Computer Vision

MSc Informatics option GVR

James L. Crowley

Fall Semester 4 November 2010

Lesson 5

Describing Local Appearance with Gaussian Derivatives in Scale Space

Lesson Outline:

1 Describing Local Appearance

Appearance is what you see.

We seek a set of K local basis functions, $f_k(x,y)$ to describe "appearance" around a point in an image. Each basis function, $f_k(x,y)$ gives an image "feature", x_k, describing appearance at the pixel (x_o, y_o).

$$x_k = \sum_{x=-R}^{R} \sum_{y=-R}^{R} p(x_o + x, y_o + y) f_k(x,y)$$

Projection of the image neighborhood $P(x_o,y_o)$ onto this set of functions gives a "feature" vector for appearance, $\vec{X} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_K \end{pmatrix}$

Ideally these functions should be orthogonal

$$\sum_{x=-R}^{R} \sum_{y=-R}^{R} f_m(x_o + x, y_o + y) f_n(x,y) = 0 \quad \text{if n} \neq \text{m}$$

to minimize the number of features. This is not strictly necessary.
Such a basis can be obtained by a Taylor series representation of the image function P(x) around a point, a:

$$f(x) = f(a) + \frac{1}{1!} f_x (x-a) + \frac{1}{2!} f_{xx} (x-a)^2 + \frac{1}{3!} f_{xxx} (x-a)^3 + \dots$$

The basis set for a taylor series is the series of local derivatives.
This is called the "local jet". The local jet is a basis for describing appearance.

Because the image is a sampled signal, we need a kernel filter to compute the local Jet. The Gaussian function provides a Kernel with many useful properties.

2 Scale and Rotation Invariant Image Description

The Sampled Gaussian Functions

2D Gaussian Receptive Field: $G(i,j,\sigma) = \frac{1}{A} W_R(i,j) \cdot e^{-\frac{(i^2+j^2)}{2\sigma^2}}$

where

$$w_R(i,j) = \begin{cases} 1 & \text{for } -R \le i \le R \text{ and } -R \le j \le R \\ 0 & \text{otherwise} \end{cases}$$

Finite windw, $w_R(i,j)$ has $N^2 = (2R+1)^2$ coefficients

Typically: for R should be $\geq 3\sigma$. Recommend $R=4\sigma$

(Attention: i, j are INTEGERS, not complex numbers!)

The normalization factor $A = \sum_{i=-R}^{R} \sum_{j=-R}^{R} e^{-\frac{(i^2+j^2)}{2\sigma^2}} \approx 2\pi\sigma$

Using the Gaussian to compute image derivatives

For an image P(i, j), the derivatives can be approximated by convolution with Derivatives of Gaussians

$$P_x(i,j) \approx P * G_x(i,j,\sigma)$$
$$P_y(i,j) \approx P * G_y(i,j,\sigma)$$
$$P_{xx}(i,j) \approx P * G_{xx}(i,j,\sigma)$$
$$P_{xy}(i,j) \approx P * G_{xy}(i,j,\sigma)$$
$$P_{yy}(i,j) \approx P * G_{yy}(i,j,\sigma)$$

Note: it is NECESSARY to specify σ. Small σ is not necessarily best.

The Gradient $\vec{\nabla}P(i,j)$ is calculated by $P(i,j) * \vec{\nabla}G(i,j,\sigma)$

where $\vec{\nabla}G(i,j,\sigma) = \begin{pmatrix} G_x(i,j,\sigma) \\ G_y(i,j,\sigma) \end{pmatrix}$

Gradient: $\vec{\nabla}P(i,j) = \begin{pmatrix} P_x(i,j) \\ P_y(i,j) \end{pmatrix} \approx \vec{\nabla}(P * G(i,j,\sigma)) = P * \vec{\nabla}G(i,j,\sigma) = \begin{pmatrix} P * G_x(i,j,\sigma) \\ P * G_y(i,j,\sigma) \end{pmatrix}$

Laplacien: $\nabla^2 P(i,j) = P * \nabla^2 G(i,j,\sigma) = P_{xx}(i,j) + P_{yy}(i,j) \approx P * G_{xx}(i,j,\sigma) + P * G_{yy}(i,j,\sigma)$

Gaussian Derivatives are Steerable:

$$G_1^{\theta}(x,y,\sigma) = \cos(\theta) \cdot G_x(x,y,\sigma) + \sin(\theta) \cdot G_y(x,y,\sigma)$$

Thus:

1st order $P_1^{\theta}(i,j) = Cos(\theta) P_x(i,j) + Sin(\theta) P_y(i,j)$

2nd order $P_1^{\theta}(i,j) = Cos(\theta)^2 P_{xx}(i,j) + Sin(\theta)^2 P_{yy}(i,j) + 2Cos(\theta)Sin(\theta) P_{xy}(i,j)$

3rd order

$$P_3^{\theta}(i,j) = Cos(\theta)^3 P_{xxx}(i,j) + Cos(\theta)^2 Sin(\theta) P_{xxy}(i,j) + Cos(\theta)Sin(\theta)^2 P_{xyy}(i,j) + Sin(\theta)^3 P_{yyy}(i,j)$$

By steering the derivatives to the local orientation, we obtain an "invariant" measure of local contrast. We can also "steer" in scale to obtain invariance to size.

Note, we can NOT steer the mixed derivatives, i.e $P_{xy}(i,j)$

3 Image Scale Space

Continuous Case.

Let P(x,y) be the image.
Let G(x, y, σ) by a Gaussian function of scale $\sigma=2^s$

Image Scale space is a 3D continuous space P(x,y,s)

$P(x, y, s) = P(x,y)* G(x, y,\sigma)$

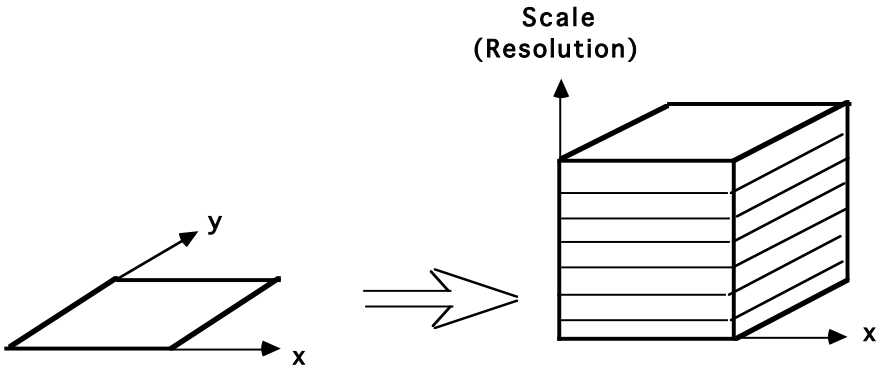
Note that the scale axis (s) is logarithmic. $s = Log_2(\sigma) = Log_2(2^s)$

Scale invariance: If a shape in an image is made larger by $D = 2^d$

$p(x,y) \rightarrow p(x2^d, y2^d)$

Then the scale space projection of appearance is shifted by s

$P(x,y,s+d) = p(x2^d, y2^d) * G(x2^d, y2^d, 2^d)$



The appearance of a pattern in the image results in a unique structure in P(x, y, s).
This structure is "equivariant" in position, scale and rotation.
Translate the pattern by Δx, Δy and the structure translates by Δx, Δy in P(x, y, s).

Rotate by θ in x, y and the structure rotates by θ in P(x, y, s).

Scale by a factor of 2^s, and the structure translates by s in P(x, y, s).

Scale space :
Separates global structure from fine detail.
Provides context for recognition.
Provides a description that is invariant to position, orientation and scale.

Discrete Scale Space - Scale invariant impulse response.

In a computer, we need to discretize (sample) x, y, and s.

Let P(i,j) be a discrete representation for $P(x,y) = P(i\Delta x, j\Delta x)$
Suppose P(i,j) is an image array of size M x M pixels.

We propose to sample scale with a step size of $\Delta\sigma = 2^{1/2}$ so that $\sigma_k = 2^{k/2}$

Note that scale space "dilates" the Gaussian impulse response by 2^s.

$$P(x,y,s) = P(x,y) * G(x, y, 2^s)$$

As the Gaussian impulse response dilates, the sample density can also dilate.

$$P(x,y,s) = p(i\,\Delta x_k, j\,\Delta x_k, k) \quad \text{such that } \Delta x_k = 2^{k/2}$$

For a Gaussian Kernel filter $G(i,j,k) = G(x, y, \sigma_k = 2^{k/2})$

The image pyramid becomes :

$$P(i,j,k) = p(i\,2^{k/2}, j\,2^{k/2}, 2^{k/2}) = P(i,j) * G(i\,2^{k/2}, j\,2^{k/2}, 2^{k/2})$$

for $1 \le k \le M-4$

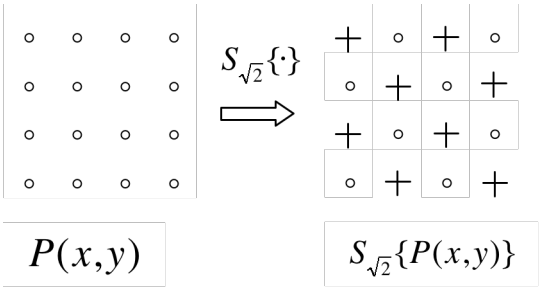
$(M = \text{Log}_2(\max\{\text{width, height}\}))$

Diagonal, Square root of two Sampling

Problem : How can we sample an image for for odd k? $\Delta x = 2^{k/2} = 2^{(k-1)/2}\sqrt{2}$

for k odd, $\Delta x_k = \{1, 2, 4, 8\ldots\}$
for k even, $\Delta x_k = \{\sqrt{2}, 2\sqrt{2}, 4\sqrt{2}, 8\sqrt{2}, \ldots\}$

How ? with the diagonal sampling operator $S_{\sqrt{2}}\{\}$

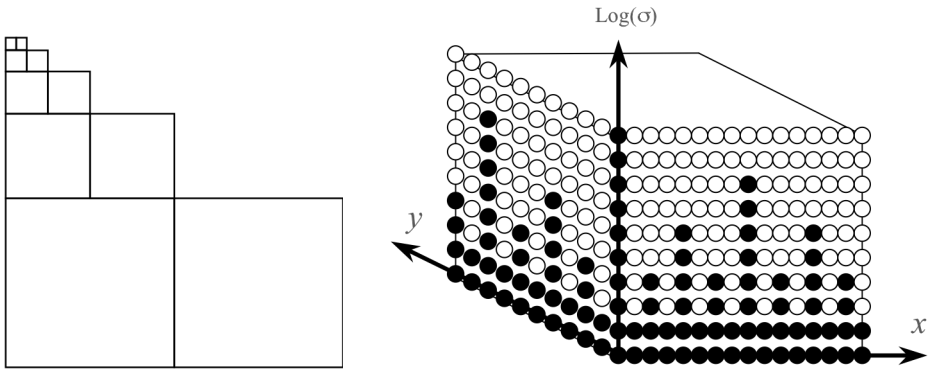


For k even, the √2 resampling operator, $S_{\sqrt{2}}{}^{k}\{\}$, selects even columns of even rows and odd columns of odd rows.

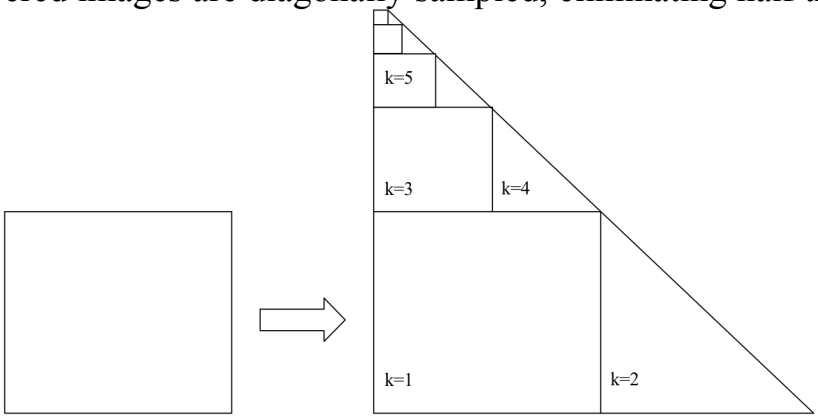
For k odd, diagonal sample operator eliminates every second column (starting with even columns on even rows and odd columns on odd rows). For k odd, resampling eliminates every second row (odd rows).

$$S_{\sqrt{2}^k}\{P\{x,y)\} = \begin{cases} P(x,y) & \text{if } (x+y)^2 \text{ Mod } 2^{k-1} = 0 \\ 0 & \text{otherwise} \end{cases}$$

Data Structure



The even numbered images are diagonally sampled, eliminating half the pixels.



For an image of size MxM, number of pixels is

$$P = MxM \times (1 + ½ + ¼ + \ldots) = 2M^2$$

4 Scale Invariant Pyramid Algorithm

Cost of computing p(i,j,k) is

$$C= O(M^2((N_0+1)^2+(N_1+1)^2+(N_2+1)^2+\ldots+(N_{M-4}+1)^2))$$

if we use "seperable" convolution:

$$P(i,j) * G(i,j, 2^{k/2}) = P(i,j) * G(i, 2^{k/2}) * G(j, 2^{k/2})$$

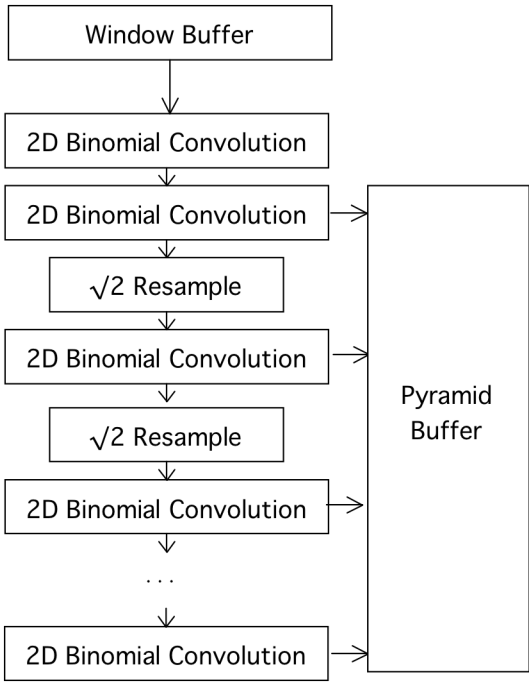
then

$$C= O(M^2 \cdot 2(N_0+N_1+N_2+N_3+\ldots+N_{M-4}+M-4+1)$$

$$C= O(M^2 \cdot 2(8+16+32+64+\ldots+ N_{M-4})+6).$$

Practically, the computational cost is exorbitant.

We can use Cascade Convolution Methods to reduce cost to O(N)



Each binomial convolution provides convolution with G(x,y,1).

Cumulative Variance is 1,2,4,8,16,32,64,…

Cumulative standard deviation (σ) is $1, \sqrt{2}, 2, 2\sqrt{2}, 4, 4\sqrt{2}, 8, 8\sqrt{2}, 16, \ldots$

Using the Pyramid to compute image derivatives

For an image P(i, j), the derivatives can be approximated by convolution with Derivatives of Gaussians

$$P_x(i,j) \approx P * G_x(i,j,\sigma)$$
$$P_y(i,j) \approx P * G_y(i,j,\sigma)$$
$$P_{xx}(i,j) \approx P * G_{xx}(i,j,\sigma)$$
$$P_{xy}(i,j) \approx P * G_{xy}(i,j,\sigma)$$
$$P_{yy}(i,j) \approx P * G_{yy}(i,j,\sigma)$$

Note: it is NECESSARY to specify σ. Small σ is not necessarily best.

With the Pyramid, derivatives can be obtained directly by sum and difference.

$$P_x(i,j,k) =< P(i,j), G_x(i,j,2^{k/2}) \approx P(i+1,j,k) - P(i-1,j,k)$$
$$P_y(i,j,k) =< P(i,j), G_x(i,j,2^{k/2}) \approx P(i,j+1,k) - P(i,j-1,k)$$
$$P_{xx}(i,j,k) =< P(i,j), G_{xx}(i,j,2^{k/2}) \approx P(i+1,j,k) - 2P(i,j,k) + P(i-1,j,k)$$
$$P_{yy}(i,j,k) =< P(i,j), G_{yy}(i,j,2^{k/2}) \approx P(i,j+1,k) - 2P(i,j,k) + P(i,j-1,k)$$

$$P_{xy}(i,j,k) =< P(i,j), G_{xy}(i,j,2^{k/2}) >$$
$$\approx P(i+1,j+1,k) - P(i-1,j+1,k) - P(i+1,j-1,k) + P(i-1,j-1,k)$$

The Gradient $\vec{\nabla}P(i,j)$ is calculated by $P(i,j) * \vec{\nabla}G(i,j,\sigma)$

where $\vec{\nabla}G(i,j,\sigma) = \begin{pmatrix} G_x(i,j,\sigma) \\ G_y(i,j,\sigma) \end{pmatrix}$

Gradient: $\vec{\nabla}P(i,j) = \begin{pmatrix} P_x(i,j) \\ P_y(i,j) \end{pmatrix} \approx \vec{\nabla}(P * G(i,j,\sigma)) = P * \vec{\nabla}G(i,j,\sigma) = \begin{pmatrix} P * G_x(i,j,\sigma) \\ P * G_y(i,j,\sigma) \end{pmatrix}$

Laplacien: $\nabla^2 P(i,j) = P * \nabla^2 G(i,j,\sigma) = P_{xx}(i,j) + P_{yy}(i,j) \approx P * G_{xx}(i,j,\sigma) + P * G_{yy}(i,j,\sigma)$

For a pyramid

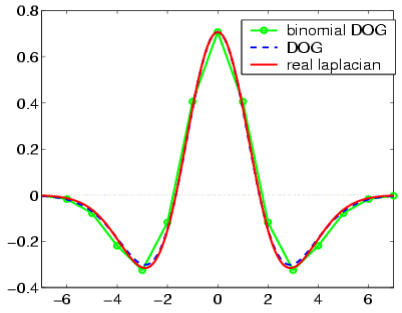
Diffusion Equation: $\nabla^2 G_x(i,j,\sigma) = G_{xx}(i,j,\sigma) + G_{yy}(i,j,\sigma) = \frac{\partial G(i,j,\sigma)}{\partial \sigma}$

As a consequence: $\nabla^2 G(i,j,\sigma) \approx G(i,j,\sigma_1) - G(i,j,\sigma_2)$

This typically requires $\sigma_1 \geq \sqrt{2}\ \sigma_2$

Thus it is common to use:

$$\nabla^2 P(i,j,k) = <p(i,j), \nabla^2 G(i,j,\sigma_k)> \approx P(i,j,k) - P(i,j,k-1)$$



Oriented Derivatives

Gaussian Derivatives are Steerable:

$$G_1^{\theta}(x,y,\sigma) = \cos(\theta)\cdot G_x(x,y,\sigma) + \sin(\theta)\cdot G_y(x,y,\sigma)$$

Thus in the pyramid

1st order $P_1^{\theta}(i,j,k) = Cos(\theta)P_x(i,j,k) + Sin(\theta)P_y(i,j,k)$

2nd order $P_1^{\theta}(i,j,k) = Cos(\theta)^2 P_{xx}(i,j,k) + Sin(\theta)^2 P_{yy}(i,j,k) + 2Cos(\theta)Sin(\theta)P_{xy}(i,j,k)$

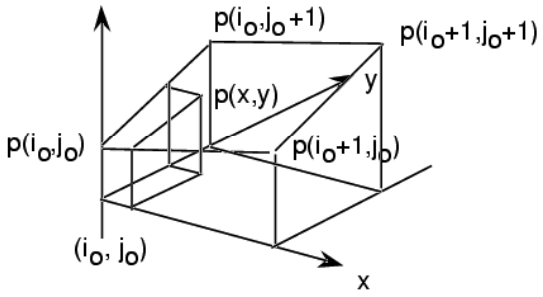
3rd order

$$P_3^{\theta}(i,j,k) = Cos(\theta)^3 P_{xxx}(i,j,k) + Cos(\theta)^2 Sin(\theta)P_{xxy}(i,j,k) + Cos(\theta)Sin(\theta)^2 P_{xyy}(i,j,k) + Sin(\theta)^3 P_{yyy}(i,j,k)$$

By steering the derivatives to the local orientation, we obtain an "invariant" measure of local contrast. We can also "steer" in scale to obtain invariance to size.

Note, we can NOT steer the mixed derivatives, i.e $P_{xy}(i,j,k)$

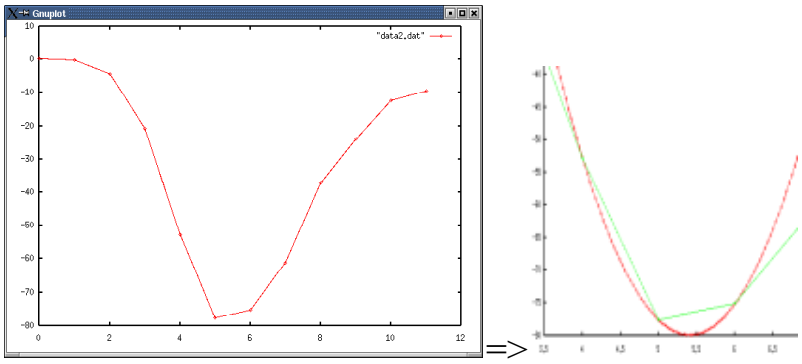
Interpolation



$$a = p(i_0+1, j_0) - p(i_0, j_0)$$
$$b = p(i_0, j_0+1) - p(i_0, j_0)$$
$$c = a + p(i_0, j_0+1) - p(i_0+1, j_0+1)$$

$$p(x,y) = a(x-i_0) + b(y-j_0) + c(x-i_0)(y-j_0) + p(i_0, j_0)$$

Scale Inerpolation

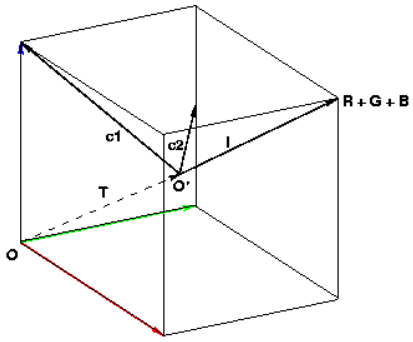


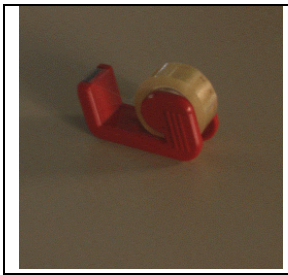
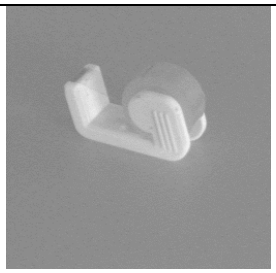
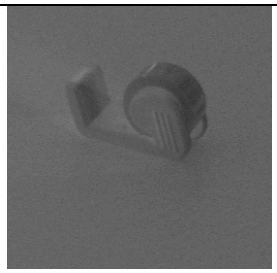
$$x_{\max} = x_2 + \frac{f(x_1) - f(x_3)}{2(f(x_1) + f(x_3) - 2f(x_2))}$$

Color Opponent Scale Space

$$(R, G, B) \Rightarrow (L, C_1, C_2) \qquad \begin{pmatrix} L \\ C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} 0.33 & 0.33 & 0.33 \\ -0.5 & -0.5 & 1 \\ 0.5 & -0.5 & 0 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

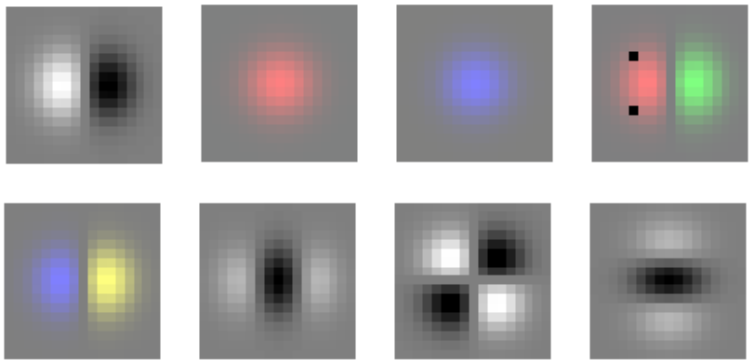
This representation separates luminance and chrominance.



			
RGB	B-W	R-G	R+G-B

This makes it possible to "steer" the chrominance to an illumination color

$$\begin{pmatrix} L \\ C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} 0.33 & 0.33 & 0.33 \\ -0.5 & -0.5 & 1 \\ 0.5 & -0.5 & 0 \end{pmatrix} \begin{pmatrix} \alpha_1 R \\ \alpha_2 G \\ \alpha_3 B \end{pmatrix}$$



We then compute 3 pyramids : L(i,j,k), C_1(i,j,k), and C_2(i,j,k),

This gives us a feature vector for appearance:

$$\vec{\varphi} = \begin{bmatrix} \varphi_0 \\ \vdots \\ \varphi_k \end{bmatrix} = \begin{bmatrix} G_x^L \\ G^{C_1} \\ G^{C_2} \\ G_x^{C_1} \\ G_x^{C_2} \\ G_{xx}^L \\ G_{xy}^L \\ G_{yy}^L \end{bmatrix}$$

5 Scale Invariant Interest Points

Maximal points in the image derivatives provide keypoints.
In an image scale space, these points are scale invariant.

Example: maxima in the lapacian as invariant "interest points"

Recall the Laplacian of the image :

$$\nabla^2 P(x,y,s) = P * \nabla^2 G(x,y,\sigma) = P * G_{xx}(x,y,\sigma) + P * G_{yy}(x,y,\sigma) \approx P * \nabla^2 G(x,y,\sigma_1) - P * \nabla^2 G(x,y,\sigma_2)$$

Scale invariant interest points are given by

$$X(x,y,s) = \underset{x,y,s}{local-\max}\{\nabla^2 P(x,y,s)\}$$

Where Local-Max detect points that are larger then all neighboring points within a radius of size R.

Because $\nabla^2 P(i,j,k) = \langle P(i,j), \nabla^2 G(i,j,\sigma_k) \rangle \approx P(i,j,k) - P(i,j,k-1)$

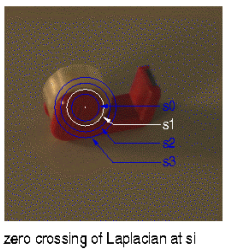
This is referred to as a "Difference of Gaussian" or DoG detector.

We can detect scale invariant interest points as

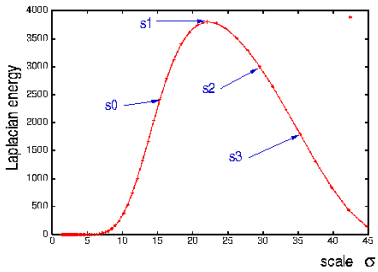
$$X(i,j,k)_n = \underset{i,j,k}{Local-\max}\{\nabla^2 P(i,j,k)\}$$

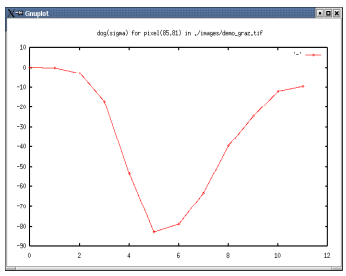
with R=2.

Examples:



zero crossing of Laplacian at s1





Maximally stable scale invariant points are found as :

$$X(i,j,k) = \underset{i,j,k}{Local-\max}\{P(i,j,k) - P(i,j,k-1)\}$$

Such points are used for tracking, for image registration, and as feature points for recognition.

The scale of the maximal Laplacian is an invariant at ALL image points.

The scale σ_i is an "invariant" for the appearance at P(i,j).

$$\sigma_i = \arg\underset{\sigma}{-\max}\{P * \nabla^2 G(i,j,\sigma)\}$$
$$\sigma_i = \arg\underset{\sigma}{-\max}\{\nabla^2_{\sigma=2^k} P(i,j)\}$$
$$\sigma_i = \arg\underset{\sigma}{-\max}\{P(i,j,k) - P(i,j,k-1)\}$$

Other popular detectors for scale invariant interest points include:

Gradient Magnitude: $X(i,j,k) = \underset{i,j,k}{Local-\max}\{\| P_x(i,j,k), P_y(i,j,k) \|\}$

and Determinant of the Hessian: $X(i,j,k) = \underset{i,j,k}{Local-\max}\left\{\det\begin{pmatrix} P_{xx}(i,j,k) & P_{xy}(i,j,k) \\ P_{xy}(i,j,k) & P_{yy}(i,j,k) \end{pmatrix}\right\}$

$$X(i,j,k) = \underset{i,j,k}{Local-\max}\left\{P_{xx}(i,j,k)P_{yy}(i,j,k) - P_{xy}(i,j,k)^2\right\}$$