianza, [96] realizaron el estado del arte de este tipo de métodos, en el cual fueron evaluados los principales detectores presentes en la literatura teniendo en cuenta transformaciones planares (cambios de escala, afines, y desenfoque) en imágenes. En este grupo el detector de Harris-Laplace muestra gran repetitividad y precisión en la localización, heredado del detector de Harris, sin embargo en la estimación de la escala resulta menos preciso debido a la naturaleza multiescala de las esquinas. El Hessian-Laplace es más robusto que la versión de escala única [96], debido principalmente a que los blob se encuentran mejor localizados que las esquinas a través de las escalas. Las implementaciones eficientes de estos operadores (DoG y SURF), que fueron ideadas para ganar en costo computacional, presentan sin embargo resultados comparables a sus versiones originales, siendo este último hasta tres veces más rápido que el operador DoG y cinco veces más rápido que el Hessian-Laplace [125]. Todo lo contrario sucede con las regiones salientes

[75], de ellas se debe extraer un histograma y calcular su entropía por cada región candidata en el espacio de escalas y en el afín. Por su parte, el detector basado en bordes [98-99] presenta buenos resultados en cuanto a repetitividad y precisión, pero extrae muy pocas regiones por imagen. El detector del MSER, presenta los mejores resultados entre los detectores de regiones afines, tanto como en repetitividad como en eficiencia computacional. [11] realizó la evaluación de estos mismos métodos combinados con descriptores sobre una colección de objetos en 3D, tomados en diferentes posiciones. Los resultados del desempeño de este tipo de métodos varían al ser tomados en cuenta en este tipo de escena, por ejemplo: el método de MSER combinado con los descriptores de forma y SIFT que obtuvieron el mejor resultado en escenas planas, fueron superados por el Hessian-Afin y las diferencias de Gaussianas en escenas 3D.

Al analizar estos detectores invariantes hay que tener en cuenta que el poder discriminativo de estas características disminuye al aumentar el nivel de invarianza, ya que al tener que ser tomados más patrones como equivalentes hay que estimar un mayor número de parámetros, lo que produce que el proceso de extracción sea más complejo computacionalmente y sujeto a mayores fuentes de ruido. Por esta razón, es aconsejable seleccionar un detector con el nivel de invarianza necesario para la aplicación en cuestión. Teniendo en cuenta esto, los detectores DoG y SURF presentan resultados comparables a sus contrapartes afines, siendo a su vez mucho más eficientes. Otra variante puede ser combinar estos operadores con descriptores que le proporcionen un mayor poder de discriminación.

3 Principales instituciones que trabajan en el tema

Las características locales son utilizadas en la actualidad en una amplia gama de aplicaciones relacionadas con la visión por computadora, entre las que podemos mencionar: interpretación y recuperación de imágenes y video, reconocimiento de objetos, Localización y Mapeo Simultaneo de Robots (SLAM) y seguimiento de puntos en video. La investigación en esta área ha tenido un auge en estos últimos años, existiendo prestigiosos investigadores y numerosas instituciones interesados en este tema. A continuación mencionaremos unos cuantos que consideramos se encuentran entre los más relevantes en esta área.

– **Dr. Krystian Mikolajczyk:** Investigador principal del grupo de *Visión por Computadora* (Robot Vision), del departamento de Ingeniería Electrónica de la Universidad de Surrey (<http://info.ee.surrey.ac.uk>). En la actualidad se encuentra asociado a dos proyectos relacionados con las características locales:

- **EPSRC:** *Reconocimiento de Categorías de Objetos y Escenas* (<http://info.ee.surrey.ac.uk/Personal/K.Mikolajczyk/projects/rocs/>) e intenta desarrollar un sistema general de reconocimiento visual capaz de lidiar con un gran número de escenas y objetos pertenecientes a entornos genéricos tanto bajo techo como al aire libre. Este proyecto se centra en el desarrollo de nuevas representaciones de imágenes que posibiliten el modelado simultáneo de escenas y categorías de objetos (mediante una representación de categorías con modelos jerárquicos múltiples de apariencia y estructura), nuevos métodos para la extracción de características locales robustos al cambio del punto de observación (introduciendo segmentación de imagen bottom-up) y desorden del fondo, nuevas estructuras de datos, agrupamiento y técnicas de búsqueda para un reconocimiento eficiente (se plantearon investigar estructuras arbóreas para la búsqueda en espacio de grandes dimensiones).

- *CLARET II: Clasificación y Recuperación de Imágenes* (<http://claret.wikidot.com/start>). Implementa un método para reconocer y localizar simultáneamente múltiples clases de objetos en una imagen. Este reconocimiento es basado en clases de apariencia, construidas a partir de características locales.
- **Dr. Cordelia Schmid:** Directora de Investigación del Instituto Nacional de Investigaciones en Informática y Automática (INRIA). Esta investigadora está al frente de un grupo de investigadores que conforman el proyecto Aprendizaje y Reconocimiento en Visión (LEAR) (<http://lear.inrialpes.fr/>). Este proyecto se centra en combinar aprendizaje de máquina con los métodos de análisis de imagen al reconocimiento de objetos, interpretación de escena, detección de categorías, recuperación de imágenes. Ya ha obtenidos buenos resultados en las evaluaciones Trecvid 2008 y PASCAL VOC 2008.
- **Dr. Tine Tuytelaars:** Investigadora Principal *del Laboratorio Visión para Industria Comunicaciones y Servicios* (VISICS) de la Universidad Católica de Leuven, en Bélgica. Actualmente se encuentra como investigadora en el proyecto CLASS: Anotación a Nivel Cognitivo utilizando Estructuras Estadísticas Latentes (<http://class.inrialpes.fr/>), el cual busca desarrollar una habilidad cognitiva básica para ser utilizada en el análisis de contenido inteligente para el descubrimiento automático de categorías y atributos. Se persigue combinar modelos de estructuras latentes y aprendizaje semisupervisado con descriptores visuales, para el análisis a tres niveles de abstracción: objetos individuales, clases de objetos y atributos, categorías jerárquicas y relaciones entre entidades.
- **Dr. C. Kenney, junto con el Prof B.S Manjunath y M. Zuliani:** en el *Laboratorio de Investigaciones de Visión por Computadora* (Vision Research Lab), del departamento del Electrónica e Ingeniería en Computación de la Universidad de California. Trabajan en un proyecto de detección de esquinas (<http://www.vision.ece.ucsb.edu/~zuliani/Research/CornerDetection/CornerDetection.shtml>), en el cual buscan desarrollar una plataforma unificada para analizar diferentes detectores de esquinas basados en las propiedades espectrales de la matriz de autocorrelación. Plantean utilizar la plataforma desarrollada para extender los detectores más utilizados a píxeles con dimensión mayor que 2 (ejemplo imágenes tomográficas), e intensidad mayor que 1 (imágenes a color o multispectrales). Utilizando la teoría de la condición pretenden investigar la estructura intrínseca del vecindario de una esquina, para a partir de aquí definir una plataforma de detección de esquinas multiescala.
- **Dr. Andrew Zisserman, Dr Jiri Matas, Dr. Radu Horaud, Prof. Pascal Fua y Oliver Faugeras** pertenecen al proyecto VISIONTRAIN: Sistema Computacional de Visión Cognitiva (<http://visiontrain.inrialpes.fr/>). Este proyecto se plantea como objetivos principales desarrollar teorías computacionales y métodos para la visión de bajo nivel; análisis de movimiento a partir de secuencia de imágenes, aprendizaje y reconocimiento de formas, objetos y categorías; modelos cognitivos que describan acciones.
- **Dr. Pascal Fua y Dr. Vincent Lepetit** del laboratorio de *Visión por Computadora (CVLAB)*, perteneciente a la Escuela Politécnica Federal de Laussana. Se encuentran en la actualidad investigando en un conjunto de temas relacionados con la extracción de características locales:
 - *Detección de Objetos en Tiempo Real y Estimación de Pose* (<http://cvlab.epfl.ch/research/detect/ferns/>): Se encuentran investigando sobre métodos eficientes de reconocimiento de puntos para ser utilizados en el contexto de SLAM.

- *Ampliación de Superficies Deformadas* (<http://cvlab.epfl.ch/research/detect/deformable/>). A partir de un método de detección de puntos desarrollan técnicas en tiempo real para detectar superficies deformadas y encontrando correspondencias entre imágenes le permiten obtener la deformación.
- **Dr. Luc Van Gool** perteneciente al *Laboratorio de Visión por Computadora*, perteneciente al departamento de Tecnologías de la Información e Ingeniería Eléctrica de la Universidad de Zurich. Se encuentra actualmente en un proyecto de Reconocimiento de Objetos y Escena, a partir de regiones afines extraídas a partir de un conjunto de imágenes.
- **Dr. David Lowe** de la Universidad de British Columbia ha desarrollado en estos últimos años un conjunto de proyectos relacionados con la extracción de puntos de interés entre los que se pueden mencionar: Reconocimiento de objetos utilizando regiones invariantes, Realidad aumentada en escenas naturales, creación de Imágenes panorámicas de manera automática, Reconocimiento de locaciones y construcción de mapas para robots móviles (<http://www.cs.ubc.ca/~lowe/research.html>).

4 Conclusiones

Como se ha podido apreciar, existe una gran cantidad de trabajos concernientes a la detección de características locales. Seleccionar cual o cuales de estos operadores son los que presentan un mejor desempeño, resulta una tarea compleja debido principalmente a la gran variedad de patrones detectados por estos métodos (distintos tipos de esquinas: uniones tipo T, tipo L, etc; blob, regiones de intensidad uniforme, regiones con alta entropía). La mayoría de estos algoritmos no son capaces de lidiar con todo este conjunto de características, existiendo operadores que extraen conjuntos de puntos que son complementarios entre ellos. Además, las regiones seleccionadas presentan diversos grados de invarianza ante distintas transformaciones. La amplia mayoría de los detectores presentados en este estudio son invariantes desde el punto de vista traslacional y rotacional [67, 85, 92, 100, 113]. Varios de ellos han sido extendidos para lidiar con el cambio de escala, principalmente mediante la adaptación de los operadores para realizar la búsqueda en el espacio de escala [125, 129, 166]. Mientras que los detectores con mayor grado de invarianza son aquellos que se presentan invariantes ante las transformaciones afines [75, 94, 99, 129].

Producto de la creciente utilización de los rasgos locales en diversos tipos de aplicaciones de visión por computadora, el desarrollo de métodos más robustos, eficientes e invariantes se ha convertido en el objetivo de un conjunto de investigadores. Aunque para seleccionar un extractor se deben tener en cuenta las características particulares del problema, muchos autores se han decantado por aquellos que presentan un buen compromiso entre eficiencia computacional y grado de invarianza.

A pesar de que este tema ha sido investigado por más de 50 años, todavía se considera que queda por avanzar en temas como dotar a los detectores de una interpretación semántica, o realizar métodos automáticos de combinación y selección de detectores. A continuación listamos un conjunto de problemas que permanecen sin resolver y que están siendo atacados por las principales instituciones que trabajan en esta temática.

4.1 Problemas abiertos

Limitada Repetitividad: A pesar del gran número de detectores presentes en la literatura y de los resultados obtenidos en diversos tipos de aplicaciones, la repetitividad de los principales operadores apenas sobrepasa el 50%. Los métodos de extracción se comportan inestables para pequeñas regiones y aunque para regiones mayores producen un buen estimado de la escala y la forma, otros factores como la oclusión y el desorden del fondo afectan los resultados.

Aumento de la Eficiencia: A pesar de que existen aproximaciones eficientes como el DoG y el SURF de los operadores invariantes a escala y que estas se comportan bien ante pequeñas transformaciones afines, se hace necesario la definición de detectores eficientes capaces de lidiar con este tipo de transformaciones puesto que de los operadores existentes solo el MSER presenta un buen balance entre eficiencia y resultados, y aún así no es factible su utilización en aplicaciones en línea.

Falta de interpretación semántica: Las falta de una definición formal de lo que constituyen las características locales (uniones, regiones de intensidad uniforme, regiones con alta incertidumbre, puntos relevantes en el dominio de frecuencia), provoca la falta de una interpretación semántica de las regiones extraídas. Estas solo pueden ser interpretadas como regiones de la imagen, las cuales pueden corresponder a partes significativas de la misma. Puede ser utilizado algún tipo de conocimiento a priori que permita el diseño de detectores capaces de extraer información significativa de las partes de los objetos.

Selección automática del conjunto de detectores óptimos: Como ya se ha discutido, todos los operadores no extraen el mismo tipo de características, así como que no se comportan igual en el dominio de las aplicaciones. Muchos autores utilizan diversos detectores de forma paralela, provocando un aumento de la complejidad computacional. Se hace necesaria la definición de herramientas que permitan a una aplicación específica evaluar y sugerir el detector o la combinación de detectores óptimos para esta, teniendo en cuenta que muchos de ellos extraen regiones que son complementarias entre sí.

Referencias bibliográficas

1. Attneave, F., Some informational aspects of visual perception. Psychol. Rev, 1954. vol. 61(no. 3): p. 183-193.
2. Chum, O. and J. Matas, Web Scale Image Clustering. Large Scale Discovery of Spatially Related Images. 2008, Czech Technical University Prague.
3. Tian, Q., et al., Image Retrieval Using Wavelet Based Salient Points. Journal of Electronic Imaging, Special Issue on Storage and Retrieval of Digital Media, 2001. vol. 10(no. 4).
4. Lowe, D.G. Object recognition from scale invariant features in 7th International Conference on Computer Vision. 1999. Kerkyra, Greece.
5. Rothganger, F., et al., 3D Object Modeling and Recognition Using Local Affine-Invariant Image Descriptors and Multi-View Spatial Constraints. International Journal of Computer Vision, 2006. vol. 66(no. 3): p. 231 - 259.
6. Rosten, E., High performance rigid body tracking. 2006, University of Cambridge.
7. Rosten, E. and T. Drummond, Fusing points and lines for high performance tracking, in 10th IEEE International Conference on Computer Vision, Springer-Verlag., Editor. 2005: Beijing, China. p. 1508 - 1515.
8. Tomasi, C. and T. Kanade, Detection and tracking of point features. 1991, Carnegie Mellon University.

9. Tuytelaars, T. and K. Mikolajczyk, Local Invariant Feature Detectors: A Survey. *Foundation and Trends in Computer Graphics and Vision*, 2007. 3(3): p. 177-280.
10. Mikolajczyk, K. and C. Schmid, A Performance Evaluation of Local Descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005. vol. 27(no. 10): p. 1615 - 1630.
11. Moreels, P. and P. Perona, Evaluation of Features Detectors and Descriptors based on 3D Objects. *International Journal of Computer Vision*, 2006. vol. 73(no. 3): p. 263 - 284.
12. Rosenfeld, A. and E. Johnston, Angle detection on digital curves. *IEEE Transactions on Computers*, 1973. vol. C-22: p. 875-878.
13. Rosenfeld, A. and M. Thurston, Edge and curve detection for digital scene analysis. *IEEE Transactions on Computers*, 1971. C-20: p. 562-569.
14. Rosenfeld, A., M. Thurston, and Y.H. Lee, Edge and curve detection: Further experiments. *IEEE Transactions on Computers*, 1972. C-21: p. 677-715.
15. Rosenfeld, A. and J.S. Weszka, An improved method of angle detection on digital curves. *IEEE Transactions on Computers*, 1975. vol. C-24(no. 9): p. 940-941.
16. Freeman, H., Computer processing of line drawing images. *Surveys*, 1974. vol. 6(no. 1): p. 57-97.
17. Freeman, H. and L.S. Davis, A corner-finding algorithm for chain-coded curves. *IEEE Transactions on Computers*, 1977. vol. C-26(no. 3): p. 297-303.
18. Beus, H.L. and S.S.H. Tiu, An improved corner detection algorithm based on chain-coded plane curves. *Pattern Recognition*, 1987. vol. 20(no. 3): p. 291-296.
19. O’Gorman, L., Curvilinear feature detection from curvature estimation, in 9th International Conference on Pattern Recognition. 1988. p. 1116-1119.
20. Chetverikov, D., A Simple and Efficient Algorithm for Detection of High Curvature Points in Planar Curves, in CAIP. 2003, Springer- Verlag Berlin Heidelberg. p. 746-753.
21. Anderson, I.M. and J.C. Bezdek, Curvature and tangential deflection of discrete arcs: A theory based on the commutator of scatter matrix pairs and its application to vertex detection in planar shape data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1984. PAMI-6: p. 27-40.
22. Seeger, U. and R. Seeger, Fast corner detection in grey-level images. *Pattern Recognition Letters*, 1994. vol. 15(no. 7): p. 669-675.
23. Arrebola, F., et al., Corner detection by local histograms of contour chain code. *Electronics Letters*, 1997. vol. 33(no. 21): p. 1769-1771.
24. Arrebola, F., et al., Corner detection and curve representation by circular histograms of contour chain code. *Electronics Letters*, 1999. vol. 35(no. 13): p. 1065-1067.
25. Teh, C.-H. and R. Chin, On the detection of dominant points on digital curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989: p. 859-872.
26. Langridge, D.J., Curve encoding and detection of discontinuities. *Computer Vision, Graphics and Image Processing*, 1987. vol. 20(no. 1): p. 58-71.
27. Ogawa, H., Corner detection on digital curves based on local symmetry of the shape. *Pattern Recognition*, 1989. vol. 22(no. 4): p. 351-357.
28. Rosenberg, B., The analysis of convex blobs. *Computer Graphics and Image Processing*, 1972. vol. 1: p. 183-192.
29. Bandera, A., et al., Corner detection by means of adaptively estimated curvature function. *Electronic Letters*, 2000. vol. 36 (no. 2): p. 124-126.
30. Urdiales, C., et al., Corner detection based on adaptively filtered curvature function. *Electronics Letters*, 2003. vol. 32(no. 5): p. 426-428.
31. Sohn, K., et al., Curvature estimation and unique corner point detection for boundary representation. *IEEE International Conference on Robotics and Automation*, 1992. vol. 2(.): p. 1590-1595.
32. Ansari, N. and E.J. Delp, On detecting dominant points. *Pattern Recognition*, 1991. vol. 24(no. 5): p. 441-451.

33. Rattarangsi, A. and R.T. Chin, Scale-based detection of corners of planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992. vol. 14(no. 4): p. 430-449.
34. Rosin, P.L., Measuring corner properties. *Computer Vision and Image Understanding: CVIU*, 1999. vol. 73(no. 2): p. 291-307.
35. Lindeberg, T., Feature detection with automatic scale selection. *International Journal of Computer Vision*, 1998. vol. 30(no. 2): p. 79-116.
36. Ray, B.K. and R. Pandyan, Acord-an adaptive corner detector for planar curves. *Pattern Recognition Letters*, 2003. vol. 36(no. 3): p. 703-708.
37. Mokhtarian, F. and R. Suomela, Robust image corner detection through curvature scale space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998. vol. 20(no. 12): p. 1376-1381.
38. He, X. and N. Yung, Curvature scale space corner detector with adaptive threshold and dynamic region of support, in *17th International Conference on Pattern Recognition*. 2004. p. 791-794
39. Medioni, G. and Y. Yasumoto, Corner detection and curve representation using cubic b-splines. *Computer Vision, Graphics and Image Processing*, 1987. vol. 39(no. 3): p. 279-290.
40. Dunham, J., Optimum uniform piecewise linear approximation of planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986. vol. 8,(no. 1): p. 67-75.
41. Pavlidis, T., Problems in recognition of drawings. *Syntactic and Structural Pattern Recognition*, 1988. vol. 45(103-113).
42. Ji, Q. and R.M. Haralick, Corner detection with covariance propagation. *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 1997: p. 362-367.
43. Sankar, P. and C. Sharma, A parallel procedure for the detection of dominant points on a digital curve. *Computer Graphics and Image Processing*, 1978. vol. 7(no. 4,): p. 403-412.
44. Cooper, J., S. Venkatesh, and L. Kitchen, The dissimilarity corner detector, in *5th International Conference on Advanced Robotics*. 1991. p. 1377-1382.
45. Haralick, R.M. and L.G. Shapiro, *Computer and robot vision*, ed. Adison-Wesley. Vol. vol. 1. 1993.
46. Luo, B., A.D.J. Cross, and E.R.H. . Corner detection via topographic analysis of vector potential, in *9th British Machine Vision Conference*,. 1998.
47. Luo, B. and D. Pycock, Unified multi-scale corner detection, in *4th IASTED International Conference on Visualisation Imaging and Image Processing*,. 2004.
48. Davies, E.R., Application of the generalised hough transform to corner detection. *IEEE Proceedings on Computers and Digital Techniques*, 1988. vol. 135(no. 1): p. 49-54.
49. Ballard, D.H., Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 1981. vol. 13(no. 2): p. 111-122.
50. Shen, F. and H. Wang, Corner detection based on modified Hough transform. *Pattern Recognition Letters*, 2002. vol. 23(8): p. 1039-1049.
51. Kang, S.K., Y.C. Choung, and J.A. Park. Image corner detection using hough transform. in *IbPRIA 2005 : Iberian conference on pattern recognition and image analysis*. 2005. Estoril , PORTUGAL: Springer, Berlin.
52. Lee, J., Y. Sun, and C. Chen, Wavelet transform for corner detection, in *IEEE Conference on Systems Engineering*. 1992.
53. Lee, J.-S., Y.-N. Sun, and C.-H. Chen, Multiscale corner detection by using wavelet transform. *IEEE Transactions on Image Processing*, 1995. vol. 4(no. 1): p. 100-104.
54. Quddus, A. and M. Fahmy, Fast wavelet-based corner detection technique. *Electronics Letters*, 1999. vol. 35(no. 4): p. 287-288.
55. Chen, C.-H., J.-S. Lee, and Y.-N. Sun†, Wavelet transformation for gray-level corner detection. *Pattern Recognition*, 1995. vol 28(no. 6): p. 853-861
56. Quddus, A. and M. Gabbouja, Wavelet-based corner detection technique using optimal scale *Pattern Recognition Letters*, 2002. vol. 23(no 1-3): p. 215-220.
57. Gao, X., et al., Multiscale contour corner detection based on local natural scale and wavelet transform. *Image and Vision Computing*, 2006. vol. 25(no.6): p. 890-898.

58. Yeh, C.-H., Wavelet-based corner detection using eigenvectors of covariance matrices. *Pattern Recognition Letters*, 2003. vol. 24(no. 15): p. 2797-2806.
59. Liu, Q. and T.-S. Lin, The Corner Detection Algorithm Based on 2-D Discrete Wavelet Transform, in *Proceedings of the 2008 3rd International Conference on Innovative Computing Information and Control - Volume 00*. 2008, IEEE Computer Society.
60. Noble, J.A., Images as functions and sets. *Image and Vision Comput.*, 1992. 10(1): p. 19-29.
61. Shapiro, L.G., R.S. Macdonald, and S.R. Sternberg, Ordered structural shape matching with primitive extraction by mathematical morphology. *Pattern Recognition*, 1987. 20.
62. Lin, R.-S., C.-H. Chu, and Y.-C. Hsueh, A modified morphological corner detector. *Pattern Recognition Letters*, 1998. vol. 19(no. 3): p. 279-286.
63. Laganieri, R., A morphological operator for corner detection. *Pattern Recognition*, 1998. vol. 31(no. 11): p. 1643-1652.
64. Chen, M. and P. Yan, A multiscaling approach based on morphological filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989. vol 1(no 7): p. 694-700.
65. Paler, K., et al., Local ordered grey levels as an aid to corner detection. *Pattern Recognition* 1984. vol. 17(no. 5): p. 535-543.
66. Shen, F. and H. Wang, Real time gray level corner detector, in *6th International Conference on Control, Automation, Robotics and Vision*. 2000.
67. Smith, S.M. and J.M. Brady, SUSAN - a new approach to low level image processing. *International Journal of Computer Vision*, 1997. vol. 23(no. 1): p. 45-78.
68. Bae, S.C., I.S. Kweon, and C.D. Yoo, Cop: a new corner detector. *Pattern Recognition Letters*, 2002. vol. 23(no. 11): p. 1349-1360.
69. Trajković, M. and M. Hedley, Fast corner detection. *Image and Vision Computing*, 1998. vol. 16(no. 2): p. 75-87.
70. Rosten, E., G. Reitmayr, and T. Drummond, Real-time video annotations for augmented reality. *International Symposium on Visual Computing*, 2005.
71. Lepetit, V. and P. Fua, Keypoint recognition using randomized trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006. vol. 28(no. 9): p. 1465-1479.
72. Loy, G. and A. Zelinsky, A fast radial symmetry transform for detecting points of interest, in *7th European Conference on Computer Vision*. 2002. p. 358-368.
73. Gilles, S., *Robust Description and Matching of Image*. 1998, University of Oxford.
74. Kadir, T. and M. Brady, Scale, saliency and image description. *International Journal of Computer Vision*, 2001. vol. 45,(no. 2): p. 83-105.
75. Kadir, T., M. Brady, and A. Zisserman, An affine invariant method for selecting salient regions in images. *Proceedings of the European Conference on Computer Vision, Graphics and Image Processing*, 2004: p. 345-457.
76. Guiducci, A., Corner characterization by differential geometry techniques. *Pattern Recognition Letters*, 1988. vol. 8(no. 5): p. 311-318.
77. Rohr, K., Recognizing corners by fitting parametric models. *International Journal of Computer Vision*, 1992. vol. 9(no. 3): p. 213-230.
78. Ghosal, S. and R. Mehrotra, Zernike moment-based feature detectors, in *1st International Conference on Image Processing*. 1994. p. 934-938.
79. Liu, S.-T. and W.-H. Tsai, Moment-preserving corner detection. *Pattern Recognition*, 1990. vol. 23(no. 5): p. 441-460.
80. Canny, J., A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986. vol. 8(no. 6): p. 679-698.
81. Rangarajan, K., M. Shah, and D.v. Brackley, Optimal corner detection, in *2nd IEEE International Conference on Computer Vision and Image Understanding: CVIU*. 1988.
82. Xie, X., R. Sudhakar, and H. Zhuang, Corner detection by a cost minimization approach. *Pattern Recognition*, 1993. vol. 26(no. 8): p. 1235-1243.
83. Chen, W.-C. and P. Rockett, Bayesian labelling of corners using a grey-level corner image model, in *4th International Conference on Image Processing*,. 1997. p. 687-690.

84. Dias, P., A. Kassim, and V. Srinivasan, A neural network based corner detection method. in IEEE International Conference on Neural Networks, 1995. vol. 4: p. 2116-2120.
85. Rosten, E. and T. Drummond. Machine Learning for High Speed Corner Detection. in Computer Vision – ECCV 2006. 2006.
86. Kienzle, W., et al., Learning an interest operator from human eye movements, in 18th IEEE Conference on Computer Vision and Pattern Recognition Workshop. 2005.
87. Ebner., M. On the evolution of interest operators using genetic programming. in In Proc. EuroGP'98. 1998. Paris, France.
88. Moravec, H., Obstacle avoidance and navigation in the real world by a seeing robot rover. 1980: in tech. report CMU-RI-TR-80-03, Robotics Institute, Carnegie Mellon University & doctoral dissertation, Stanford University. Carnegie Mellon University, , available as Stanford AIM-340, CS-80-813 and republished as a Carnegie Mellon University Robotics Institute Technical Report to increase availability.
89. Ebner, M. and A. Zell. Evolving a task specific image operator. in In Proc. EvoIASP'99. 1999. Goteborg, Sweden.
90. Trujillo, L. and G. Olague, Synthesis of interest point detectors through genetic programming, in 8th annual conference on Genetic and evolutionary computation. 2006. p. 887-894.
91. Trujillo, L. and G. Olague. Using Evolution to Learn How to Perform Interest Point Detection. in ICPR 2006. 2006.
92. Beaudet, P., Rotational invariant image operators, in 4th International Conference on Pattern Recognition. 1978. p. 579-583.
93. Trujillo, L. and G. Olague, Scale Invariance for Evolved Interest Operators, in Applications of Evolutionary Computing, S.B. Heidelberg, Editor. 2007. p. 423 - 430.
94. Matas, J., et al. Robust wide-baseline stereo from maximally stable extremal regions. in Proceedings of the British Machine Vision Conference. 2002.
95. Matas, J., S. Obdrzalek, and O. Chum. Local affine frames for wide-baseline stereo. in Proceedings of 16th International Conference Pattern Recognition. 2002.
96. Mikolajczyk, K., et al., A comparison of affine region detectors. International Journal of Computer Vision, 2005. vol. 65(no. 1-2): p. 43-72.
97. Perdoch, M., J. Matas, and S. Obdrzalek. Stable affine frames on isophotes. in Proceedings of the International Conference on Computer Vision. 2007.
98. Tuytelaars, T. and L.V. Gool, Wide baseline stereo matching based on local, affinely invariant regions. Proceedings of the British Machine Vision Conference, 2000: p. 412-425.
99. Tuytelaars, T. and L.V. Gool, Matching widely separated views based on affine invariant regions. International Journal of Computer Vision, 2004. 1(59): p. 61 - 85.
100. Harris, C. and M. Stephens, A combined corner and edge detector, in Alvey Vision Conference. 1988. p. 147-151.
101. Schmid, C., R. Mohr, and C. Bauckhage, Evaluation of interest point detectors. International Journal of Computer Vision, 2000. vol. 37(no. 2): p. 151-172.
102. Schmid, S.M., R. Mohr, and C. Bauckhage. Comparing and evaluating interest point. in International Conference on Computer Vision. 1998.
103. Noble, J.A., Descriptions of image surfaces. 1989, Department of Engineering Science, University of Oxford.
104. Forstner, W., A feature-based correspondence algorithm for image matching. International Archive of Photogrammetry and Remote Sensing, 1986. vol. 26: p. 150-166.
105. Rohr, K., Localization properties of direct corner detectors. Journal of Mathematical Imaging and Vision, 1994. 4(2): p. 139-150.
106. Shi, J. and C. Tomasi, Good features to track, in in 9th IEEE Conference on Computer Vision and Pattern Recognition. 1994.
107. Kenney, C.S., et al., A condition number for point matching with application to registration and postregistration error estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003. vol. 25(no. 11): p. 1437-1454.

108. Kenney, C., M. Zuliani, and B. Manjunath, An axiomatic approach to corner detection, in 18th IEEE Conference on Computer Vision and Pattern Recognition. 2005. p. 191-197.
109. Zuliani, M., C. Kenney, and B. Manjunath, A mathematical comparison of point detectors, in Second IEEE Image and Video Registration Workshop (IVR). 2004.
110. Schmid, C. and R. Mohr, Combining gray-value invariants with local constraints for object recognition. Proceedings of the Conference on Computer Vision and Pattern Recognition., 1996: p. 872-877.
111. Zheng, Z., H. Wang, and E. Teoh, Analysis of gray level corner detection. Pattern Recognition Letters, 1999. vol. 20: p. 149-162.
112. Mikolajczyk, K., Scale and Affine Invariant Interest Point Detectors. 2002, INRIA: Grenoble.
113. Kitchen, L. and A. Rosenfeld, Gray-level corner detection. Pattern Recognition Letters, 1982. vol. 1(no. 2.): p. 95-102.
114. Nagel, H.H., Displacement vectors derived from second order intensity variations in images sequences. Computer Vision Graphics and Image Processing, 1983. 21: p. 85 - 117.
115. Dreschler, L. and H.H. Nagel, On the selection of critical points and local curvature extrema of region boundaries for interframe matching. International Conference on Pattern Recognition, 1982: p. 542 - 544.
116. Singh, A. and M. Shneier, Grey level corner detection: A generalization and a robust real time implementation. Computer Vision, Graphics and Image Processing., 1990. vol. 51(no. 1): p. 54-69.
117. Zuniga, O. and R. Haralick, Corner detection using the facet model, in 1st IEEE Conference on Computer Vision and Pattern Recognition. 1983. p. 30-37.
118. Shah, M.A. and R. Jain, Grey Level corner detection: A generalization and a robust real time implementation. Computer Vision and Image Understanding: CVIU, 1984. 51: p. 54 - 69.
119. Deriche, R. and G. Giraudon, A computational approach for corner and vertex detection. International Journal of Computer Vision, Graphics and Image Processing, 1993. vol. 10(no. 2): p. 101-124.
120. Lipschutz, M.M., Differential Geometry, ed. M.-. Hill. 1969, New York.
121. Dreschler, L. and H.-H. Nagel, Volumetric model and 3d trajectory of a moving car from monocular tv frames sequence of a street scene. Computer Graphics and Image Processing, 1982. vol. 20(no. 3): p. 199-228.
122. Wang, H. and M. Brady, Real-time corner detection algorithm for motion estimation. Image and Vision Computing, 1995. vol. 13(no. 9): p. 695-703.
123. Crowley, J.L., A representation for visual information. 1981, Carnegie Mellon University.
124. Crowley, J.L. and A. Parker, A representation for shape based on peaks and ridges in the difference of low pass transform. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1984. 6(2): p. 156- 170.
125. Bay, H., et al., Speeded-up robust features (SURF). International Journal on Computer Vision and Image Understanding, 2008. vol. 110(no. 3): p. 346-359.
126. Triggs, B., Detecting keypoints with stable position, orientation and scale under illumination changes., in in 8th European Conference on Computer Vision., 2004. p. 100-113.
127. Lindeberg, T., Scale-Space Theory in Computer Vision. 1994, Kluwer Academic Publishers.
128. Dufournaud, Y., C. Schmid, and R. Horaud, Image Matching with Scale Adjustment. Computer Vision and Image Understanding: CVIU, 2004 93(2): p. 175 - 194.
129. Mikolajczyk, K. and C. Schmid, Scale and Affine Invariant Interest Point Detectors. International Journal of Computer Vision, 2004. 60(1): p. 63 - 86.
130. Baumberg, A. Reliable feature matching across widely separated views. in Proceedings of the Conference on Computer Vision and Pattern Recognition. 2000.
131. Lindeberg, T. and J. Garding, Shape-adapted smoothing in estimation of 3-d shape cues from affine deformations of local 2-d brightness structure. . IVC, 1997. 15(6): p. 415-434.
132. Tuytelaars, T. and L.V. Gool, Content-based image retrieval based on local affinity invariant regions, in International Conference on Visual Information Systems. 1999. p. 493-500.

133. Treisman, A.M. and G. Gelade, A feature integration theory of attention. *Cognitive Psychology*, 1980. 12: p. 97 - 136.
134. Wolfe, J.M., Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review*, 1994. 1(2): p. 202 - 238.
135. Wolfe, J.M., K. Cave, and S. Franzel, Guided Search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance.*, 1989. 15: p. 419 - 433.
136. Wolfe, J.M. and G. Gancarz, Guided Search 3.0: Basic and clinical applications of vision science. Dordrecht, Nethrland: Kluwer Academic., 1996: p. 189 - 192.
137. Wolfe, J.M., Guided search 4.0: A guided search model that does not require memory for rejected distractors. *Journal of Vision, Abstracts of the 2001 VSS Meeting*, 2001. 1(3).
138. Koch, C. and S. Ullman, Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, 1985. vol. 4(no. 4): p. 219-227.
139. Milanese, R., Detecting Salient Regions in an Image: From Biological Evidence to Computer Implementation. 1993, University of Geneva.: Switzerland.
140. Milanese, R., et al. Integration of bottom-up and top-down cues for visual attention using non-linear relaxation. in *IEEE Conference on Computer Vision and Pattern Recognition*. 1994.
141. Itti, L. and C. Koch, Feature combination strategies for saliency based visual attention systems. *Journal of Electronic Imaging*, 2001. 10(1): p. 161 - 169.
142. Itti, L., C. Koch, and E. Niebur, A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998. 20(11): p. 1254-1259.
143. Navalpakkam, V., J. Rebesch, and L. Itti, Modeling the influence of task on attention. *Vision Research*, 2005. 45(2): p. 205 - 231.
144. Draper, B. and A. Lionelle. Evaluation of selective attention under similarity transform. in *International Workshop on Attention and Performance in Computer Vision*. 2003. Graz, Austria.
145. Frintrap, S., VOCUS: A Visual Attention System for Object Detection and Goal-directed Search. 2005, University of Bonn
146. Backer, G., B. Mertsching, and M. Bollman, Data and model-driven gaze control for an active-vision system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001. 23(12): p. 1415 - 1429.
147. Frintrap, S., P. Jensfelt, and H. Christensen, Simultaneous Robot Localization and Mapping Based on a Visual Attention System. *Attention in Cognitive Systems, Lecture Notes on Artificial Intelligence (LNAI)*, 2007. 4840.
148. Frintrap, S. and M. Kessel. Cognitive Data Association for Visual Person Tracking. in *IEEE Workshop on Human Detection from Mobile Platforms (HDMP '08)*. 2008. Pasadena, California.
149. Sun, Y. and R. Fisher, Object-based visual attention for computer vision. *Artificial Intelligence*, 2003. 146(1): p. 77 - 123.
150. Frintrap, S., The High Repeatability of Salient Regions, in *ECCV Workshop "Vision in Action: Efficient Strategies for Cognitive Agents in Complex Environments"*. 2008: Marseille, France
151. Ouerhani, N., Visual Attention: From Bio-Inspired Modeling to Real-Time Implementation, in *Institut de Microtechnique Université de Neuchâtel*. 2003: Switzerland.
152. Johnson, A. and M. Hebert. Object recognition by matching oriented points. in *Proceedings of the Conference on Computer Vision and Pattern Recognition*. 1997. Puerto Rico, USA.
153. Zabih, R. and J. Woodfill. Non-parametric local transforms for computing visual correspondance. in *Proceedings of the 3rd European Conference on Computer Vision*. 1994. Stockholm, Sweden.
154. Ke, Y. and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. in *Proceedings of the Conference on Computer Vision and Pattern Recognition*. 2004. Washington, USA.
155. Belongie, S., J. Malik, and J. Puzicha, Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002. vol. 24(4): p. 509 - 522.

156. Vetterli, J.K.M., *Wavelets and Subband Coding*. Prentice Hall, 1995.
157. Florack, L., et al. General intensity transformations and second order invariants. in *Proceedings of the 7th Scandinavian Conference on Image Analysis*,. 1991. Aalborg, Denmark.
158. Freeman, W. and E. Adelson, The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991. vol. 13(9): p. 891 - 906.
159. Schaffalitzky, F. and A. Zisserman. Multi-view matching for unordered image sets. in *Proceedings of the 7th European Conference on Computer Vision*. 2002. Copenhagen, Denmark.
160. Horaud, R., T. Skordas, and F. Veillon, Finding geometric and relational structures in an image. *ECCV*, 1990: p. 374 - 384.
161. Heitger, F., et al., Simulation of neural contour mechanisms: From simple to end-stopped cells. *Vision Research*, 1992. vol. 32(no. 5): p. 963-981.
162. Sebe, N., et al., Evaluation of salient point techniques. *Image and Vision Comput.*, 2001. vol. 21(no. 13-14): p. 1087 - 1095.
163. Mohanna, F. and F. Mokhtarian. Performance evaluation of corner detection algorithms under affine and similarity transforms. in *Proceedings of the British Machine Vision Conference*. 2001.
164. Mokhtarian, F. and F. Mohanna, Performance evaluation of corner detectors using consistency and accuracy measures *Computer Vision and Image Understanding: CVIU*, 2006. vol. 102(no. 1): p. 81 - 94.
165. Tissainayagam, P. and D. Suter, Assessing the performance of corner detectors for point feature tracking applications *Image and Vision Computing*, 2004. vol. 22(no. 8): p. 663-679.
166. Lowe, D.G., Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004. vol. 60(no. 2): p. 91-110.

RT_019, febrero 2010

Aprobado por el Consejo Científico CENATAV

Derechos Reservados © CENATAV 2010

Editor: Lic. Lucía González Bayona

Diseño de Portada: DCG Matilde Galindo Sánchez

RNPS No. 2142

ISSN 2072-6287

Indicaciones para los Autores:

Seguir la plantilla que aparece en www.cenatav.co.cu

C E N A T A V

7ma. No. 21812 e/218 y 222, Rpto. Siboney, Playa;

Ciudad de La Habana. Cuba. C.P. 12200

Impreso en Cuba

