Maestría en Análisis y Procesamiento de Imágenes
Visión Robótica

Detectores invariantes locales

Jorge A. Sánchez
jsanchez@scdt.frc.utn.edu.ar

Centro de Investigación en Informática para la Ingeniería

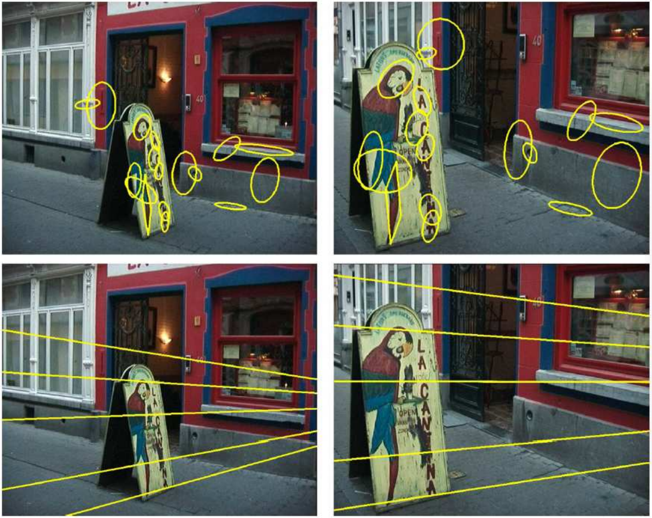
2 de Diciembre de 2011

Motivación: lograr un mayor nivel de robustez frente a cambios en la apariencia visual de los objetos/escena.

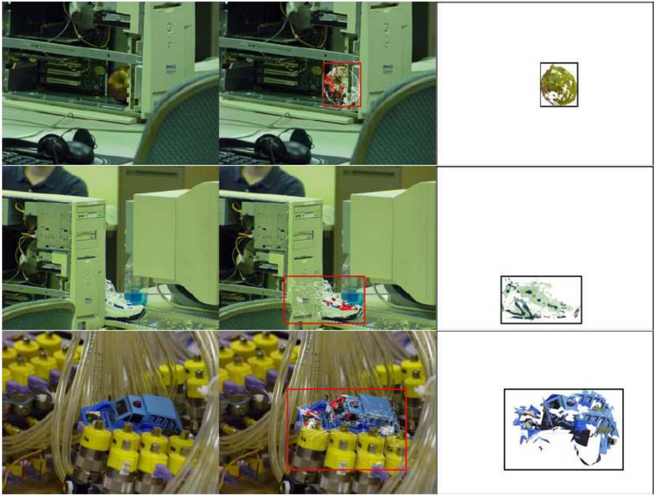
Aplicaciones:

- Estéreo de linea base amplia
- Reconocimiento de instancias particulares de objetos
- Reconstrucción de escenas a partir de fotografías
- Localización basada en apariencias
- ...

Aplicaciones
Estéreo de linea base amplia

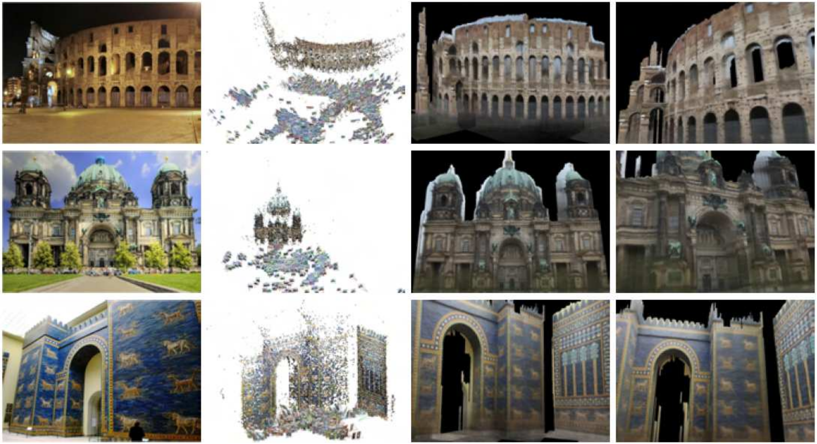


[Tuytelaars and Van Gool, 2004]



[Rothganger et al., 2006]

Aplicaciones
Reconstrucción de escenas a partir de fotografías



[Frahm et al., 2010]

[Cummings and Newman., 2010]

- **Local**, a fin de lidiar con oclusiones parciales y *background clutter*
- **Invariante/covariante** respecto de un conjunto de transformaciones determinado
- **Robusto** frente al ruido, *blur*, discretización, compresión, etc.
- **Preciso** en su localización
- **Eficiente** a fin de permitir operación en tiempo real

- **Invarianza**: la medición no cambia con la transformación

$$T(x) \sim T(Ax) \tag{1}$$

- **Covarianza**: la medición cambia de manera consistente con la transformación

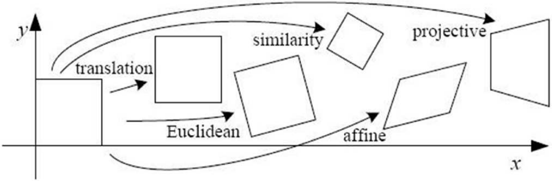
$$T(Ax) \sim AT(x) \tag{2}$$

Ejemplos:

el área de una superficie 2D es <u>invariante</u> ante rotaciones

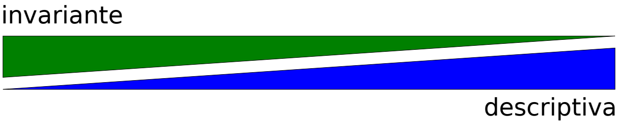
la orientación del eje mayor de un elipse es <u>covariante</u> ante rotaciones (A: rotación, T(·): eje mayor del elipse)

Geométricas:



Fotométricas:

- **invariante**: la misma respuesta aun frente a transformaciones de la imagen
- **discriminativo**: identificar una característica determinada de manera precisa



El nivel de invarianza depende de la aplicación!

- basado en la matriz de segundos momentos (SMM) / tensor de estructura / matriz de autocorrelación

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

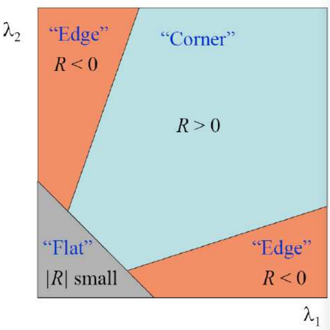
en donde:

$$I_x(\mathbf{x}, \sigma) = \frac{\partial}{\partial x}(g(\sigma) * I) = \left(\frac{\partial}{\partial x} g(\sigma)\right) * I$$

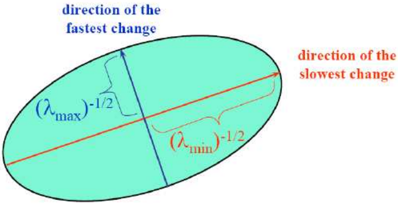
$$g(\sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \ \mathbf{x} = (x, y)^T$$

La medida de Harris:

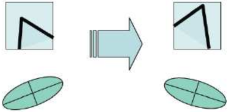
$$R = \det(M) - k(\operatorname{trace}(M))^2 = \lambda_1\lambda_2 - k(\lambda_1 + \lambda_2)^2$$



M caracteriza la distribución del gradiente en un entorno local del punto. La forma cuadrática $\mathbf{x}^T M \mathbf{x}$ se puede asociar a un elipse:

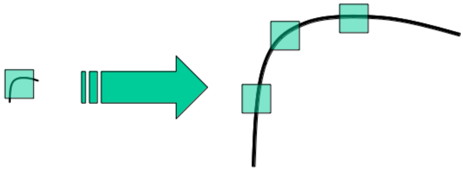


Si bien M cambia con rotaciones, no así sus autovalores $\Rightarrow$ la medida de Harris es invariante ante rotaciones de la imagen



Iluminación: covariante ante transformaciones afines

Escala: esquina o borde de acuerdo a la ventana de observación (escala de integración, σ_I)!



$\Rightarrow$ *sería deseable adaptar la "escala" de los operadores a la estructura local de la imagen.*

La representación $L : \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ de una imagen $f : \mathbb{R}^2 \to \mathbb{R}$ mediante un espacio-escala se obtiene como solución de la ecuación de difusión:

$$\partial_t L = \frac{1}{2}\nabla^2 L$$
$$L(\cdot; 0) = f$$

cuya solución está dada por

$$L(\mathbf{x}; t) = \int f(\mathbf{x} - \eta) g(\mathbf{x}; t) d\eta$$

con $g(\mathbf{x}; t)$ el núcleo Gaussiano de parámetro $\sigma = \sqrt{t}$.

de (izq.,arriba) a (der.,abajo): imagen original, espacio-escala para $t = 1, 8, 64$ respectivamente.

Ejemplo 1D: considérese la función:

$$f(x) = \sin(\omega_0 x)$$

con $w_0 = 2\pi/\lambda_0$. Se puede demostrar que:

$$L(x; t) = e^{-\omega_0^2 t/2} \sin(\omega_o x)$$

La amplitud de la representación:

$$L_{\max}(t) = e^{-\omega_0^2 t/2},$$

así como la de su derivada espacial de orden m:

$$L_{x^m,\max}(t) = \omega_0^m e^{-\omega_0^2 t/2},$$

decrecen exponencialmente con la escala.

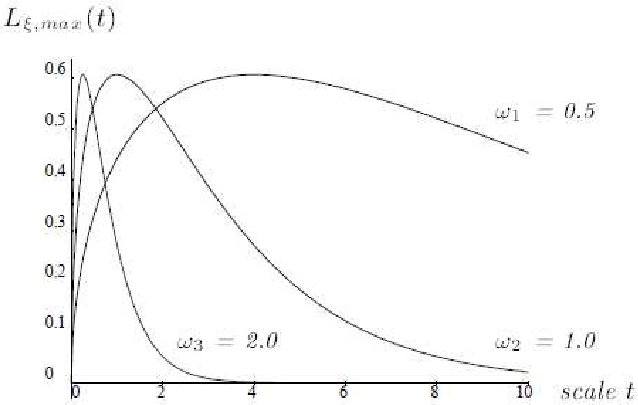
Introdúzcase el concepto de *derivada normalizada*, definida como:

$$\partial_{\xi^m,\mathrm{norm}} = t^{m/2}\partial_{x^m}$$

La amplitud de la derivada *normalizada* de orden m resulta:

$$L_{\xi^m,\max}(t) = t^{m/2}\omega_o^m e^{-\omega_0^2 t/2},$$

que corresponde a una función que <u>crece y luego decrece</u>.


$L_{\xi,max}(t)$
0.6
0.5
0.4
0.3
0.2
0.1
0
0
2
4
6
8
10
scale t
$\omega_1 = 0.5$
$\omega_2 = 1.0$
$\omega_3 = 2.0$

$L_{\xi^m,\max}$ alcanza un máximo (en este caso único) en:

$$t_{max} = \frac{m}{\omega_0^2}$$

o, en forma equivalente, en

$$\sigma_{max} = t_{max} = \frac{\sqrt{m}}{2\pi}\lambda_0$$

$\Rightarrow$ *la ubicación del máximo es proporcional a la "escala" de la función.*

En t_{max}, el valor de $L_{\xi^m,\max}$ resulta

$$L_{\xi^m,\max}(t_{max}) = m^{m/2}e^{-m/2}$$

$\Rightarrow$ *la magnitud del máximo es independiente de ω_o.*

Dada una imagen y su representación mediante un espacio-escala, la idea de la "selección automática de escalas" para un punto dado $x \in \mathbb{R}^2$, consiste en asignar al mismo una *escala característica* que en algún sentido refleje la extensión espacial de la estructura a la que pertenece.

Principio de selección automática de escalas [Lindeberg, 1998]: *se asume como "escala característica" la ubicación en escalas del máximo local de una combinación (posiblemente no lineal) de derivadas normalizadas.*

Seleccción automática de escalas

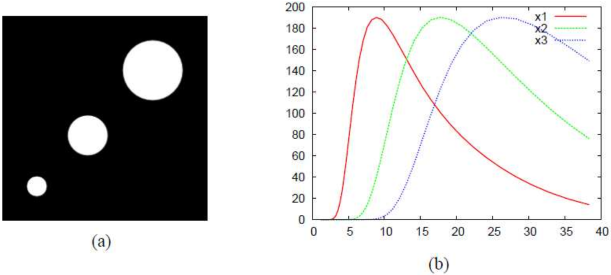


Figura 1: Evolución en la magnitud del Laplaciano γ-normalizado con la escala para estructuras (blobs) de diferente tamaño, computada en sus centros respectivos, x_i (i=1,2,3).

Invarianza rotacional + escala $\Rightarrow$ operadores diferenciales normalizados rotacionalmente simétricos

Laplaciano normalizado (LoG) (blobs)

$$\nabla^2_{norm} L = t(L_{xx} + L_{yy})$$

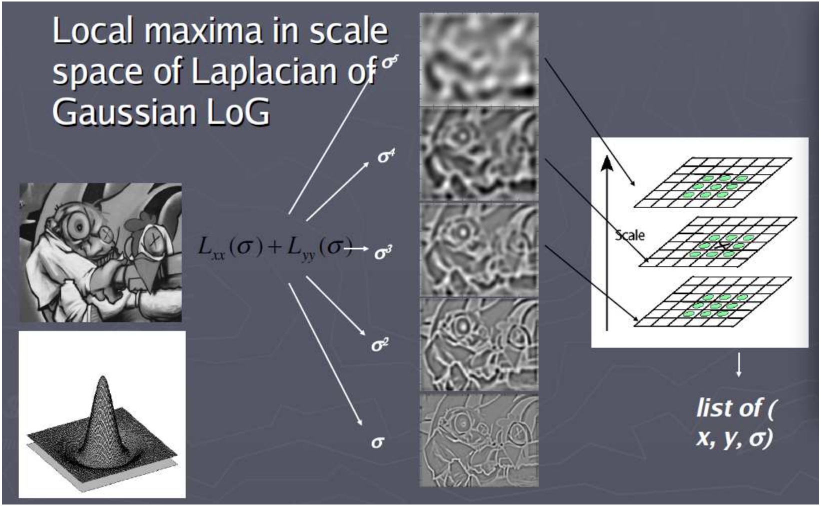
Determinante del Hessiano (blobs, puntos de ensilladura)

$$det(\mathcal{H}_{norm} L) = t^2 \left(L_{xx} L_{yy} - L_{xy}^2\right)$$

Diferencia de Gaussianas (DoG) (blobs, aproximación a LoG)

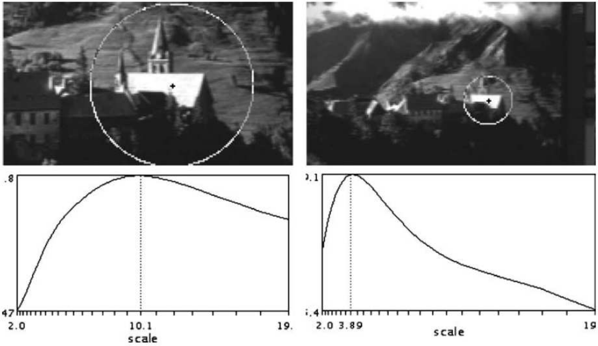
$$\nabla^2_{norm} L(\cdot; t_n) \approx L(\cdot; t_n) - L(\cdot; t_{n-1}) \quad (t_n = s t_{n-1})$$

Detección: *máximos locales en la respuesta espacial y en escala.*



T. Tuytetaars, ECCV'06 Local Invariant Features Tutorial

Seleccción automática de escalas



SMM-normalizada

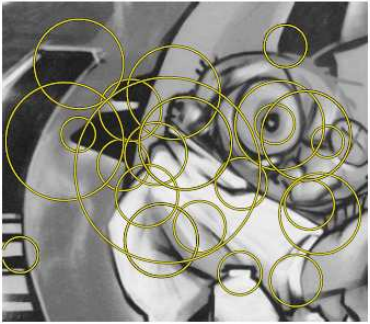
$$M_{norm} = t_D g(t_I) * \begin{pmatrix} L_x^2(\mathbf{x}, t_D) & L_x(\mathbf{x}, t_D)L_y(\mathbf{x}, t_D) \\ L_x(\mathbf{x}, t_D)L_y(\mathbf{x}, t_D) & L_y^2(\mathbf{x}, t_D) \end{pmatrix}$$

t_D: escala de derivación

t_I: escala de integración. Por lo general $t_D = st_I$, $s \leq 1$

Detector Harris-Laplace:

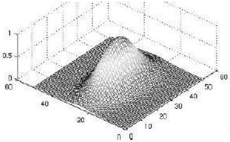
- Selección de t_I utilizando LoG
- Máximos locales (espaciales) en $R_{norm} = \det(M_{norm}) - k(\mathrm{trace}(M_{norm}))^2$
- (opcional) Refinamiento iterativo



⇒ *es posible contemplar cambios en la estructura más allá de la escala?.*

Espacio-escala afín: resulta de relajar la isotropía del núcleo Gaussiano

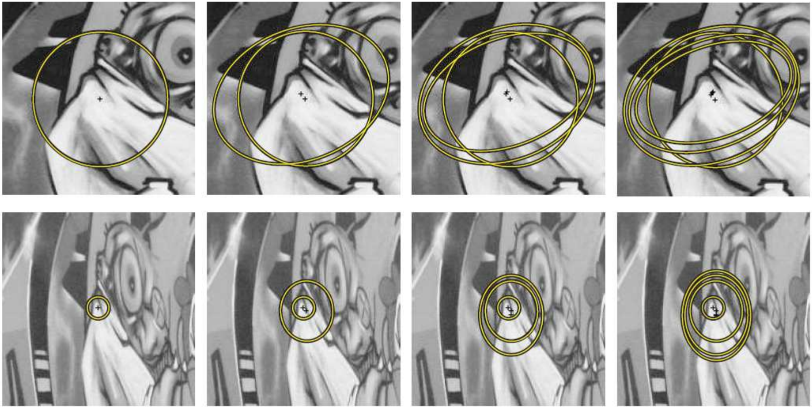
$$g(\cdot;\Sigma) = \frac{1}{2\pi\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}\mathbf{x}^{x}\Sigma^{-1}\mathbf{x}\right) \qquad (3)$$



SMM afín: $M_{afin} = \det(\Sigma_D)g(\Sigma_I) * \left((\nabla L(\cdot;\Sigma_D))(\nabla L(\cdot;\Sigma_D))^T\right)$

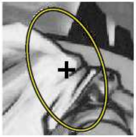
Adaptación afín: utilizar M_{afin} para estimar la estructura afín asociada a una región de la imagen [Lindeberg and Gaerding, 1997]

Adaptación afín

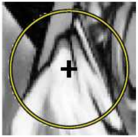


$\Rightarrow$ *transformación estimada hasta una rotación arbitraria*

Adaptación afín



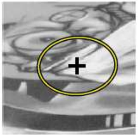
$$\mathbf{x}_L \longrightarrow M_L^{-1/2}\mathbf{x}'_L$$



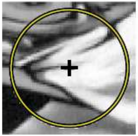
$$\downarrow$$

$$\mathbf{x}'_L \longrightarrow R\mathbf{x}'_R$$

$$\downarrow$$



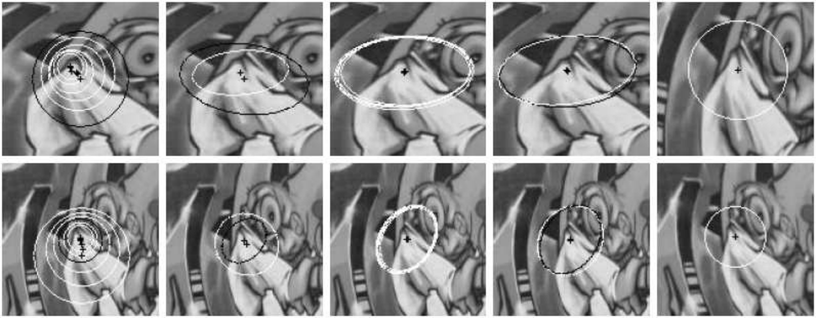
$$\mathbf{x}_R \longrightarrow M_R^{-1/2}\mathbf{x}'_R$$



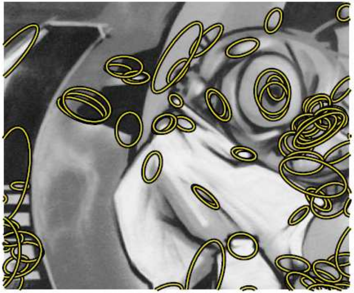
$\Rightarrow$ *Solución: asignar una orientación basada en alguna propiedad local (p.ej. distribución de la orientación del gradiente espacial)*

Detector Harris-affine:

- Detección de puntos de Harris multiescala
- Selección de escalas usando LoG
- Adaptación afín usando M_{afin}
- Refinar la ubicación espacial del punto

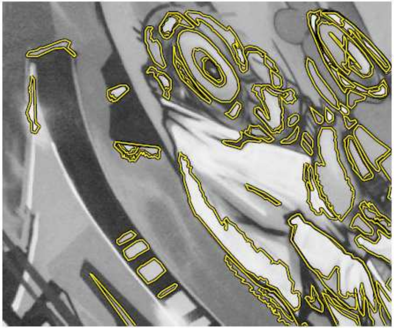


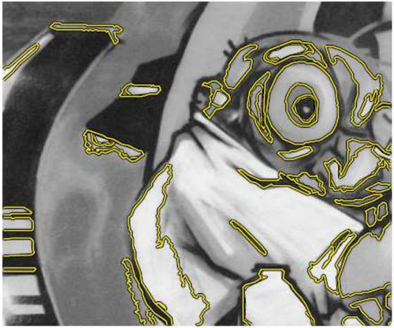
Detector Harris-affine



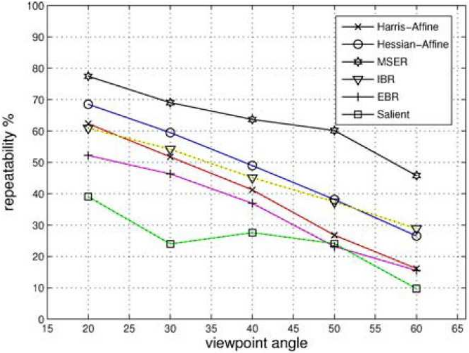
Maximally Stable Extrema Regions (MSER)

- Basados en algoritmo de *watershed*
- Regiones estables al aplicar umbrales graduales



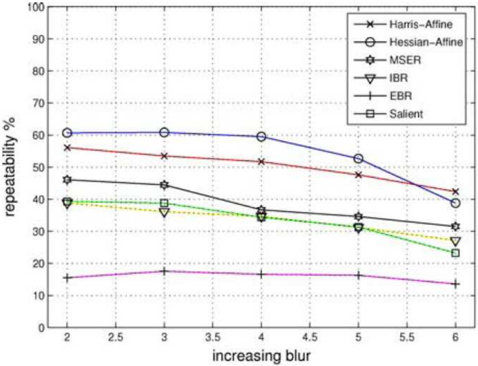


Criterio: repetibilidad (puntos de vista, zoom+rotación, blur, compresión JPEG, iluminación) [Mikolajczyk et al. 2005]

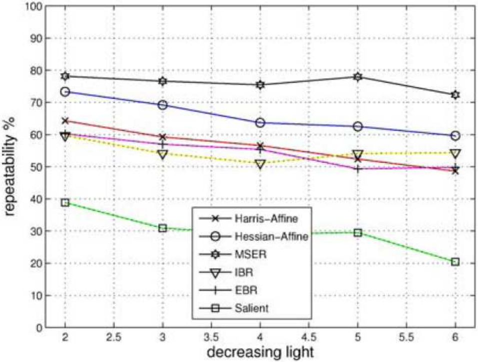


Comparación

Criterio: repetibilidad (puntos de vista, zoom+rotación, blur, compresión JPEG, iluminación) [Mikolajczyk et al. 2005]

Criterio: repetibilidad (puntos de vista, zoom+rotación, blur, compresión JPEG, iluminación) [Mikolajczyk et al. 2005]



Conclusiones:

- Complementariedad entre detectores
- El rendimiento depende del tipo de escena
- Mayor nivel de invarianza $\Rightarrow$ mayor complejidad computacional
- Muchos (realmente muchos) detectores en la literatura

Otras estrategias:

- Pirámides de resolución
- Muestro aleatorio
- Muestro regular