Harris Interest Operator

(lecture notes)

Martin Urban

January 28, 2003

1 Notations

x, y - pixel coordinates
$I(x, y)$ - pixel value (intensity)
$\nabla I(x, y)$ - image gradient, $\nabla I(x, y) = \left(\frac{\partial I(x,y)}{\partial x}, \frac{\partial I(x,y)}{\partial y} \right)^\top$
$N(x_0, y_0, r)$ - neighbourhood of a pixel (x_0, y_0)

$$N(x_0, y_0, r) = \left\{ (x, y) \in \mathbb{R}^2 \mid \| (x, y) - (x_0, y_0) \| \leq r \right\}$$

$(\Delta x, \Delta y)$ - elementary shift, $(\Delta x, \Delta y) \in \mathbb{R}^2 - \{(0, 0)\}$

D_ϵ - set of all shifts of size ϵ, $\epsilon > 0$

$$D_\epsilon = \left\{ (\Delta x_0, \Delta y_0) \in \mathbb{R}^2 \mid \| (\Delta x_0, \Delta y_0) \| = \epsilon, \ \epsilon > 0 \right\}$$

2 Basic Principle

Assume we have an image patch N_0, $N_0 = N(x_0, y_0, r)$. We want to express the minimum of square pixel difference between an image patch N_0 and a shifted image patch $N(x_0 + \Delta x, y_0 + \Delta y, r)$ over all directions $(\Delta x, \Delta y) \in D_\epsilon$:

$$f(N_0) = \min_{(\Delta x, \Delta y) \in D_\epsilon} \sum_{(x,y) \in N_0} \left(I(x, y) - I(x + \Delta x, y + \Delta y) \right)^2 \qquad (1)$$

1. Let us approximate $I(x + \Delta x, y + \Delta y)$ by the first two terms of Taylor expansion :

$$I(x + \Delta x, y + \Delta y) \approx I(x, y) + \nabla I(x, y)^\top \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \qquad (2)$$

2. Substituting the approximation into (1) we receive:

$$
\begin{aligned}
f\left(N_0\right) &\approx \min_{(\Delta x, \Delta y) \in D_\epsilon} \sum_{(x,y)\in N_0} \left[\ \Delta x, \Delta y\ \right] \nabla I\left(x,y\right) \nabla I\left(x,y\right)^\top \left[\begin{array}{c} \Delta x \\ \Delta y \end{array}\right] \\
&= \min_{(\Delta x, \Delta y) \in D_\epsilon} \left[\ \Delta x, \Delta y\ \right] \left(\sum_{(x,y)\in N_0} \nabla I\left(x,y\right) \nabla I\left(x,y\right)^\top \right) \left[\begin{array}{c} \Delta x \\ \Delta y \end{array}\right] \\
&= \min_{(\Delta x, \Delta y) \in D_\epsilon} \left[\ \Delta x, \Delta y\ \right] \mathbf{A} \left[\begin{array}{c} \Delta x \\ \Delta y \end{array}\right] ,
\end{aligned}
$$

where $\mathbf{A}$ is a symmetric positive semi-definite 2×2 matrix:

$$
\begin{aligned}
\mathbf{A} &= \sum_{(x,y)\in N_0} \nabla I\left(x,y\right) \nabla I\left(x,y\right)^\top && (3) \\
&= \sum_{(x,y)\in N_0} \left[\ \frac{\partial\, I(x,y)}{\partial\, x}, \frac{\partial\, I(x,y)}{\partial\, y}\ \right] \left[\begin{array}{c} \frac{\partial\, I(x,y)}{\partial\, x} \\ \frac{\partial\, I(x,y)}{\partial\, y} \end{array}\right] && (4)
\end{aligned}
$$

3. Since $\mathbf{A}$ is a symmetric matrix

$$
\min_{(\Delta x, \Delta y) \in D_\epsilon} \left[\ \Delta x, \Delta y\ \right] \mathbf{A} \left[\begin{array}{c} \Delta x \\ \Delta y \end{array}\right] = \|\left(\Delta x, \Delta y\right)\|^2 \lambda_{\min} , \tag{5}
$$

where $\lambda_{\min}$ is the minimal eigenvalue of $\mathbf{A}$ (see [GVL89], or [Kra00], or other linear algebra textbook).

4. Hence, we have derived that the minimal difference $f\left(N_0\right)$ is proportional (approximately) to the minimal eigenvalue of $\mathbf{A}$:

$$
f\left(N_0\right) \approx \|\left(\Delta x, \Delta y\right)\|^2 \lambda_{\min} = \epsilon^2 \lambda_{\min} \tag{6}
$$

3 Relation between image patch N_0 and matrix A properties

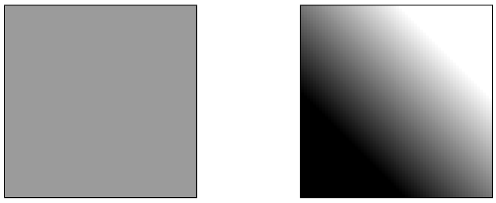
3.1 Intensity $I\left(x,y\right)$ is close to a constant function on N_0

If intensity $I\left(x,y\right)$ is close to a constant function on N_0 (Figure 1(a)) then image gradients $\nabla I\left(x,y\right),\ (x,y) \in N_0$, are close to zero vector, $\mathbf{A}$ is close to a zero matrix and thus both eigen-values λ_1 and λ_2 goes to 0.

3.2 Image gradients $\nabla I\left(x,y\right)$ have the same direction on N_0

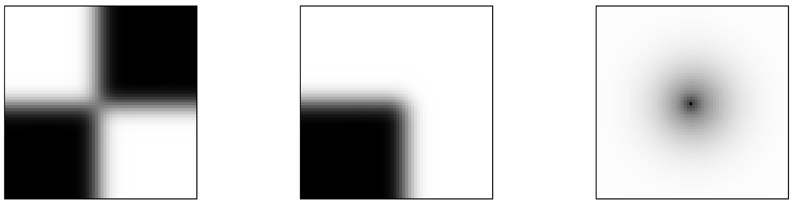
If all gradients $\nabla I\left(x,y\right)$ are of the same direction on N_0, then matrix $\mathbf{A}$ is a singular matrix, $\lambda_{\min} = 0$.

This happen for example when N_0 lies on an intensity edge (Figure 1(b)).



(a) (b)

Figure 1: Example of an image patch with a constant intensity function (a), and with an intensity edge (b).



(a) (b) (c)

Figure 2: Three examples of an image patch with gradients of different directions.

3.3 There are gradients of different directions on N_0

If N_0 contains gradients of different directions (Figure 2), then $\mathbf{A}$ is a full rank matrix, $\lambda_{\text{min}} > 0$.

4 Implementation

4.1 Image gradients

Evaluation of gradients is sensitive to noise in input data. Therefore before computing the gradients we usually smooth the image by some low pass filter. Typical example is image convolution by a Gaussian

$$I_s = I * G \ ,$$

where

$$G(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{\frac{x^2}{2\sigma_x^2}+\frac{y^2}{2\sigma_y^2}}$$

This smoothes the image and reduces the level of noise (removing high frequencies, see [ŠHB93]).

Image gradient $\nabla I\left(x,y\right)$ can be evaluated e.g. as

$$\nabla I\left(x,y\right)=\left[\begin{array}{c} I_s\left(x+1,y\right)-I_s\left(x-1,y\right) \\ I_s\left(x,y+1\right)-I_s\left(x,y-1\right) \end{array}\right] .$$

Strong image smoothing reduces the size of gradients and vice-versa. The level of image smoothing (standard deviation $\sigma_x,\ \sigma_y$ in this case) is called a *derivative scale* in this context.

4.2 Matrix A

In Section 2 we have posed

$$\mathbf{A}=\sum_{(x,y)\in N_0}\nabla I\left(x,y\right)\nabla I\left(x,y\right)^\top . \tag{7}$$

Small N_0 increases the probability that $\mathbf{A}$ will be close to a singular matrix, i.e. that $\lambda_{\text{min}}\to 0$. In other words, that there will exist a direction $(\Delta x,\Delta y)$ causing a neglecting difference (1).

Large N_0 increases the probability that all directions $(\Delta x,\Delta y)$ will cause a significant difference (1), i.e. that $\mathbf{A}$ will be a full rank matrix.

Summation members in (7) are usually weighted to enforce gradients belonging to the center of N_0, e.g. using a 2D Gaussian:

$$\mathbf{A}=\sum_{(x,y)\in N_0}G_{\sigma_I}(x-x_0,y-y_0)\nabla I\left(x,y\right)\nabla I\left(x,y\right)^\top .$$

The size of the neighborhood N_0 (or standard deviation σ_I) is called the *integration scale.*

4.3 Response function

From previous paragraphs follows that we are looking for places in the image where matrix $\mathbf{A}$ is a full rank matrix. Therefore we need a function which will effectively measure/estimate rank deficiency of matrix $\mathbf{A}$. Let us called such a function a *response function.*

4.3.1 Harris response function:

R. Harris proposed in [HS88] the following form of response function :

$$R=\det\mathbf{A}-\kappa\,\text{trace}^2\mathbf{A} , \tag{8}$$

where $\kappa=0.04$, (Figure 3).

Since

$$\det\mathbf{A}=\lambda_1\lambda_2=\mathbf{A}(1,1)\mathbf{A}(2,2)-\mathbf{A}(1,2)^2$$

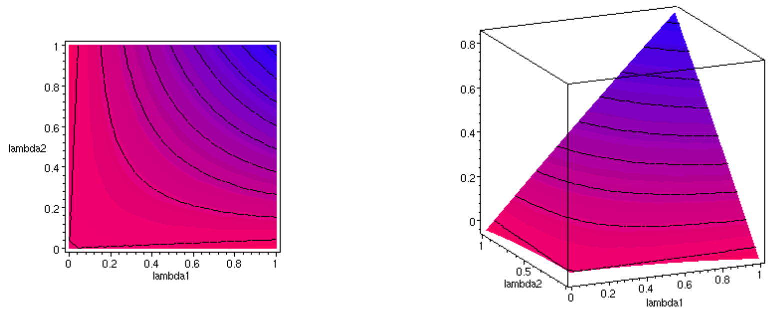


Figure 3: Harris response function.

and

$$\text{trace}\mathbf{A} = \lambda_1 + \lambda_2 = \mathbf{A}(1,1) + \mathbf{A}(2,2) \ ,$$

value of R can be computed directly from $\mathbf{A}$ without requiring evaluating eigen values λ_1, λ_2.

Here are some limit cases:

$$\begin{aligned}
\lambda_{\text{max}} \to 0 \ \wedge \ \lambda_{\text{min}} \to 0 \quad &\Rightarrow \quad R \to 0 \\
\lambda_{\text{min}} \to 0 \quad &\Rightarrow \quad R \to -4\kappa\lambda_{\text{max}}^2 \\
\lambda_{\text{min}} \to \lambda_{\text{max}} \quad &\Rightarrow \quad R \to (1-4\kappa)\lambda_{\text{max}}^2 = \max R
\end{aligned}$$

4.3.2 Other examples of response functions:

a) Harralick [HS93]:

$$\frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}$$

b)

$$\left(\frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}\right)^2$$

5 Algorithm overview

1. Smooth an image, e.g. by convolving the image with Gaussian G_{σ_x,σ_y}.
2. For each pixel compute image gradient $\nabla I(x, y)$.

3. For each pixel and a given size of N_0 (or the integrative scale σ_I), compute 2×2 matrix $\mathbf{A}$.

4. For each pixel evaluate response function $R(x, y)$.

5. Choose the interest point as local maximums of function $R(x, y)$, (e.g. by non-maximum suppression algorithm).

6 Remarks

Derivative scale: It is obvious that image smoothing reduce the size of image gradients and consequently the number detected interest points (and vice versa).

Integrative scale: As was mentioned, too small integrative scale increases probability of rank deficiency of $\mathbf{A}$. Larger integrative scale increases probability of full rank $\mathbf{A}$. However it also smoothes response function $R(x, y)$ (over image coordinates) and too large integrative scale can suppress number of local maximums of $R(x, y)$ (i.e. number of detected interest points, see the following figure).

References

[GVL89] Gene H. Golub and Charles F. Van Loan. *Matrix Computations.* The John Hopkins University Press, second edition, 1989.

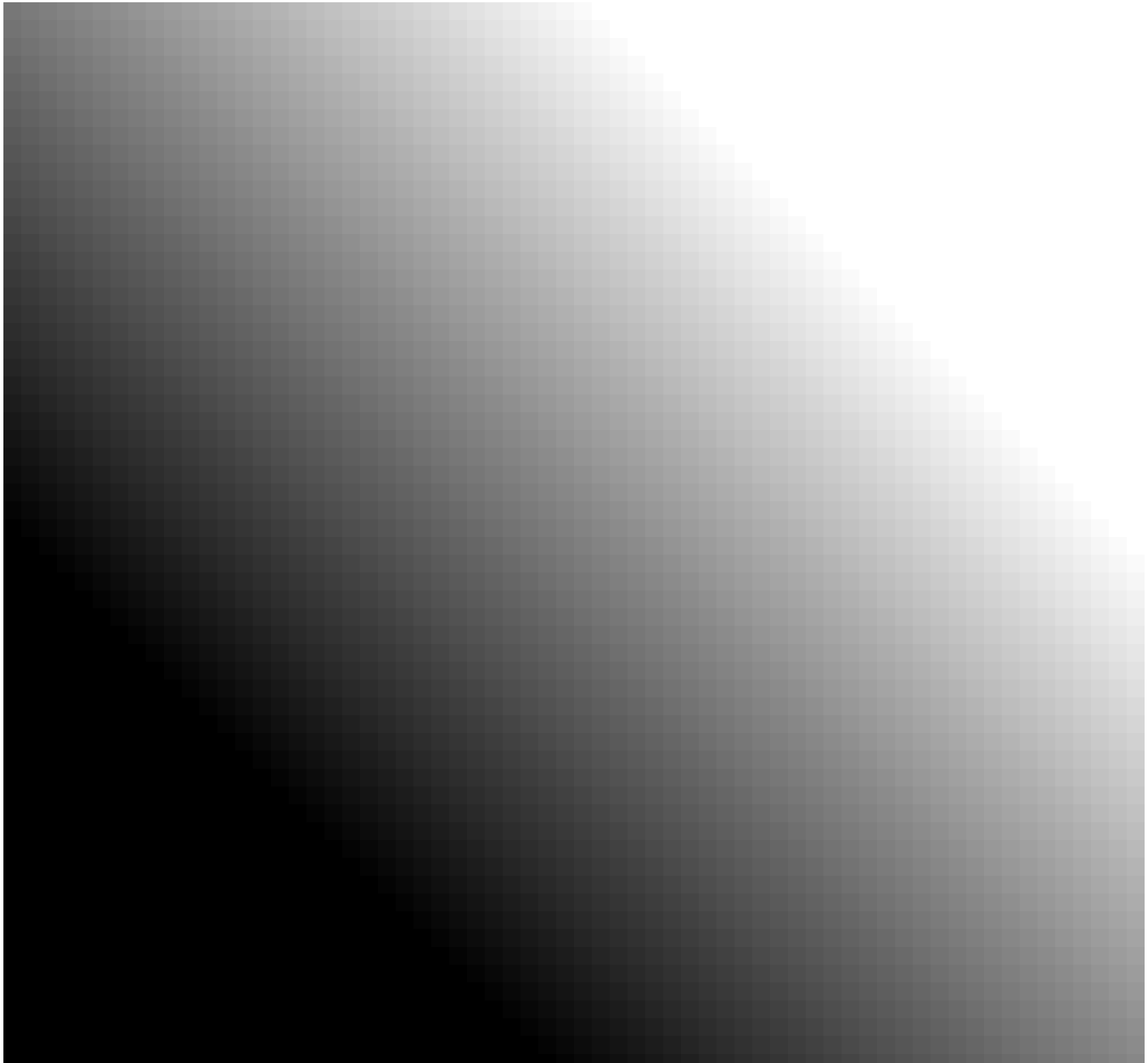
[HS88] C. Harris and M. Stephen. A combined corner and edge detection. In M. M. Matthews, editor, *Proceedings of the 4th ALVEY vision conference*, pages 147–151, University of Manchester, England, September 1988.

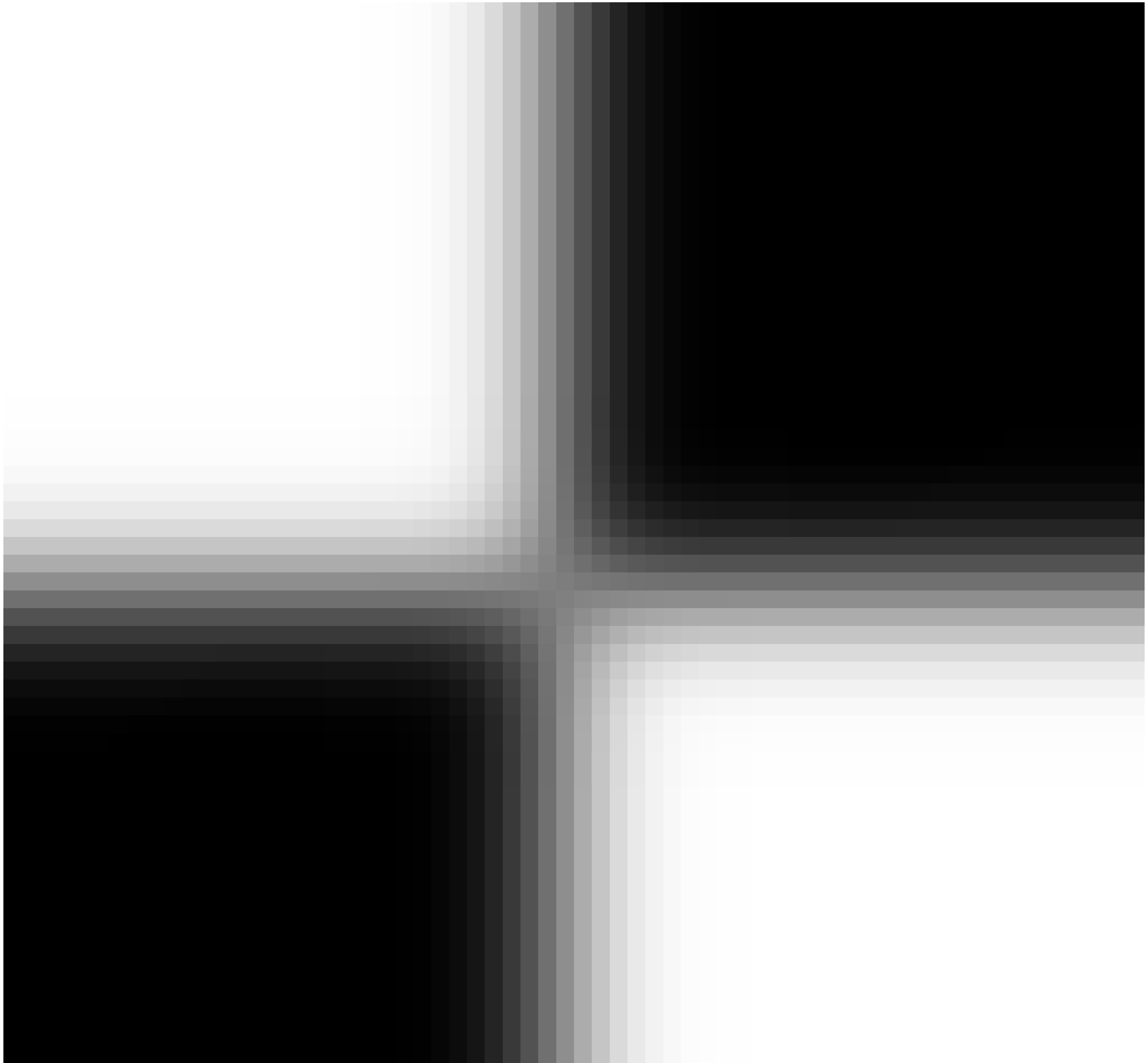
[HS93] Robert M. Haralick and Linda G. Shapiro. *Computer and Robot Vision*, volume 2. Addison-Wesley Publishing Company, 1993.

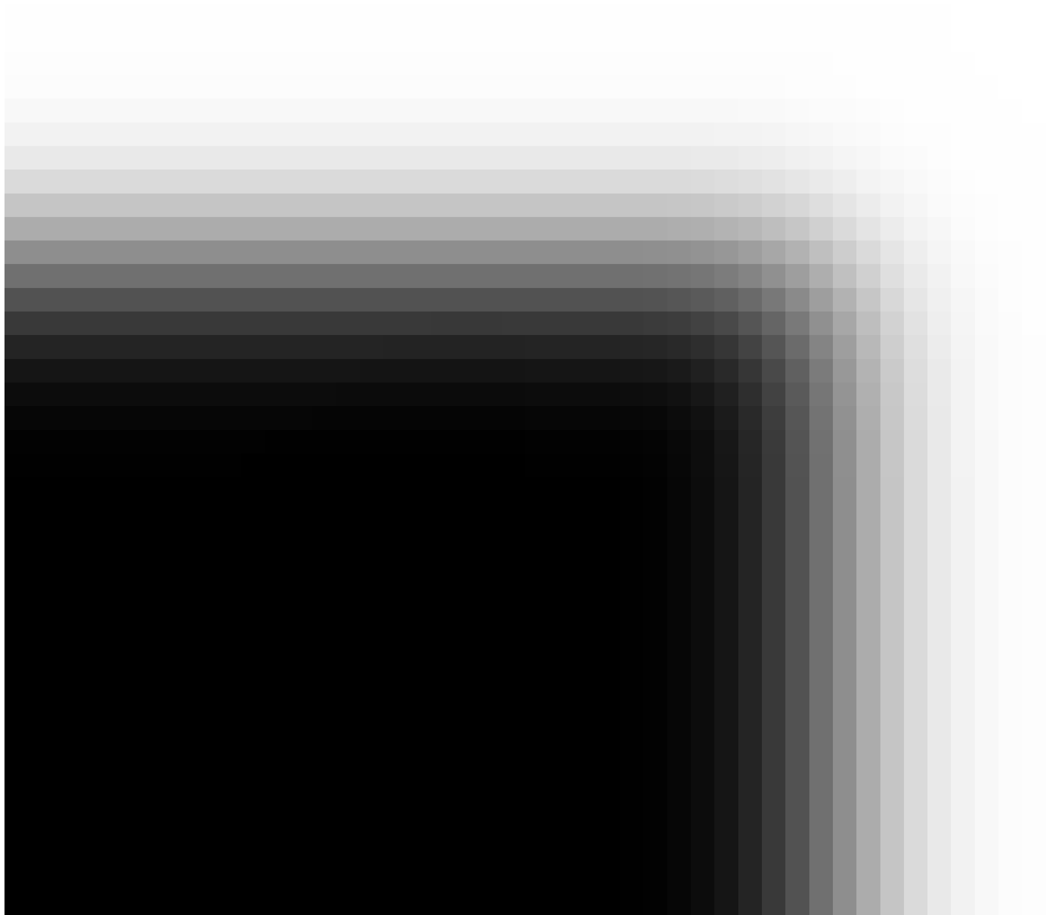
[Kra00] Eduard Krajník. *Maticový počet.* Vydavatelství ČVUT, Praha, Czech Republic, 1 edition, March 2000.

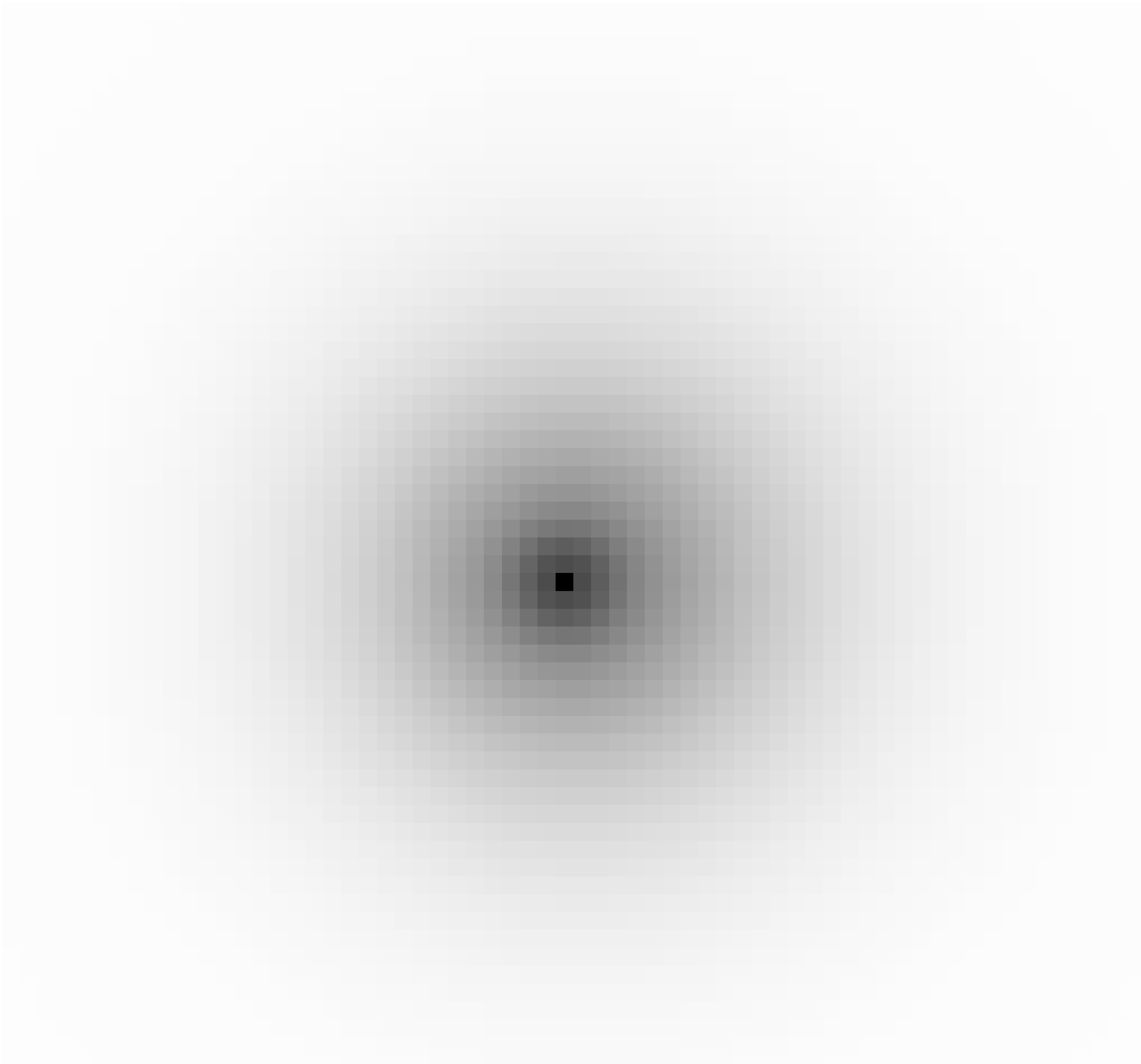
[ŠHB93] Milan Šonka, Václav Hlaváč, and Roger Boyle. *Image Processing, Analysis, and Machine Vision.* Chapman and Hall Computing, 1993.

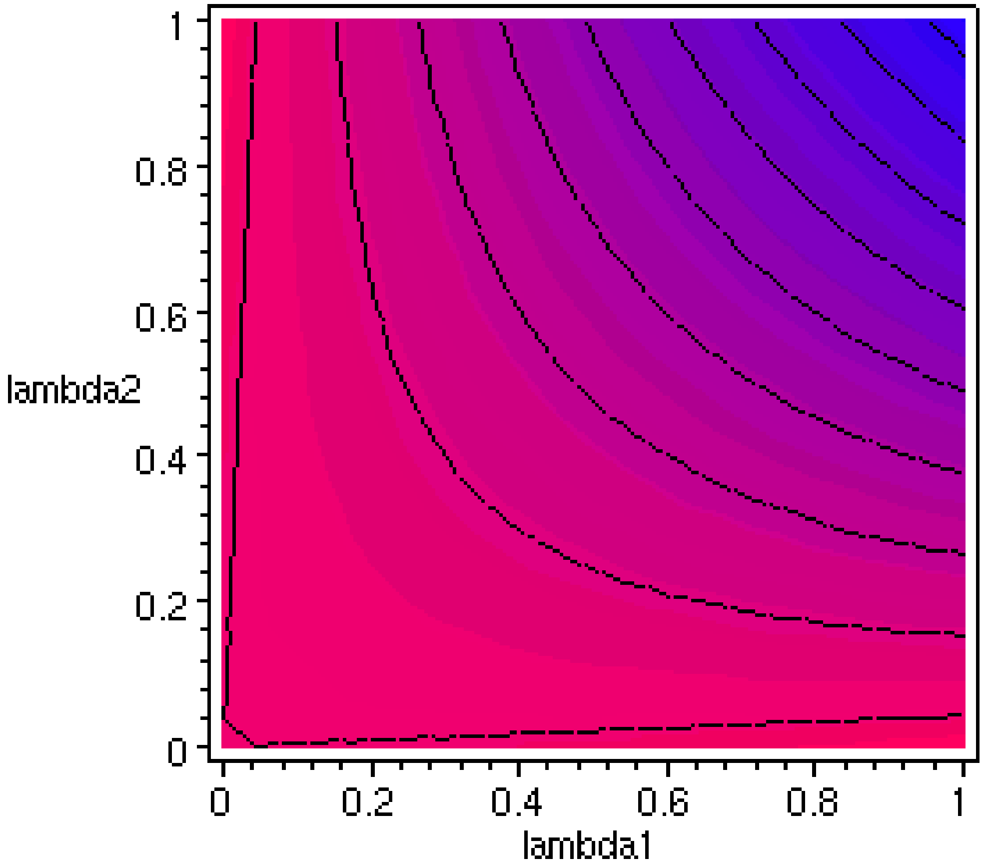
2
1
0
derivative scale σ_d
0.5
1
2
integrative scale σ_i



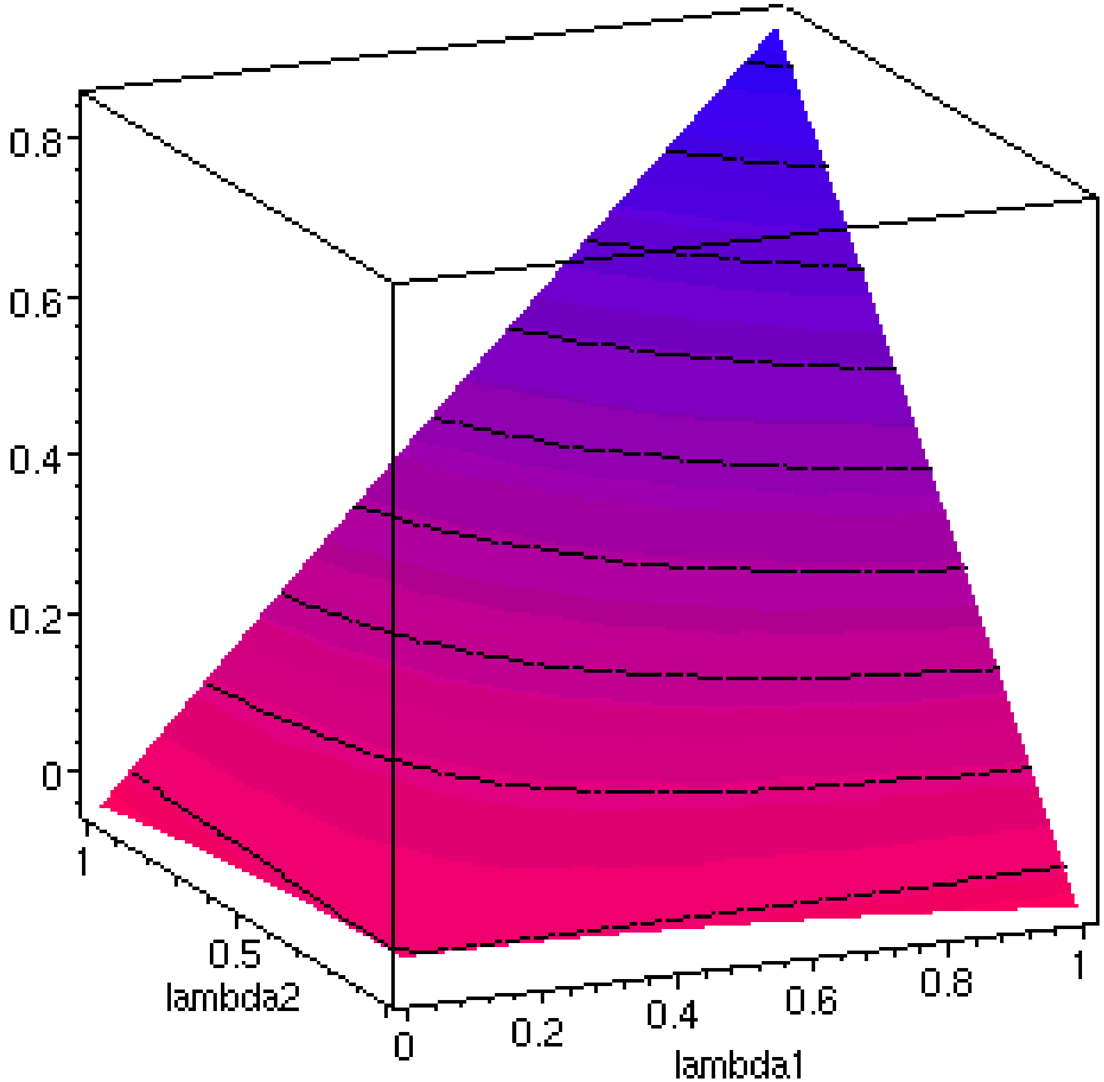








1
0.8
0.6
lambda2
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
lambda1



0.8
0.6
0.4
0.2
0
1
0.5
lambda2
0
0.2
0.4
0.6
0.8
1
lambda1