TSBB15
Computer Vision

Lecture 3

The structure tensor

Estimation of local orientation

- The gradient $\nabla f(\mathbf{x})$ can be computed with two convolutions:

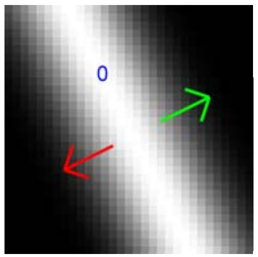
$$\nabla f(\mathbf{x}) = \begin{pmatrix} f * g_x \\ f * g_y \end{pmatrix}(\mathbf{x})$$

  g_x and g_y are derivative filters, e.g. $g_x = [-½ \;\; 0 \;\; ½]$, Sobel, or Gaussian derivatives.
- A very simple description of local orientation is given by the direction of ∇f:

$$\hat{\mathbf{n}} = \pm \frac{\nabla f}{\|\nabla f\|}$$

Estimation of local orientation

- **Problem 1:** ∇f may be zero, even though there is a well defined orientation.
- **Problem 2:** The sign of ∇f changes across a line.



Estimation of local orientation

Partial solution:

- Form the outer product of the gradient with itself: $\nabla f \, \nabla^T f$.
- This is a symmetric 2×2 matrix (tensor)
- Problem 2 solved!
- Also: The representation is unambiguous
  - Distinct orientations are mapped to distinct matrices
  - Close orientations are mapped to close matrices
  - Continuity / compatibility
- Problem 1 remains

The structure tensor

- Compute a **local average** of the outer product of the gradients :

$$\mathbf{T} = \int w(\mathbf{x})\ [\nabla f](\mathbf{x})[\nabla^T f](\mathbf{x}) d\mathbf{x}$$

- w(**x**) is some LP-filter (typically a Gaussian)
- **T** is a symmetric 2 $\times$ 2 matrix: $T_{ij} = T_{ji}$
- This construction is called the **structure tensor**
- Solves also problem 1 (why?)

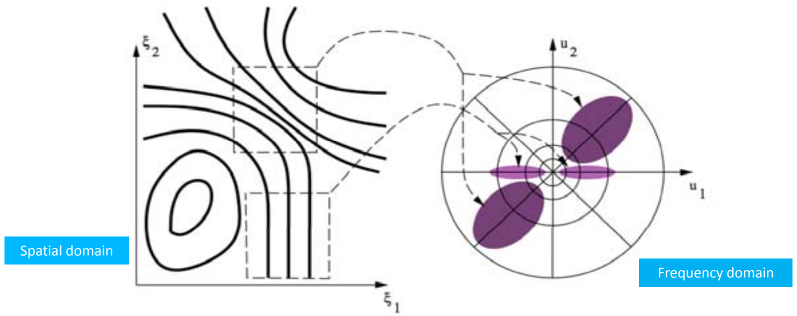
Motivation for T

- The structure tensor has been derived based on several independent approaches

- Stereo tracking (Lucas & Kanade, 1981)
- Optimal orientation (Bigün & Granlund, 1987)
- Sub-pixel refinement (Förstner & Gülch, 1987)
- Interest point detection (Harris & Stephens, 1988)

Local orientation in the Fourier domain

- Structures of different orientation end up in different places in the frequency domain



Computer Vision Laboratory

Optimal orientation estimation

Basic idea:

- The local signal $f(\mathbf{x})$ has a Fourier transform $F(\mathbf{u})$.
- We assume that f is a i1D-signal
  - F has its energy concentrated mainly on a line through the origin
- Find a line, with direction **n**, in the frequency domain that best fits the local signal

Described by Bigün & Granlund [ICCV 1987]

The structure tensor

- The solution to this optimization problem is formulated in terms of the *structure tensor*:

$$\mathbf{T} = \int w(\mathbf{x}) \, [\nabla f](\mathbf{x})[\nabla^T f](\mathbf{x}) d\mathbf{x}$$

Optimal orientation estimation

- The solution to this constrained maximization problem must satisfy

$$\mathbf{T}\,\hat{\mathbf{n}} = \lambda\,\hat{\mathbf{n}} \qquad \text{(why?)}$$

- Means: **n** is an eigenvector of **T** with eigenvalue λ
- Choose the eigenvector with the largest eigenvalue to minimize *J* (why?)
- +**n** and –**n** represent the same orientation

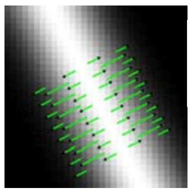
Optimal orientation estimation

- For a signal that is approximately i1D in the neighborhood of a point $\mathbf{x}_0$, with orientation ±**n**: $\nabla_w f$ is always parallel to **n** (why?)
- The gradients that are estimated around $\mathbf{x}_0$ are a scalar multiple of **n**
- The integral of their outer products results in

$$\mathbf{T} = \lambda\, \mathbf{n}\mathbf{n}^T$$

for some value λ

- λ depends on w_1, w_2, and the local signal f

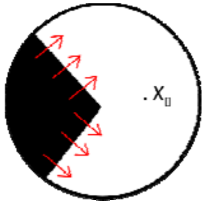


Sub-pixel refinement

- Consider a local region and let $\nabla f(\mathbf{x})$ denote the image gradient at point **x** in this region
- Let $\mathbf{x}_0$ be some point in this region
- $\langle \nabla f(\mathbf{x}) \mid \mathbf{x} - \mathbf{x}_0 \rangle$ is then a measure of compatibility between the gradient $\nabla f(\mathbf{x})$ and the point $\mathbf{x}_0$
  - Small value = high compatibility
  - High value = small compatibility

An $\mathbf{x}_0$ that lies on the edge/line that creates the gradient minimises $| \langle \nabla f(\mathbf{x}) \mid \mathbf{x} - \mathbf{x}_0 \rangle |$

Sub-pixel refinement



- In the case of more than one line/edge in the local region:
- We want to find the point $\mathbf{x}_0$ that optimally fits all these lines/edges
- We minimize

$$\epsilon(\mathbf{x}_0) = \| \langle \nabla f(\mathbf{x}) | \mathbf{x} - \mathbf{x}_0 \rangle \|^2_w$$

where *w* is a weighting function that defines the local region

Sub-pixel refinement

- The normal equations of this least squares problem are:

$$\underbrace{\begin{pmatrix} \int_\Omega w(\mathbf{x}) \left(\frac{\partial f}{\partial x_1}\right)^2 x_1 \, d\mathbf{x} + \int_\Omega w(\mathbf{x}) \frac{\partial f}{\partial x_1} \frac{\partial f}{\partial x_2} x_2 \, d\mathbf{x} \\ \int_\Omega w(\mathbf{x}) \frac{\partial f}{\partial x_1} \frac{\partial f}{\partial x_2} x_1 \, d\mathbf{x} + \int_\Omega w(\mathbf{x}) \left(\frac{\partial f}{\partial x_2}\right)^2 x_2 \, d\mathbf{x} \end{pmatrix}}_{:=\mathbf{T}} \mathbf{x}_0 = \underbrace{\int_\Omega w(\mathbf{x}) \nabla f(\mathbf{x}) \nabla^T f(\mathbf{x}) \, \mathbf{x} \, d\mathbf{x}}_{:=\mathbf{b}}$$

The structure tensor!

- Solve the linear equation: $\mathbf{T}\,\mathbf{x}_0 = \mathbf{b}$

This equation is solved for each local region of the image!

The Harris-Stephen detector

- A Taylor expansion of the image intensity *I* at point (*u*, *v*):

$$I(u + n_x, v + n_y) \approx I(u, v) + \nabla I \cdot (n_x, n_y)$$
$$\approx I(u, v) + \nabla I \cdot \mathbf{n}$$

The Harris-Stephen detector

- *S*(*x*,*y*) is a measure of how much *I*(*u*, *v*) deviates from $I(u + n_x, v + n_y)$ in a local region Ω, as a function of (n_x, n_y):

$$S(n_x, n_y) = \|I(u + n_x, v + n_y) - I(u, v)\|^2 =$$
$$= \int_\Omega w(u, v) \cdot \left| I(u + n_x, v + n_y) - I(u, v) \right|^2 du dv \approx$$
$$\approx \int_\Omega w(u, v) \cdot (\nabla I \cdot \mathbf{n})^2 du dv =$$
$$= \mathbf{n}^\top \underbrace{\left[\int_\Omega w(u, v) \cdot (\nabla I \, \nabla^\top I) du dv \right]}_{:=\mathbf{T}} \mathbf{n} = \mathbf{n}^\top \mathbf{T} \, \mathbf{n}$$

The Harris-Stephen detector

- If Ω contains a linear structure, then S is small (=0) when **n** is parallel to the line/edge
  - **T** must have one small ($\approx$ 0) eigenvalue
- If Ω contains an interest point (corner) any displacement (n_x, n_y) gives a relatively large S
  - Both eigenvalues of **T** must be relatively large
- By analyzing the eigenvalues λ_1, λ_2 of **T**:
  - If λ_1 large and λ_2 small: line/edge
  - If both λ_1 and λ_2 large: interest point
- See Harris measure below

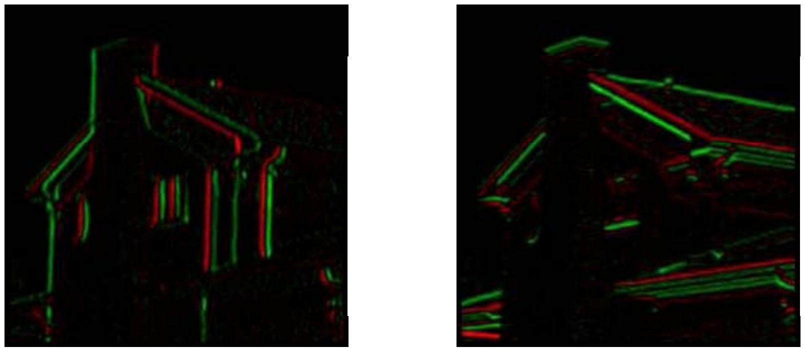
Computation of the structure tensor

- Use 1D or separable 2D filters to estimate ∇f
  - 4 1D convolutions to produce ∇f (Sobel)
- Map ∇f to $\nabla f \, \nabla^T f$
  - A point-wise operation
- Convolve T_{11}, T_{12}, and T_{22} with w
  - Choose w separable $\Rightarrow$ 6 more 1D convolutions
  - Note that $T_{21} = T_{12}$

Example: Structure tensor

Original image

Example: Structure tensor

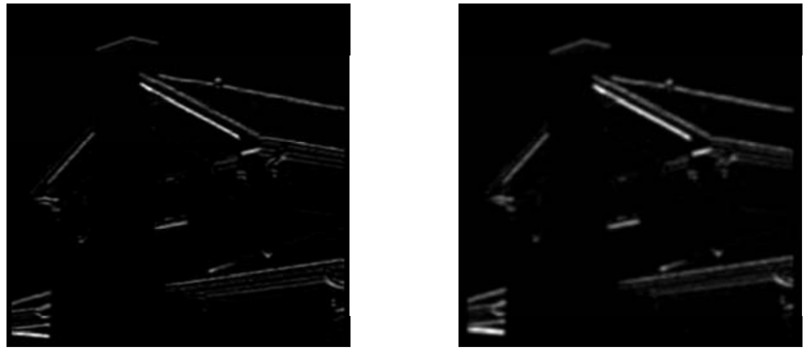


f_x Gradient images f_y

Example: Structure tensor

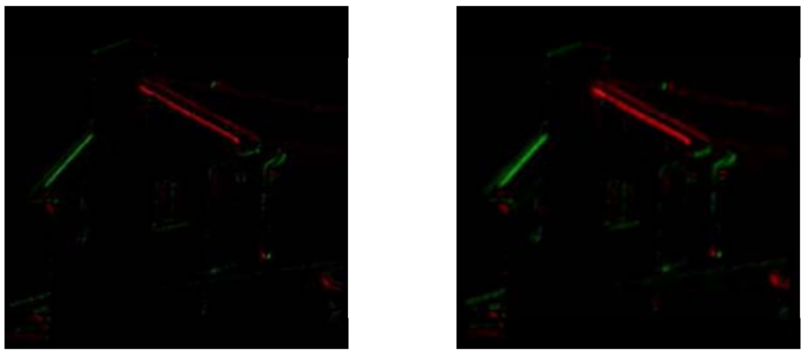
Before averaging T_{11} image After averaging

Example: Structure tensor



Before averaging T_{22} image After averaging

Example: Structure tensor



Before averaging T_{12} image After averaging

Structure tensor in 2D

- The structure tensor is computed as follows:

$$\mathbf{T} = \int w(\mathbf{x})\, [\nabla f](\mathbf{x})[\nabla^T f](\mathbf{x}) d\mathbf{x}$$

- *w* is assumed to be a positive and symmetric function that localizes the orientation estimate described by **T**, e.g. a Gaussian function

Example: Structure tensor in 2D

- In the general 2D case, we obtain

$$\mathbf{T} = \lambda_1 \, \hat{\mathbf{e}}_1 \, \hat{\mathbf{e}}_1^T + \lambda_2 \, \hat{\mathbf{e}}_2 \, \hat{\mathbf{e}}_2^T \quad \text{(why?)}$$

- where $\lambda_1 \geq \lambda_2$ are the eigenvalues of **T** and $\hat{\mathbf{e}}_1$, $\hat{\mathbf{e}}_2$ are the corresponding normalized eigenvectors
- We have already shown that for locally i1D signals we get $\lambda_1 \geq 0$ and $\lambda_2 = 0$

Example: Structure tensor in 2D

- If the local signal is not i1D, ∇f is not parallel to some **n** for all points **x** in the local region, i.e. the terms in the integral that forms **T** are not scalar multiples of each other
- Consequently: $\lambda_2 > 0$
- The idea of optimal orientation becomes less relevant the closer λ_2 gets to λ_1

Isotropic tensor

- If we assume that the orientation is uniformly distributed in the local integration support, we get $\lambda_1 \approx \lambda_2$:

$$\begin{aligned} \mathbf{T} &= \lambda_1 \, \hat{\mathbf{e}}_1 \, \hat{\mathbf{e}}_1^T + \lambda_1 \, \hat{\mathbf{e}}_2 \, \hat{\mathbf{e}}_2^T \\ &= \lambda_1 (\hat{\mathbf{e}}_1 \, \hat{\mathbf{e}}_1^T + \hat{\mathbf{e}}_2 \, \hat{\mathbf{e}}_2^T) \\ &= \lambda_1 \, \mathbf{I} \end{aligned}$$

The identity matrix

- i.e. **T** is *isotropic*: $\mathbf{n}^T \mathbf{T} \, \mathbf{n} = \mathbf{n}^T \mathbf{I} \, \mathbf{n} = 1$

Why is the parenthesis equal to **I**?

Confidence measures

- From det **T** and tr **T** we can define two confidence measures:

$$c_1 = \frac{\mathrm{tr}^2\mathbf{T} - 4 \,\det\mathbf{T}}{\mathrm{tr}^2\mathbf{T} - 2 \,\det\mathbf{T}} \qquad c_2 = \frac{2 \,\det\mathbf{T}}{\mathrm{tr}^2\mathbf{T} - 2 \,\det\mathbf{T}}$$

Confidence measures

- Using the identities

  tr **T** = $T_{11} + T_{22} = \lambda_1 + \lambda_2$

  det **T** = $T_{11}T_{22} - T_{12}^2 = \lambda_1\lambda_2$

  we obtain

$$c_1 = \frac{(\lambda_1 - \lambda_2)^2}{\lambda_1^2 + \lambda_2^2} \qquad c_2 = \frac{2\,\lambda_1\,\lambda_2}{\lambda_1^2 + \lambda_2^2}$$

  and $c_1 + c_2 = 1$ (why?)

Confidence measures

- Easy to see that
  - i1D signals give $c_1 = 1$ and $c_2 = 0$
  - Isotropic **T** gives $c_1 = 0$ and $c_2 = 1$
  - In general: an image region is somewhere between these two ideal cases
- **An advantage of these measures** is that they can be computed from **T** without explicitly computing the eigenvalues λ_1 and λ_2

Decomposition of T

- We can always decompose **T** into an i1D part and an isotropic part:

$\lambda_1 \geq \lambda_2$

$$\begin{aligned}\mathbf{T} &= \lambda_1\,\hat{\mathbf{e}}_1\,\hat{\mathbf{e}}_1^T + \lambda_2\,\hat{\mathbf{e}}_2\,\hat{\mathbf{e}}_2^T \\ &= (\lambda_1 - \lambda_2)\,\hat{\mathbf{e}}_1\,\hat{\mathbf{e}}_1^T + \lambda_2\,(\hat{\mathbf{e}}_1\,\hat{\mathbf{e}}_1^T + \hat{\mathbf{e}}_2\,\hat{\mathbf{e}}_2^T) \\ &= (\lambda_1 - \lambda_2)\,\hat{\mathbf{e}}_1\,\hat{\mathbf{e}}_1^T + \lambda_2\,\mathbf{I}\end{aligned}$$

Double angle representation

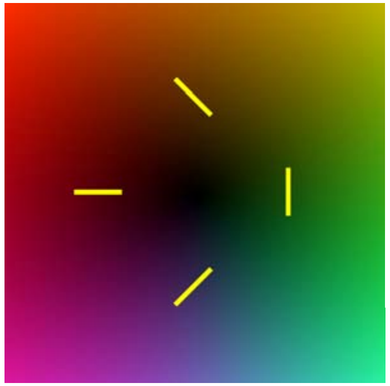
With this result at hand:

Remember: $\lambda_1 \geq \lambda_2$

$$\begin{aligned}\mathbf{z} &= \begin{pmatrix} T_{11} - T_{22} \\ 2\,T_{12} \end{pmatrix} = \\ &= (\lambda_1 - \lambda_2)\begin{pmatrix} \cos^2\alpha - \sin^2\alpha \\ 2\ \cos\alpha\ \sin\alpha \end{pmatrix} = \\ &= (\lambda_1 - \lambda_2)\begin{pmatrix} \cos 2\alpha \\ \sin 2\alpha \end{pmatrix}\end{aligned}$$

z is a *double angle representation* of the local orientation

Color coding of the double angle representation

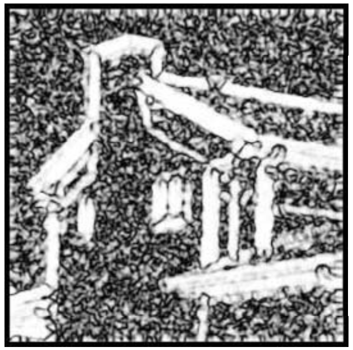


Example

trace of **T**

determinant of **T**

Example

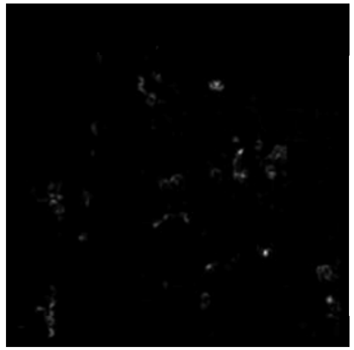


c_1

c_2

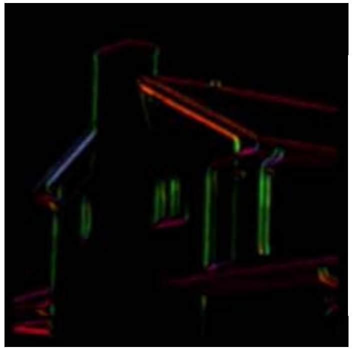
Example

λ_1



λ_2

Example



Double angle descriptor

Rank measures

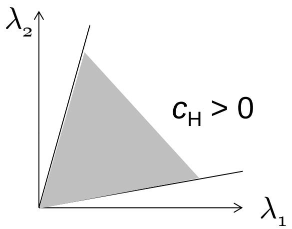
- The rank of a matrix (linear map) is defined as the dimension of its range
- We can think of c_1 and c_2 as (continuous) rank measures, since
  - i1D signal ⇒ **T** has rank 1 ⇒ $c_1 = 1$ and $c_2 = 0$.
  - Isotropic signal ⇒ **T** has rank 2 ⇒ $c_1 = 0$ and $c_2 = 1$.

Harris measure

- The Harris-Stephen detector is based on c, defined as

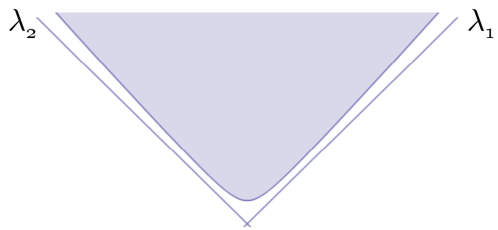
$$c_{\mathrm{H}} = \det \mathbf{T} - \kappa \left(\mathrm{trace} \mathbf{T}\right)^2, \quad \kappa \approx 0.05$$
$$= \lambda_1 \lambda_2 - \kappa(\lambda_1 + \lambda_2)^2$$

Different values for κ have been proposed in the literature!



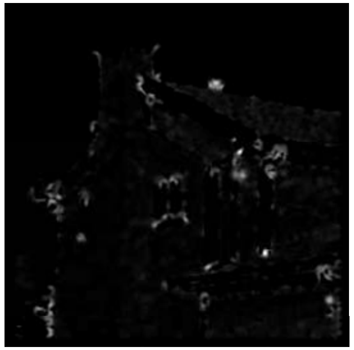
Harris measure

- By detecting points of local maxima in c_H, where $c_H > \tau$, we assure that the eigenvalues of **T** at such a point lie in the colored region below



Example

Original



Harris