

# Descriptores

Guillermo Steiner

Centro de Investigación en Informática para la Ingeniería  
Universidad Tecnológica Nacional, F.R.C.

<http://cii.frc.utn.edu.ar>

Córdoba, Argentina



Diciembre 2011

# Introducción

Un descriptor es usado para transformar una imagen a un espacio de características, generalmente de alta dimensionalidad, que permita resaltar alguna característica de la imagen de forma de poder ser calificada en conjuntos definibles.

## Forma Simple

Vector que contenga los píxel de la misma, luego por cross-correlation puedo buscar similitud entre regiones.

## Desventajas

- Alta dimensionalidad del descriptor.
- La invariancia a la iluminación la implementa el buscador de similitud.
- No es invariante a Rotación, Escala ni Transformación afina.

# Introducción

Un descriptor es usado para transformar una imagen a un espacio de características, generalmente de alta dimensionalidad, que permita resaltar alguna característica de la imagen de forma de poder ser calificada en conjuntos definibles.

## Forma Simple

Vector que contenga los píxel de la misma, luego por cross-correlation puedo buscar similitud entre regiones.

## Desventajas

- Alta dimensionalidad del descriptor.
- La invariancia a la iluminación la implementa el buscador de similitud.
- No es invariante a Rotación, Escala ni Transformación afina.

# Introducción

Un descriptor es usado para transformar una imagen a un espacio de características, generalmente de alta dimensionalidad, que permita resaltar alguna característica de la imagen de forma de poder ser calificada en conjuntos definibles.

## Forma Simple

Vector que contenga los píxel de la misma, luego por cross-correlation puedo buscar similitud entre regiones.

## Desventajas

- Alta dimensionalidad del descriptor.
- La invariancia a la iluminación la implementa el buscador de similitud.
- No es invariante a Rotación, Escala ni Transformación afina.

# Tipos de descriptores

- 1 Descriptores basados en distribución.
- 2 Técnicas de Espacio-Frecuencia.
- 3 Descriptores Diferenciales.
- 4 Otras Técnicas.

# Descriptores basados en distribución

Es uno de los tipos de descriptores mas utilizados.

Emplea histogramas para representar diferentes características que aparece en la figura.

## Ejemplo Simple

Un simple descriptor sería la distribución de intensidad de píxel representado por un histograma.

# Descriptores basados en distribución

Es uno de los tipos de descriptores mas utilizados.

Emplea histogramas para representar diferentes características que aparece en la figura.

## Ejemplo Simple

Un simple descriptor sería la distribución de intensidad de píxel representado por un histograma.

# Descriptores basados en distribución

Es uno de los tipos de descriptores mas utilizados.

Emplea histogramas para representar diferentes características que aparece en la figura.

## Ejemplo Simple

Un simple descriptor sería la distribución de intensidad de píxel representado por un histograma.

# Descriptores basados en distribución Ejemplos

Johnson and Hebert. Su representación es un histograma de las posiciones de puntos en la vecindad de un punto de interés 3D. El histograma en 2D representa la distancia del centro del punto al píxel y su intensidad.

- 1 SIFT(Scale-invariant feature transform)
- 2 SURF(Speeded-Up Robust Features)
- 3 DAISY
- 4 HOG (Histograms of Oriented Gradients)

# Técnicas de Espacio-Frecuencia

Describen el contenido de frecuencia en una imagen. La transformada de Fourier descompone la imagen en sus funciones básicas pero la relación espacial entre puntos no es explícita y las funciones básicas son infinitas esto dificulta el enfoque de local.

## Ejemplo

La transformada de Gabor supera esos problemas, pero un gran número de filtros Gabor es requerido para capturar pequeños cambios en la frecuencia y orientación. Tanto Gabor como Wavelets son a menudo desarrolladas en el contexto de clasificación de textura.

# Técnicas de Espacio-Frecuencia

Describen el contenido de frecuencia en una imagen. La transformada de fourier descompone la imagen en sus funciones básicas pero la relación espacial entre puntos no es explícita y las funciones básicas son infinitas esto dificulta el enfoque de local.

## Ejemplo

La transformada de Gabor supera esos problemas, pero un gran número de filtros Gabor es requerido para capturar pequeños cambios en la frecuencia y orientación. Tanto Gabor como Wavelets son a menudo desarrolladas en el contexto de clasificación de textura.

# Descriptores Diferenciales

Un set de derivadas de imagenes calculada hasta un cierto orden proximas a un punto vecino.

## Ejemplos

Las propiedades de derivatives local fueron investigadas por Koenderink y Van Doorn. Florack invariantes diferencial, el cual combina componentes de el local jet para obtener invariancia a la rotación.

## Ejemplos

Freeman y Adelson desarrolló filtros dirigibles, que dirigen las derivadas en una dirección determinada dada por los componentes de el jet local. Una estimación estable de las derivadas es obtenida por convolución con las derivadas gaussianas.

# Descriptores Diferenciales

Un set de derivadas de imagenes calculada hasta un cierto orden proximas a un punto vecino.

## Ejemplos

Las propiedades de derivatives local fueron investigadas por Koenderink y Van Doorn. Florack invariantes diferencial, el cual combina componentes de el local jet para obtener invariancia a la rotación.

## Ejemplos

Freeman y Adelson desarrolló filtros dirigibles, que dirigen las derivadas en una dirección determinada dada por los componentes de el jet local. Una estimación estable de las derivadas es obtenida por convolución con las derivadas gaussianas.

# Descriptores Diferenciales

Un set de derivadas de imagenes calculada hasta un cierto orden proximas a un punto vecino.

## Ejemplos

Las propiedades de derivatives local fueron investigadas por Koenderink y Van Doorn. Florack invariantes diferencial, el cual combina componentes de el local jet para obtener invariancia a la rotación.

## Ejemplos

Freeman y Adelson desarrolló filtros dirigibles, que dirigen las derivadas en una dirección determinada dada por los componentes de el jet local. Una estimación estable de las derivadas es obtenida por convolución con las derivadas gaussianas.

# Otras Tecnicas

Momentos invariantes generalizados el cual describe la naturaleza multispectral de los datos de la imagen. El momento central invariante conbinado definido por  $M_{pq}^a = \iint_{\Omega} x^p y^q [I(x, y)]^a dx dy$  de orden  $p + q$  y grado  $a$ . El momento caracteriza forma e intensidad de la distribución en una región. Ellos son independiente y puede ser fácilmente computado para cualquier orden y grado. De todas maneras los momentos de alto orden y grado son sensibles a pequeñas distorsiones geométricas y fotométricas. computando las invariantes, reduce el numero de dimensiones.

# SIFT(Scale-invariant feature transform)

Un método patentado por Lowe que permite detectar y describir puntos de interés.

- 1 Permite detectar objetos mas complejos utilizando una serie de puntos de interés que lo forman.
- 2 Los puntos de interés detectados por SIFT son puntos fuertes que permanecen estable en diferentes escalas.
- 3 Esto permite localizar en un mismo objeto los mismos puntos de interés independientemente de la escala que se capture el objeto.
- 4 por otro lado el descriptor del punto es invariante a la escala y a la orientación y parcialmente invariante a distorsiones afinas y cambios de iluminación.
- 5 Esto permite la descripción de un objeto, utilizando para ello la descripción de todos sus puntos de interés, la muy buena calidad de los descriptores obtenidos, permite detectar un nuevo objeto con solo encontrar correspondencia en pocos puntos de interés, permitiendo detecciones con alto nivel de oclusión.

# SIFT(Scale-invariant feature transform)

Un método patentado por Lowe que permite detectar y describir puntos de interés.

- 1 Permite detectar objetos mas complejos utilizando una serie de puntos de interés que lo forman.
- 2 Los puntos de interés detectados por SIFT son puntos fuertes que permanecen estable en diferentes escalas.
- 3 Esto permite localizar en un mismo objeto los mismos puntos de interés independientemente de la escala que se capture el objeto.
- 4 por otro lado el descriptor del punto es invariante a la escala y a la orientación y parcialmente invariante a distorsiones afinas y cambios de iluminación.
- 5 Esto permite la descripción de un objeto, utilizando para ello la descripción de todos sus puntos de interés, la muy buena calidad de los descriptores obtenidos, permite detectar un nuevo objeto con solo encontrar correspondencia en pocos puntos de interés, permitiendo detecciones con alto nivel de oclusión.

# SIFT(Scale-invariant feature transform)

Un método patentado por Lowe que permite detectar y describir puntos de interés.

- 1 Permite detectar objetos mas complejos utilizando una serie de puntos de interés que lo forman.
- 2 Los puntos de interés detectados por SIFT son puntos fuertes que permanecen estable en diferentes escalas.
- 3 Esto permite localizar en un mismo objeto los mismos puntos de interés independientemente de la escala que se capture el objeto.
- 4 por otro lado el descriptor del punto es invariante a la escala y a la orientación y parcialmente invariante a distorsiones afinas y cambios de iluminación.
- 5 Esto permite la descripción de un objeto, utilizando para ello la descripción de todos sus puntos de interés, la muy buena calidad de los descriptores obtenidos, permite detectar un nuevo objeto con solo encontrar correspondencia en pocos puntos de interés, permitiendo detecciones con alto nivel de oclusión.

# SIFT(Scale-invariant feature transform)

Un método patentado por Lowe que permite detectar y describir puntos de interés.

- 1 Permite detectar objetos mas complejos utilizando una serie de puntos de interés que lo forman.
- 2 Los puntos de interés detectados por SIFT son puntos fuertes que permanecen estable en diferentes escalas.
- 3 Esto permite localizar en un mismo objeto los mismos puntos de interés independientemente de la escala que se capture el objeto.
- 4 por otro lado el descriptor del punto es invariante a la escala y a la orientación y parcialmente invariante a distorsiones afinas y cambios de iluminación.
- 5 Esto permite la descripción de un objeto, utilizando para ello la descripción de todos sus puntos de interés, la muy buena calidad de los descriptores obtenidos, permite detectar un nuevo objeto con solo encontrar correspondencia en pocos puntos de interés, permitiendo detecciones con alto nivel de oclusión.

# SIFT(Scale-invariant feature transform)

Un método patentado por Lowe que permite detectar y describir puntos de interés.

- 1 Permite detectar objetos mas complejos utilizando una serie de puntos de interés que lo forman.
- 2 Los puntos de interés detectados por SIFT son puntos fuertes que permanecen estable en diferentes escalas.
- 3 Esto permite localizar en un mismo objeto los mismos puntos de interés independientemente de la escala que se capture el objeto.
- 4 por otro lado el descriptor del punto es invariante a la escala y a la orientación y parcialmente invariante a distorsiones afinas y cambios de iluminación.
- 5 Esto permite la descripción de un objeto, utilizando para ello la descripción de todos sus puntos de interés, la muy buena calidad de los descriptores obtenidos, permite detectar un nuevo objeto con solo encontrar correspondencia en pocos puntos de interés, permitiendo detecciones con alto nivel de oclusión.

# SIFT(Scale-invariant feature transform)

Un método patentado por Lowe que permite detectar y describir puntos de interés.

- 1 Permite detectar objetos mas complejos utilizando una serie de puntos de interés que lo forman.
- 2 Los puntos de interés detectados por SIFT son puntos fuertes que permanecen estable en diferentes escalas.
- 3 Esto permite localizar en un mismo objeto los mismos puntos de interés independientemente de la escala que se capture el objeto.
- 4 por otro lado el descriptor del punto es invariante a la escala y a la orientación y parcialmente invariante a distorsiones afinas y cambios de iluminación.
- 5 Esto permite la descripción de un objeto, utilizando para ello la descripción de todos sus puntos de interés, la muy buena calidad de los descriptores obtenidos, permite detectar un nuevo objeto con solo encontrar correspondencia en pocos puntos de interés, permitiendo detecciones con alto nivel de oclusión.

# Método

## Detección de extremos en Espacio de Escala

Una búsqueda en toda la imagen y en toda las escalas de puntos potenciales de interés.

## Localización de puntos de interés

Se determina los puntos de interés basados en los puntos potenciales que posean una buena estabilidad.

## Asignación de Orientación

A cada punto de interés se asigna una o varias direcciones, las futuras operaciones se realizarán sobre la imagen transformada según la orientación, escala y posición donde lo que provee invariancia a esas transformaciones

## Descripción del punto de interés

Se mide alrededor del punto de interés y a la escala determinada los gradientes de la imagen, generando una representación que permite distorsiones y cambios en la iluminación.

# Método

## Detección de extremos en Espacio de Escala

Una búsqueda en toda la imagen y en toda las escalas de puntos potenciales de interés.

## Localización de puntos de interés

Se determina los puntos de interés basados en los puntos potenciales que posean una buena estabilidad.

## Asignación de Orientación

A cada punto de interés se asigna una o varias direcciones, las futuras operaciones se realizarán sobre la imagen transformada según la orientación, escala y posición donde lo que provee invariancia a esas transformaciones

## Descripción del punto de interés

Se mide alrededor del punto de interés y a la escala determinada los gradientes de la imagen, generando una representación que permite distorsiones y cambios en la iluminación.

# Método

## Detección de extremos en Espacio de Escala

Una búsqueda en toda la imagen y en toda las escalas de puntos potenciales de interés.

## Localización de puntos de interés

Se determina los puntos de interés basados en los puntos potenciales que posean una buena estabilidad.

## Asignación de Orientación

A cada punto de interés se asigna una o varias direcciones, las futuras operaciones se realizarán sobre la imagen transformada según la orientación, escala y posición donde lo que provee invariancia a esas transformaciones

## Descripción del punto de interés

Se mide alrededor del punto de interés y a la escala determinada los gradientes de la imagen, generando una representación que permite distorsiones y cambios en la iluminación.

# Método

## Detección de extremos en Espacio de Escala

Una búsqueda en toda la imagen y en toda las escalas de puntos potenciales de interés.

## Localización de puntos de interés

Se determina los puntos de interés basados en los puntos potenciales que posean una buena estabilidad.

## Asignación de Orientación

A cada punto de interés se asigna una o varias direcciones, las futuras operaciones se realizarán sobre la imagen transformada según la orientación, escala y posición donde lo que provee invariancia a esas transformaciones

## Descripción del punto de interés

Se mide alrededor del punto de interés y a la escala determinada los gradientes de la imagen, generando una representación que permite distorsiones y cambios en la iluminación.

# Detección de extremos en Espacio de Escala

## Objetivo

Encontrar puntos de interés que puedan ser detectados nuevamente desde diferentes puntos de vista y escala en el mismo objeto (invariancia en escala)

## Como

Buscando puntos estables a través de toda las posibles escalas, utilizando una función continua de escala conocida como "espacio de escala"

El único kernel posible es la función Gaussiana, si se define una espacio de escala de una imagen como  $L(x, y, \sigma)$ , esto es el resultado de la Convolución entre una Gaussiana  $G(x, y, \sigma)$  con una imagen  $I(x, y)$

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

donde

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

# Detección de extremos en Espacio de Escala

## Objetivo

Encontrar puntos de interés que puedan ser detectados nuevamente desde diferentes puntos de vista y escala en el mismo objeto (invariancia en escala)

## Como

Buscando puntos estables a través de toda las posibles escalas, utilizando una función continua de escala conocida como “espacio de escala”

El único kernel posible es la función Gaussiana, si se define una espacio de escala de una imagen como  $L(x, y, \sigma)$ , esto es el resultado de la Convolución entre una Gaussiana  $G(x, y, \sigma)$  con una imagen  $I(x, y)$

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

donde

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

# Detección de extremos en Espacio de Escala

## Objetivo

Encontrar puntos de interés que puedan ser detectados nuevamente desde diferentes puntos de vista y escala en el mismo objeto (invariancia en escala)

## Como

Buscando puntos estables a través de toda las posibles escalas, utilizando una función continua de escala conocida como “espacio de escala”

El único kernel posible es la función Gaussiana, si se define una espacio de escala de una imagen como  $L(x, y, \sigma)$ , esto es el resultado de la Convolución entre una Gaussiana  $G(x, y, \sigma)$  con una imagen  $I(x, y)$

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

donde

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

## Detección de extremos en Espacio de Escala

La detección entonces del punto de interés, se realiza por el método de la segunda derivada, el cual como se vio anteriormente se puede reducir a la resta de dos gaussianas separadas por un factor  $k$

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (3)$$

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned}$$

De esta forma el cálculo de diferencia de Gaussiana, puede ser fácilmente realiza con la resta de una imagen suavizadas por dos diferentes gaussianas.

# Detección de extremos en Espacio de Escala

La diferencia de gaussiana provee una aproximación cerrada a la Laplaciana de Gaussiana de espacio normalizado  $\sigma^2 \nabla^2 G$  donde el factor  $\sigma^2$  es requerido para normalizar la escala, esto puede ser visto por la difusión del calor, donde reemplazamos  $t = \sigma^2$

$$\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G \quad (4)$$

## Detección de extremos en Espacio de Escala

Aproximando la derivada de  $G$  entre la escala  $k\sigma$  y  $\sigma$

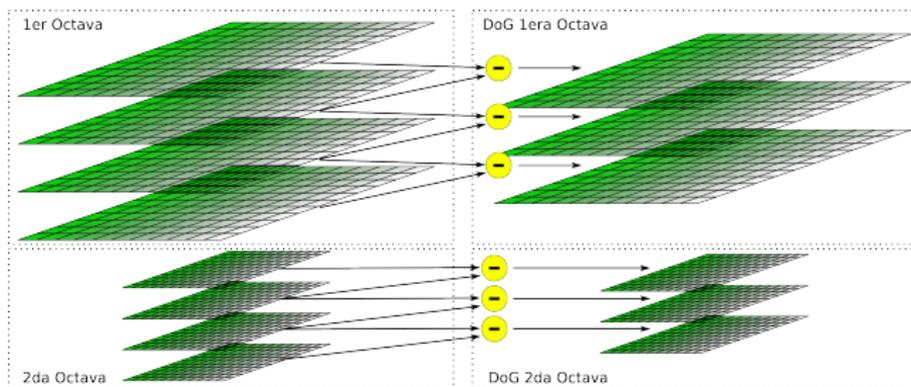
$$\sigma \nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma} \quad (5)$$

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k - 1)\sigma^2 \nabla^2 G \quad (6)$$

- 1 De esta forma se encuentra la relación entre la diferencia de gaussiana y el Laplaciano de la gaussiana, permaneciendo  $(k - 1)$  constante, puede quitarse de la ecuación,
- 2 El error de la aproximación tiende a 0 cuando  $k$  tiende a 1
- 3 Por otro lado, en la practica cambios significativos en la escala tales como  $k = \sqrt{2}$  no impactaron en la estabilidad de la detección de los puntos extremos.

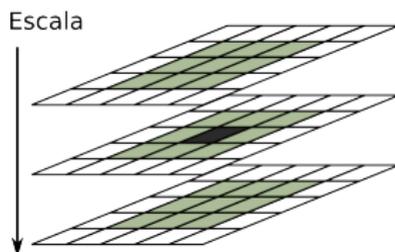
## Pirámide de resolución

- 1 Se divide el espacio en octavas separadas  $k = 2$ ,
- 2 Dentro de cada octava, se divide en intervalos  $k = 2^{1/s}$  donde  $s$  es la cantidad de niveles de la octava, y se crean  $s + 3$  imágenes convolución con el kernel gaussiano
- 3 Se realiza una resta entre imágenes adyacentes creando las diferencias de gaussiana,
- 4 Se reduce el tamaño de la imagen que posee el  $\sigma$  del doble del inicial obteniendo la primera imagen para la siguiente octava



# Detección de Extremos

Para detectar los extremos en  $D(x, y, \sigma)$ , se busca el máximo local en los  $3 \times 9$ , el máximo local será entonces el punto que resulte mayor o menor a todos estos vecinos.



## Localización de los puntos de interés

Una vez detectado los extremos, se implementa algún método para detectar máximos en subpixel y finalmente Lowe implementa un método de estimación de respuesta a corner por medio de la matriz hessiana.

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (7)$$

Se calcula la Traza y Determinante (Harris)

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta \quad (8)$$

$$Det(H) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta \quad (9)$$

Se toma un factor que relaciona ambos valores  $\alpha = r\beta$  y hacemos

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r + 1)^2}{r} \quad (10)$$

para chequear entonces la curvatura de un potencial punto de interes

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r + 1)^2}{r} \quad (11)$$

donde para los experimentos se usó  $r = 10$

## Determinación de la Orientación

Determinar de la orientación permite que el punto de interés sea invariante a la rotación, otros descriptores directamente que son invariante a la rotación, no poseen la información de la rotación necesaria como información en algunos casos.

$$G[f(x, y)] = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x} f(x, y) \\ \frac{\partial}{\partial y} f(x, y) \end{bmatrix} \quad (12)$$

$$\|G\|_{L2} = \sqrt{G_x^2 + G_y^2} \quad (13)$$

$$\angle G = \tan^{-1} \left( \frac{G_y}{G_x} \right) \quad (14)$$

Un histograma de orientación es realizado con los puntos alrededor del punto de interés. El histograma tendrá 36 bins que discretiza los 360 grados, Cada píxel sumado a el histograma es pesado por su gradiente y por una gaussiana con  $\sigma$  1,5 veces la escala.

El pico mas alto en el histograma será considerado la dirección del punto, otro picos cuyo valor supere el 80 % del pico máximo, serán también direcciones, finalmente se interpola con una parábola.

## Descripción del punto de interés

Una vez localizado el punto de interés y su orientación, se obtiene un punto con una determinada ubicación escala y orientación, esto permite encontrar un descriptor que resulte relativamente invariante a estos parámetros.

El próximo paso es computar ese descriptor para la región local de la imagen, que sea invariante a los demás parámetros tales como iluminación y cambios en el punto de vista.

Edelman, Intrator y Poggio (1997) proponen un modelo basado en la visión biológica, en particular las neuronas de la corteza visual primaria, Estas neuronas responden a el gradiente de una particular orientación y a frecuencia espacial, pero permitiendo pequeños corrimientos en la ubicación de los gradientes en la retina, permitiendo pequeños cambios en el punto de vista de objetos 3D, Lowe se inspira en esa idea.

## Descripción del punto de interés

Una vez localizado el punto de interés y su orientación, se obtiene un punto con una determinada ubicación escala y orientación, esto permite encontrar un descriptor que resulte relativamente invariante a estos parámetros.

El próximo paso es computar ese descriptor para la región local de la imagen, que sea invariante a los demás parámetros tales como iluminación y cambios en el punto de vista.

Edelman, Intrator y Poggio (1997) proponen un modelo basado en la visión biológica, en particular las neuronas de la corteza visual primaria, Estas neuronas responden a el gradiente de una particular orientación y a frecuencia espacial, pero permitiendo pequeños corrimientos en la ubicación de los gradientes en la retina, permitiendo pequeños cambios en el punto de vista de objetos 3D, Lowe se inspira en esa idea.

## Descripción del punto de interés

Una vez localizado el punto de interés y su orientación, se obtiene un punto con una determinada ubicación escala y orientación, esto permite encontrar un descriptor que resulte relativamente invariante a estos parámetros.

El próximo paso es computar ese descriptor para la región local de la imagen, que sea invariante a los demás parámetros tales como iluminación y cambios en el punto de vista.

Edelman, Intrator y Poggio (1997) proponen un modelo basado en la visión biológica, en particular las neuronas de la corteza visual primaria, Estas neuronas responden a el gradiente de una particular orientación y a frecuencia espacial, pero permitiendo pequeños corrimientos en la ubicación de los gradientes en la retina, permitiendo pequeños cambios en el punto de vista de objetos 3D, Lowe se inspira en esa idea.

# Descripción del punto de interés

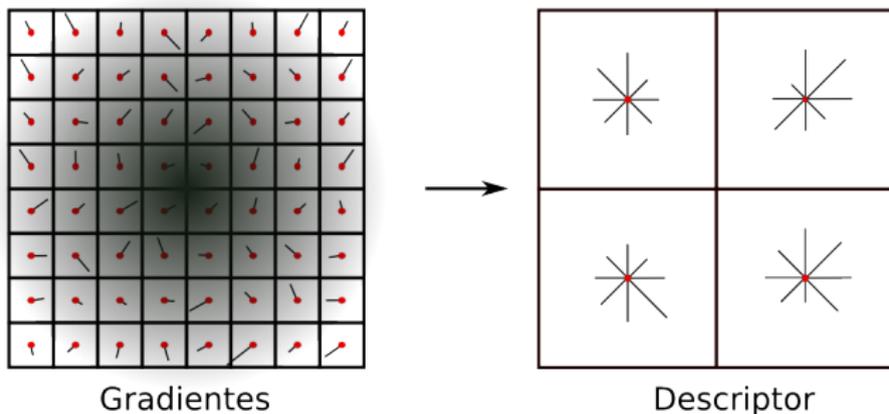


Figura: Descriptor SIFT para el caso de 4 regiones (2x2) de 4x4 gradientes cada una

## Representación del Descriptor

- Para eficiencia se calcula la magnitud y gradiente en todas la imagen y en todas las escalas
- Por cada punto de interés se calcula las magnitudes de todos los gradientes vecinos afectándolos con el peso de una gaussiana de  $\sigma$  1,5 veces la escala (permite pequeños desplazamientos y da menor importancia a los gradientes lejanos)
- Sus orientaciones, será rotadas de acuerdo la orientación que posee el punto de interés (permite al descriptor ser invariante a la rotación)
- La región del punto de interés, es dividida en regiones de por ejemplo 4x4 gradientes.
- En cada región se calcula el histograma de orientaciones llevando estos a solo 8 direcciones posibles

Finalmente se obtiene un vector de tamaño igual a 8, la cantidad de regiones, su tamaño y el numero de direcciones posibles, varia en búsqueda de lo mas óptimo, Trabajo original 4x4 regiones de 8 bins cada una  $4 \times 4 \times 8 = 128$

# Representación del Descriptor Normalización

Para reducir los efectos de la iluminación, en primer lugar el vector es normalizado a la unidad, un cambio en el constaste significa que el gradiente quedará multiplicado por esa constante la cual es anulada en una normalización, mientras que los cambios de brillo no tienen efecto en los gradientes.

Para reducir la influencia de grandes gradientes, se truncan todos aquellos que sean mayores a 0,2 (una vez normalizado), luego se renormaliza disminuyendo la influencia de estos y priorizando para el matching la distribución de los gradientes.

## Representación del Descriptor Normalización

Para reducir los efectos de la iluminación, en primer lugar el vector es normalizado a la unidad, un cambio en el constante significa que el gradiente quedará multiplicado por esa constante la cual es anulada en una normalización, mientras que los cambios de brillo no tienen efecto en los gradientes.

Para reducir la influencia de grandes gradientes, se truncan todos aquellos que sean mayores a 0,2 (una vez normalizado), luego se renormaliza disminuyendo la influencia de estos y priorizando para el matching la distribución de los gradientes.

# SURF(Speeded-Up Robust Features)

Detector y clasificador de puntos de interés, inspirado en SIFT

La diferencia principal con SIFT, es el cálculo de los gradientes, donde la aproximación se realiza mediante filtros Haar desde imágenes integrales

# SURF(Speeded-Up Robust Features)

Detector y clasificador de puntos de interés, inspirado en SIFT

La diferencia principal con SIFT, es el cálculo de los gradientes, donde la aproximación se realiza mediante filtros Haar desde imágenes integrales

# Detección de Punto de Interés

Para la detección de puntos de interés, se basa en una aproximación de la matriz hessiana, esta aproximación, permite el cálculo por medio de imágenes integrales

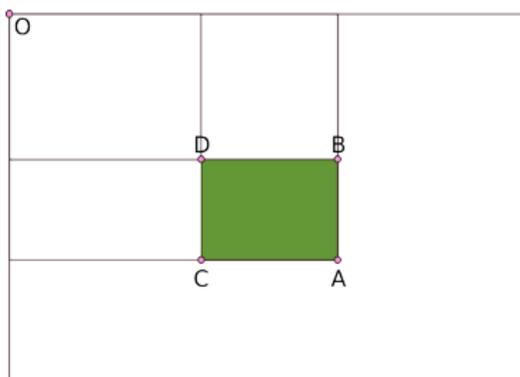
$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (15)$$

## Imágenes Integrales

Permiten el rápido cálculo de filtros de convolución tipo box.

Es una nueva imagen del mismo tamaño de la imagen original donde por cada punto  $\mathbf{x} = (x, y)^T$  tendrá la suma de las intensidades para todos los puntos con coordenada  $(i, j)$  donde  $i \leq x$  y  $j \leq y$

$$I_{\Sigma}(\mathbf{x}) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j) \quad (16)$$



**Figura:** Ejemplo de el cálculo de la sumatoria de intensidades dentro de un rectángulo

$$\Sigma = A - B - C + D$$

# Matriz Hessiana

El detector está basado en la matriz hessiana, de manera de detectar estructura de blob cuando el determinante de la misma es máximo.

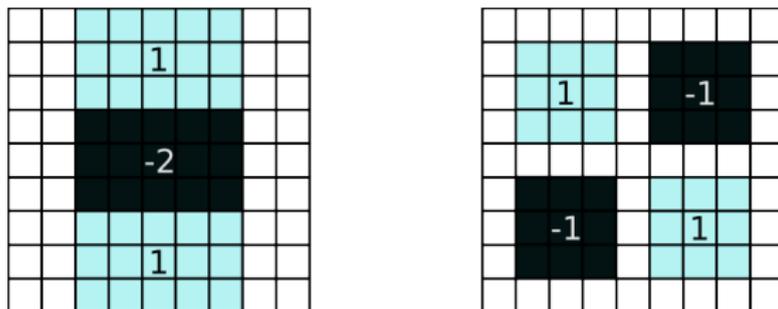
Tenemos entonces que para un punto  $\mathbf{x} = (x, y)$  y escala  $\sigma$

$$\mathcal{H}(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix} \quad (17)$$

donde por ejemplo  $L_{xx}(\mathbf{x}, \sigma)$  es la convolución del Laplaciano de la gaussiana  $\frac{\partial^2}{\partial x^2} G(\sigma)$  con la imagen  $I$

## Matriz Hessiana

Utilizando una aproximación de esta matriz hessiana a filtros box, podemos evaluarla con muy poco costo computacional e independiente del tamaño del filtro.



**Figura:** Aproximación de la derivada de segundo orden en  $y$  ( $D_{yy}$ ) y en dirección  $xy$  ( $D_{xy}$ ) el tamaño de  $9 \times 9$  corresponde a un filtro con  $\sigma = 1, 2$

Finalmente calculamos el determinante aproximado de la matriz hessiana.

$$\det(\mathcal{H}_{approx}) = D_{xx}D_{yy} - (wD_{xy})^2 \quad (18)$$

Donde  $w$  es un peso para balancear las respuesta del filtro

## Representación en el espacio de escala

Los puntos de interés son necesarios detectarlos en múltiple escala, SIFT hace uso de una pirámide de resolución, aplicando sucesivamente filtros gaussianos de un determinado  $\sigma$  y realizando un down-sampling cuando se podía,

SURF utiliza el mismo filtro discretizado, simplemente aumentando su tamaño donde el costo computacional no depende del tamaño del filtro, para el caso ejemplo con un filtro de  $9 \times 9$  corresponde a un filtro de  $\sigma = 1, 2$  (detección en máxima resolución), en las siguientes capas se aumentará gradualmente el tamaño del filtro

## Representación en el espacio de escala

Los punto de interés son necesario detectarlo en múltiple escala, SIFT hacia uso de una pirámide de resolución, aplicando sucesivamente filtros gaussianos de un determinado  $\sigma$  y realizando un down-sampling cuando se podía,

SURF utiliza el mismo filtro discretizado, simplemente aumentando su tamaño donde el costo computacional no depende del tamaño del filtro, para el caso ejemplo con un filtro de  $9 \times 9$  corresponde a un filtro de  $\sigma = 1, 2$  (detección en máxima resolución), en las siguientes capas se aumentará gradualmente el tamaño del filtro

# Descriptor

## similar a SIFT

El descriptor se basa en la distribución de intensidades en la vecindad del punto de interés.

Pero para el cálculo se utiliza la respuesta de Haar wavelets en  $x$  e  $y$ .

## Pasos

- Cálculo de la dirección predominante por medio de una región circular
- Se construye una región cuadrada alineada a la orientación previamente calculada y se extrae el descriptor

# Descriptor

## similar a SIFT

El descriptor se basa en la distribución de intensidades en la vecindad del punto de interés.

Pero para el cálculo se utiliza la respuesta de Haar wavelets en  $x$  e  $y$ .

## Pasos

- Cálculo de la dirección predominante por medio de una región circular
- Se construye una región cuadrada alineada a la orientación previamente calculada y se extrae el descriptor

# Descriptor

## similar a SIFT

El descriptor se basa en la distribución de intensidades en la vecindad del punto de interés.

Pero para el cálculo se utiliza la respuesta de Haar wavelets en  $x$  e  $y$ .

## Pasos

- Cálculo de la dirección predominante por medio de una región circular
- Se construye una región cuadrada alineada a la orientación previamente calculada y se extrae el descriptor

## Cálculo de la Orientación

- Se calcula la respuesta de Haar wavelet en  $x$  e  $y$  dentro de un círculo de radio  $6s$  donde  $s$  es la escala del punto de interés, el tamaño del filtro dependerá también de la escala (longitud  $4s$ ).
- La respuesta de los wavelets son pesadas con una gaussiana de  $\sigma = 2s$  centrada en el punto de interés,
- Representando las respuestas en un plano  $dx, dy$ , se calcula la orientación dominante, sumando las respuestas de todos los puntos que se encuentran dentro de una ventana de tamaño  $\pi/3$ ,
- La suma de las respuesta en las dos direcciones da la magnitud de un vector centrado en la ventana.
- El mayor vector sobre todas las ventanas define la orientación del descriptor.

En muchas aplicaciones la invariancia a la rotación no es necesaria, o la rotación esperada es pequeña ( $\pm 15^\circ$ ), en estos casos puede eliminarse esta sección (U-SURF)

## Cálculo de Descriptor

- Comienza con un cuadrado centrado en el punto de interés y orientado según la dirección del mismo, el tamaño de este cuadrado es  $20s$ .
- La región es dividida en  $4 \times 4$  cuadrados mas pequeños o sub regiones.
- Por cada una de esas regiones se computa la respuesta de Haar wavelet de acuerdo a  $5 \times 5$  puntos de muestreo.
- llamaremos  $d_x$  y  $d_y$  a las respuesta de Haar wavelet en x e y respectivamente, con un filtro de tamaño  $2s$  para aumentar la robustez pesamos estas respuestas con una gaussiana de  $\sigma = 3, 3s$
- Se calcula las sumatorias de cada sub-región en cada dirección, para obtener información de cambios de polaridad, calculamos además la suma de sus valores absolutos,
- Con estos resultados obtenemos un vector de describe las variaciones de la intensidad en una subregión.

$$\mathbf{v} = \left( \sum d_x, \sum d_y, \sum |d_x|, \sum |d_y| \right) \quad (19)$$

# Cálculo de Descriptor

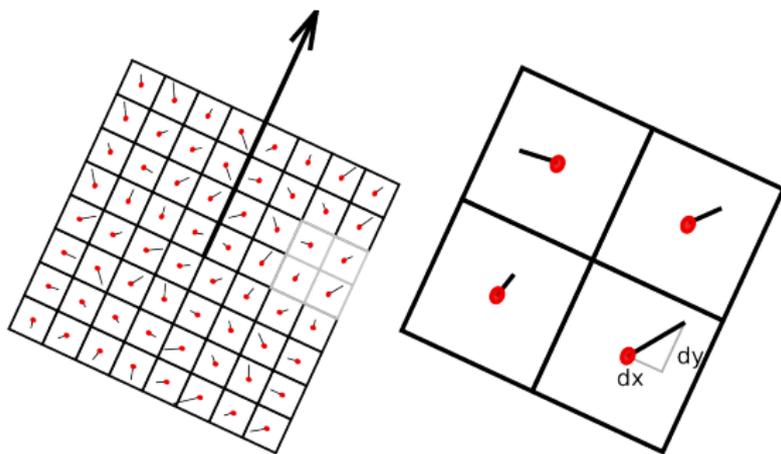


Figura: Descriptor SURF

Para optimizar el cálculo, los filtros Haar son aplicados a la imagen sin rotar y luego interpoladas de acuerdo a la rotación calculada.

# DAISY

Basados en descripción de regiones al igual que SIFT y SURF, es muy eficiente en el cálculo denso.

## En SIFT

Se calcula un gradiente y que el mismo contribuye con su magnitud multiplicada por un peso al bin que corresponda su dirección, el peso dependerá de la distancia que tiene el gradiente a el punto característico.

## DAISY

El histograma será calculado una sola vez por región y reutilizado para todo los píxel cercanos.

Para ello, se reemplaza la suma pesada de gradientes normalizados por convoluciones de gradientes en determinadas direcciones con varios filtros gaussianos.

Esto mantiene la invarianza de SIFT pero permite realizar los cálculos mucho mas rápido.

# DAISY

Basados en descripción de regiones al igual que SIFT y SURF, es muy eficiente en el cálculo denso.

## En SIFT

Se calcula un gradiente y que el mismo contribuye con su magnitud multiplicada por un peso al bin que corresponda su dirección, el peso dependerá de la distancia que tiene el gradiente a el punto característico.

## DAISY

El histograma será calculado una sola vez por región y reutilizado para todo los píxel cercanos.

Para ello, se reemplaza la suma pesada de gradientes normalizados por convoluciones de gradientes en determinadas direcciones con varios filtros gaussianos.

Esto mantiene la invarianza de SIFT pero permite realizar los cálculos mucho mas rápido.

# DAISY

Basados en descripción de regiones al igual que SIFT y SURF, es muy eficiente en el cálculo denso.

## En SIFT

Se calcula un gradiente y que el mismo contribuye con su magnitud multiplicada por un peso al bin que corresponda su dirección, el peso dependerá de la distancia que tiene el gradiente a el punto característico.

## DAISY

El histograma será calculado una sola vez por región y reutilizado para todo los píxel cercanos.

Para ello, se reemplaza la suma pesada de gradientes normalizados por convoluciones de gradientes en determinadas direcciones con varios filtros gaussianos.

Esto mantiene la invarianza de SIFT pero permite realizar los cálculos mucho mas rápido.

# DAISY

Basados en descripción de regiones al igual que SIFT y SURF, es muy eficiente en el cálculo denso.

## En SIFT

Se calcula un gradiente y que el mismo contribuye con su magnitud multiplicada por un peso al bin que corresponda su dirección, el peso dependerá de la distancia que tiene el gradiente a el punto característico.

## DAISY

El histograma será calculado una sola vez por región y reutilizado para todo los píxel cercanos.

Para ello, se reemplaza la suma pesada de gradientes normalizados por convoluciones de gradientes en determinadas direcciones con varios filtros gaussianos.

Esto mantiene la invarianza de SIFT pero permite realizar los cálculos mucho mas rápido.

## Descriptor DAISY

Para una imagen dada, se calcula  $H$  mapas de orientación  $G_i$ , donde  $1 \leq i \leq H$  uno por cada dirección cuantizada.  $G_o(u, v)$  es el gradiente normalizado de la posición  $(u, v)$  en la dirección  $o$  y los resultados negativos se igualan a 0, formalmente entonces

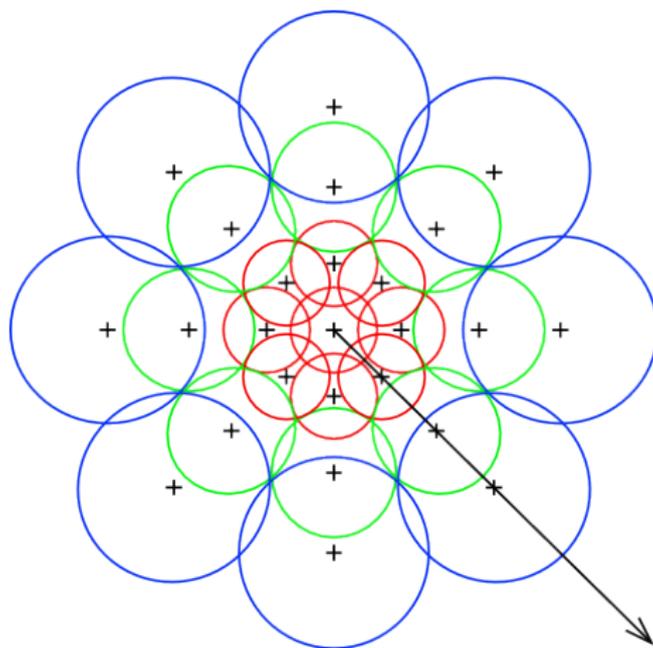
$$G_o = \left( \frac{\partial I}{\partial o} \right)^+ \quad (20)$$

Donde  $I$  es la imagen  $o$  es la orientación de la derivada y el operador  $(.)^+$  es el operador tal que  $(a)^+ = \max(a, 0)$ .

Luego cada mapa de orientación es convolucionado varias veces con filtros gaussianos de diferente  $\sigma$

$$G_o^\sigma = G_o * \left( \frac{\partial I}{\partial o} \right)^+ \quad (21)$$

# Descriptor DAISY



Es descriptor DAISY de cada píxel, es un vector construido con los valores de los mapas de orientación convolucionados ubicados en círculos concéntricos, y donde el valor del kernel Gaussiano es proporcional al radio de los círculos, el nombre del filtro proviene de la figura resultante.

## Descriptor DAISY

El computo de los diferentes mapas se realiza con convoluciones sucesivas de pequeños kernel gaussianos que no requiere mucho costo computacional. En analogía con SIFT, cada píxel resultante de la convolución de un mapa de orientación es muy similares a los bins en SIFT que representaba una suma ponderada de gradientes normalizada y computada sobre una región cercana.  $\mathbf{h}_\sigma(u, v)$  representa el vector de orientaciones de la posición  $(u, v)$  convolucionado por un kernel de desviación estándar  $\sigma$

$$\mathbf{h}_\sigma(u, v) = [G_1^\sigma(u, v), \dots, G_H^\sigma(u, v)]^T \quad (22)$$

# DAISY

Donde  $G_1^\sigma(u, v), \dots, G_H^\sigma(u, v)$  son los  $\sigma$  convolucionados mapas de orientaciones, se normaliza ese vector a la unidad  $\tilde{\mathbf{h}}_\sigma(u, v)$

La normalización es realizada en cada histograma independientemente para ser capaz de representar píxel cerca de oclusiones.

Si  $Q$  representa el numero de diferentes capas circulares, entonces el descriptor DAISY  $\mathcal{D}(u_0, v_0)$  para una ubicación  $(u_0, v_0)$  es definido como:

$$\mathcal{D}(u_0, v_0) = \left[ \begin{array}{l} \tilde{\mathbf{h}}_{\sigma_1}^T(u_0, v_0), \\ \tilde{\mathbf{h}}_{\sigma_1}^T(I_1(u_0, v_0, R_1)), \dots, \tilde{\mathbf{h}}_{\sigma_1}^T(I_T(u_0, v_0, R_1)), \\ \tilde{\mathbf{h}}_{\sigma_2}^T(I_1(u_0, v_0, R_2)), \dots, \tilde{\mathbf{h}}_{\sigma_2}^T(I_T(u_0, v_0, R_2)), \\ \vdots \\ \tilde{\mathbf{h}}_{\sigma_Q}^T(I_1(u_0, v_0, R_Q)), \dots, \tilde{\mathbf{h}}_{\sigma_Q}^T(I_T(u_0, v_0, R_Q)) \end{array} \right]$$

Donde  $I_j(u, v, R)$  es la ubicación de un punto con distancia  $R$  de  $(u, v)$  en la dirección dada por  $j$ .

# DAISY

El computo del descriptor ante rotaciones, no es complejo y no es necesario recomputar las orientaciones, simplemente se realiza un rotación de la grilla donde se realiza la muestra, por otro lado el histograma deberá ser corrido circularmente pero en ningún caso representa costo computacional.

La varianza de la gaussiana es calculada como

$$\sigma_i = \frac{R(i + 1)}{2Q} \quad (23)$$

es decir depende del tamaño del las regiones del descriptor donde  $R$  es la distancia al punto mas lejano ,  $Q$  es la cantidad de capas o círculos e  $i$  es el numero de capa

# Parámetros utilizados

Nombre	Símbolo	Valor
Distancia a el píxel mas alejado	$R$	15
Cantidad de capas	$Q$	3
Cantidad de direcciones	$T$	8
Tamaño de histograma	$H$	8
Puntos de muestras	$S$	$Q \times T + 1 = 25$
Tamaño de descriptor	$\mathcal{D}_s$	$S \times H = 200$

# HOG (Histograms of Oriented Gradients)

El descriptor HOG esta basado en el análisis de contornos mediante el cálculo de gradientes.

La idea subyacente que presenta este descriptor es que la apariencia y forma de un objeto en una imagen puede ser representado por la distribución de la orientaciones de los gradientes

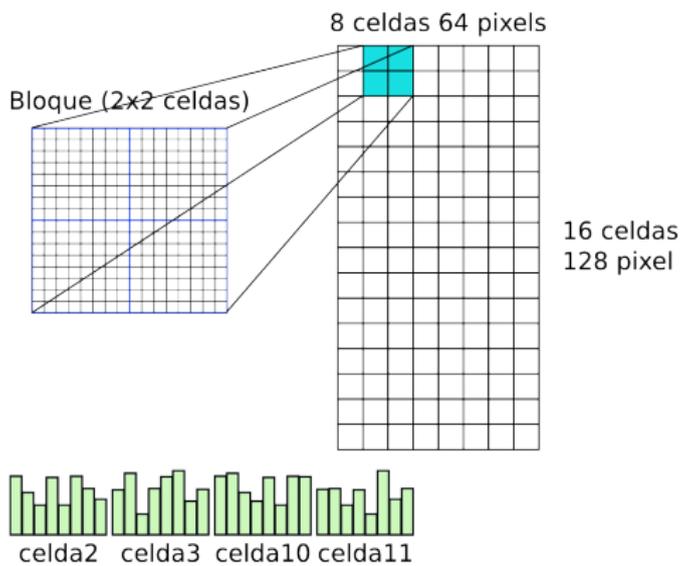
A diferencia de SIFT, HOG describe totalmente el objeto, no describe puntos de interes sino que aplica un descriptor mas pequeño en forma densa a lo largo de toda la imagen.

# Descriptor

Los pasos para obtener el descriptor son:

- 1 Se transforma la imagen a escala de grises.
- 2 Se calculan los gradientes mediante derivadas parciales y luego se calcula dirección y ángulo del gradiente de cada píxel.
- 3 Se divide la imagen en bloques descriptores. Estos están dispuestos de forma tal que cubran toda la imagen a describir con una superposición predefinida entre ellos.
- 4 Se subdivide cada bloque descriptor en celdas.
- 5 Por cada celda se calcula un vector, donde cada elemento del mismo posee la suma de las contribuciones de los gradientes para una dirección dada o bins.
- 6 Una Normalización de los vectores es usualmente requerida debido a variaciones en el contraste e iluminación.
- 7 Los vectores de bins de las celdas que forman un bloque son concatenados entre si, resultando en un vector descriptor del bloque.
- 8 Se concatenan los vectores de cada bloque en un único vector descriptor de la imagen.

# Descriptor



# Descriptor

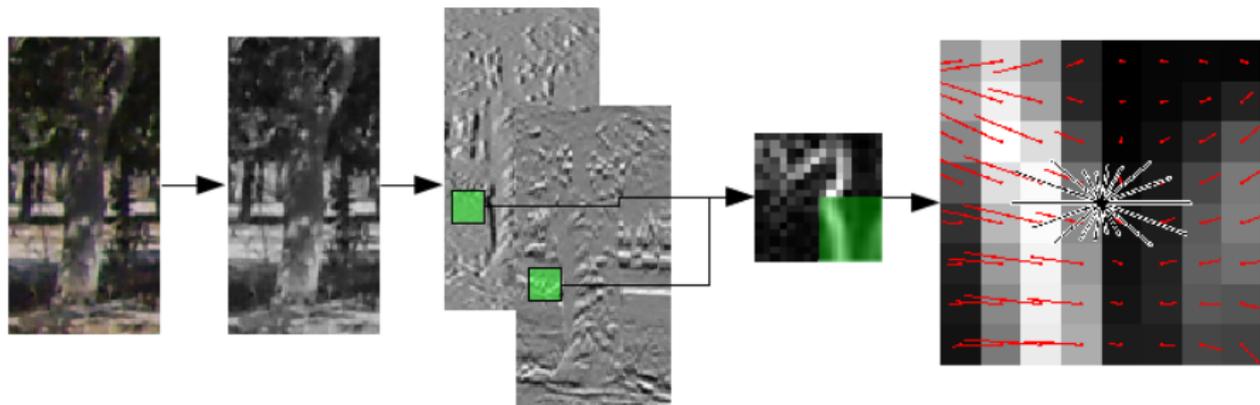


Figura: Pasos para obtener el descriptor.

FIN !!!  
¿ Preguntas ?