TSBB15 Computer Vision

Lecture 3
The structure tensor

LINKÖPING UNIVERSITY

Estimation of local orientation

- A very simple description of local orientation at image point **p** = (*u*,*v*) is given by:

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

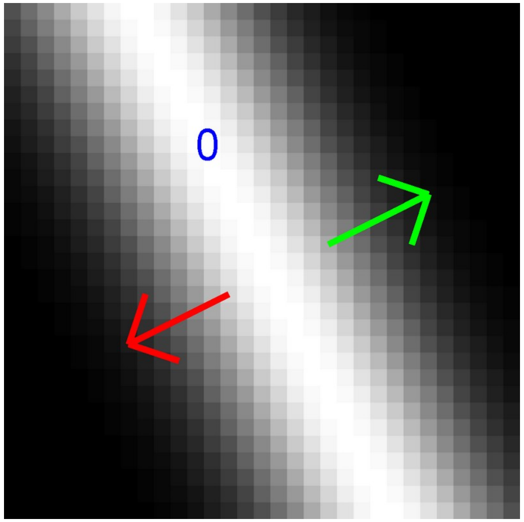
- Here, ∇*I* is the gradient at point **p** of the image intensity *I.* In practice:

$$\nabla I = \begin{pmatrix} \frac{\partial}{\partial u} \\ \frac{\partial}{\partial v} \end{pmatrix} (w_1 * I)$$

LINKÖPING UNIVERSITY

Estimation of local orientation

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



Estimation of local orientation

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

The structure tensor

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

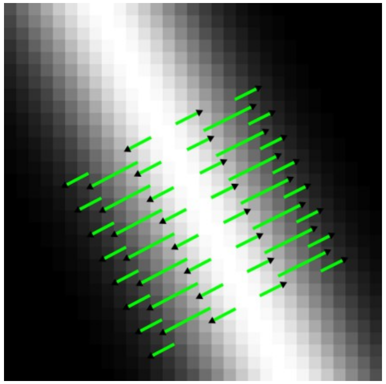
$$\mathbf{T}(\mathbf{p}) = \int w_2(\mathbf{x})[\nabla I](\mathbf{x})[\nabla^{\mathrm{T}} I](\mathbf{x})\, d\mathbf{x}$$

- Here, **x** represent an offset from **p**
- $w_2(\mathbf{x})$ is some LP-filter (typically a Gaussian)
- **T** is a symmetric 2 × 2 matrix: $T_{ij} = T_{ji}$
- This construction is called the **structure tensor**
- Solves also problem 1 (why?)
- **T** is computed for each point **p** in the image

Orientation representation

- For a signal that is approximately i1D in the neighborhood of a point **p**, with orientation ±**n**: ∇I is always parallel to **n** (why?)
- The gradients that are estimated around **p** are a scalar multiple of **n**
- The average 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 I

Motivation for **T**

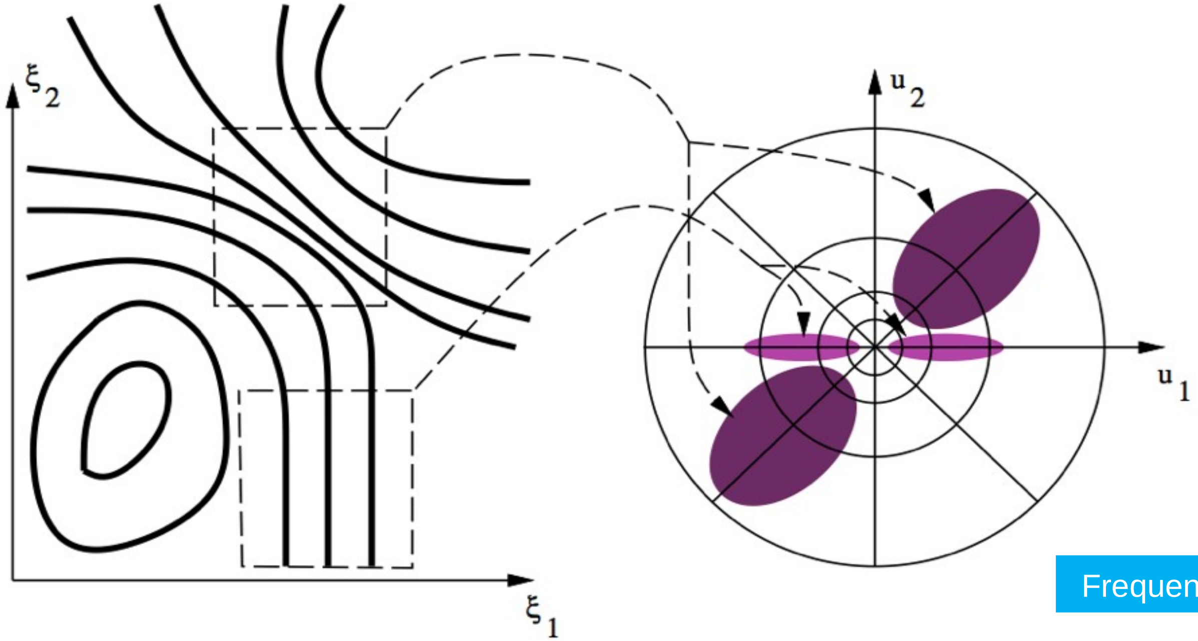
- The structure tensor has been derived based on several independent approaches

For example

- Stereo tracking (Lucas & Kanade, 1981) (Lec. 5)
- 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



Spatial domain

Frequency domain

Optimal orientation estimation

- Basic idea:
- The local signal $I(\mathbf{x})$ 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 $\mathbf{n}$, in the frequency domain that best fits the energy of F
- Described by Bigün & Granlund [ICCV 1987]

Optimal orientation estimation

- The solution to this constrained maximization problem must satisfy

$$\mathbf{T}\hat{\mathbf{n}} = \lambda\hat{\mathbf{n}}$$

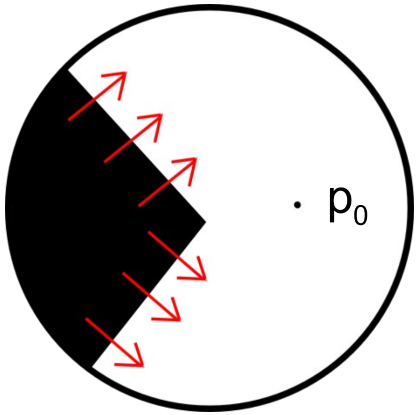
- Means: **n** is an eigenvector of **T** with eigenvalue λ
- In fact: Choose the eigenvector with the largest eigenvalue for best fit

Sub-pixel refinement

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

A $\mathbf{p}_0$ that lies on the edge/line that creates the gradient minimizes
$| \langle \nabla I(\mathbf{p}) \mid \mathbf{p} - \mathbf{p}_0 \rangle |$

Sub-pixel refinement



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

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

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

Sub-pixel refinement

- The normal equations of this least squares problem are:

$$\underbrace{\begin{pmatrix} \int_\Omega w(\mathbf{p}) \left(\frac{\partial I}{\partial u}\right)^2 \, d\mathbf{p} & \int_\Omega w(\mathbf{p}) \frac{\partial I}{\partial u} \frac{\partial I}{\partial v} \, d\mathbf{p} \\ \int_\Omega w(\mathbf{p}) \frac{\partial I}{\partial u} \frac{\partial I}{\partial v} \, d\mathbf{p} & \int_\Omega w(\mathbf{p}) \left(\frac{\partial I}{\partial v}\right)^2 \, d\mathbf{p} \end{pmatrix}}_{:=\mathbf{T}} \mathbf{p}_0 =$$

$$= \underbrace{\int_\Omega w(\mathbf{p}) \nabla I(\mathbf{x}) \nabla^T I(\mathbf{p}) \, \mathbf{p} \, d\mathbf{p}}_{:=\mathbf{b}}$$

The structure tensor!

- Solve the linear equation: $\mathbf{T}\,\mathbf{p}_o = \mathbf{b}$

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

The Harris-Stephens detector

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

$$
\begin{aligned}
I(u+n_u, v+n_v) &\approx I(u,v) + \nabla I \cdot (n_u, n_v) \\
&\approx I(u,v) + \nabla I \cdot \mathbf{n}
\end{aligned}
$$

The Harris-Stephens detector

- $S(n_u, n_v)$ is a measure of how much $I(u, v)$ deviates from $I(u+n_u, v+n_v)$ in a local region Ω, as a function of (n_u, n_v):

$$
\begin{aligned}
S(n_u, n_v) &= \|I(u+n_u, v+n_v) - I(u,v)\|^2 \\
&= \int_\Omega w(u,v) \cdot |I(u+n_u, v+n_v) - I(u,v)|^2 \, du dv \\
&\approx \int_\Omega w(u,v) \cdot (\nabla I \cdot \mathbf{n})^2 \, du dv \\
&= \mathbf{n}^\mathrm{T} \underbrace{\left[\int_\Omega w(u,v) \cdot (\nabla I \nabla^\mathrm{T} I) \, du dv \right]}_{:=\mathbf{T}} \mathbf{n} = \mathbf{n}^\mathrm{T} \mathbf{T} \mathbf{n}
\end{aligned}
$$

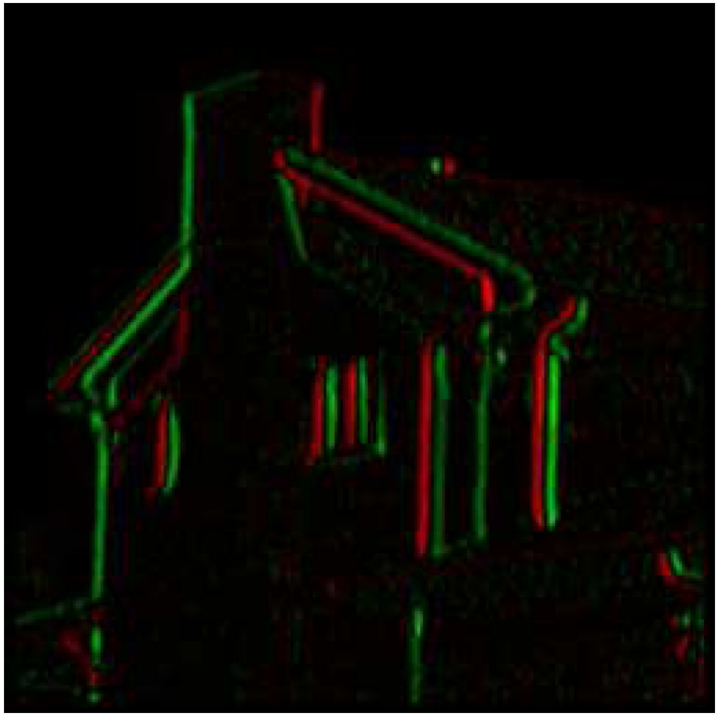
The Harris-Stephens detector

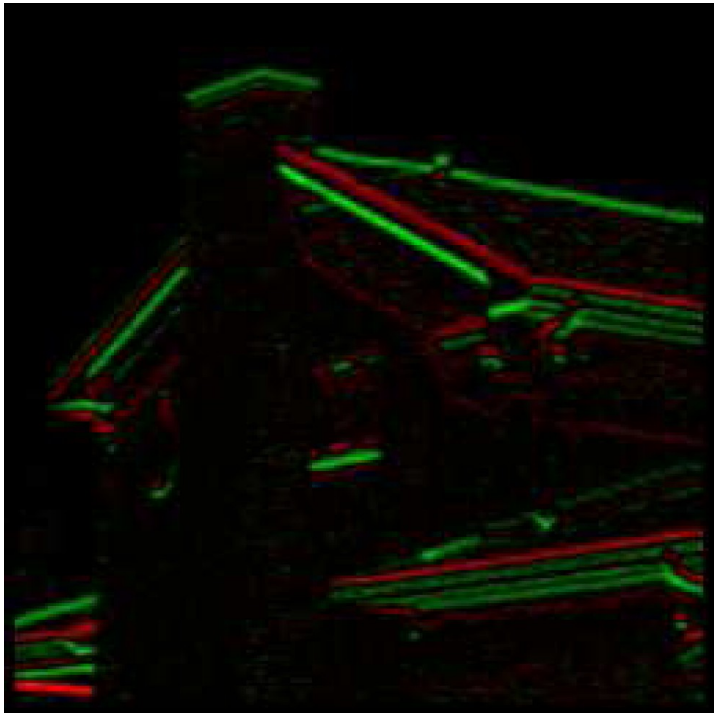
- If Ω contains a linear structure, then S is small (=0) when **n** is parallel to the line/edge
  - **T** must have one small (≈ 0) eigenvalue
- If Ω contains an interest point (corner) any displacement (n_u, n_v) gives a relatively large S
  - Both eigenvalues of **T** must be relatively large
- By 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

Example: Structure tensor

Original image

Example: Structure tensor





f_x Gradient images f_y

Example: Structure tensor

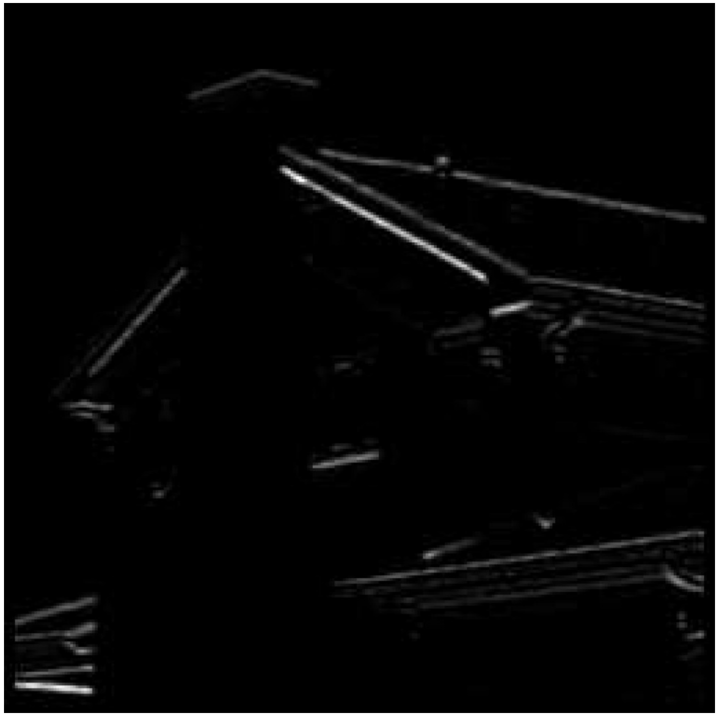
Before averaging

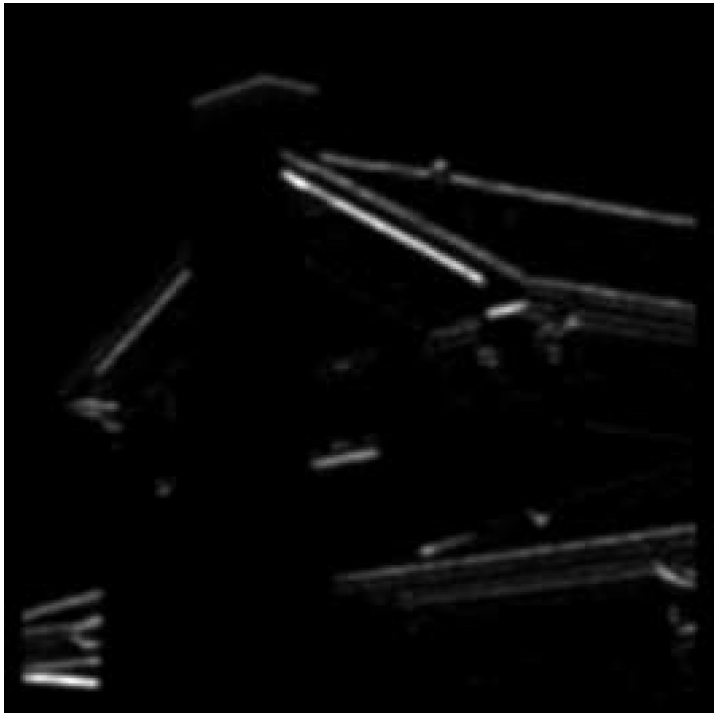
T_{11} image

After averaging

LINKÖPING UNIVERSITY

Example: Structure tensor



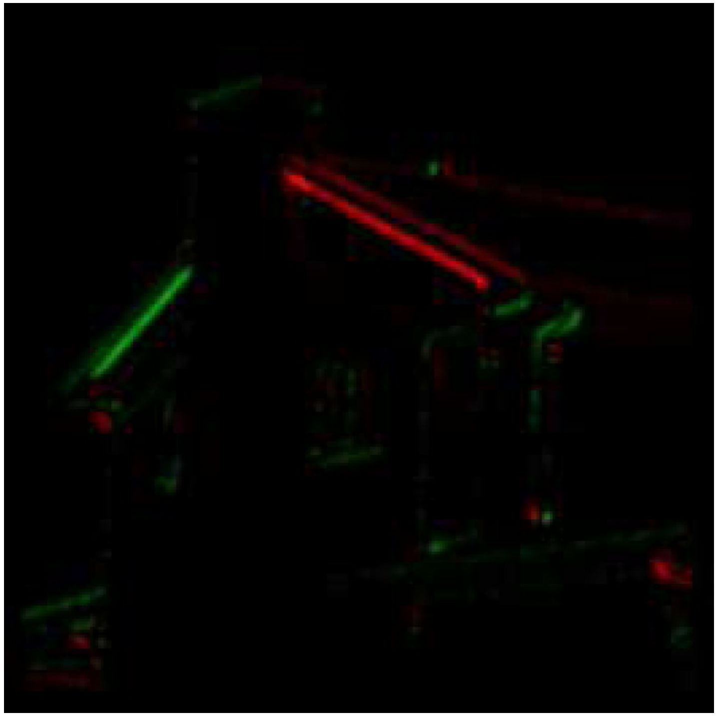


Before averaging

T_{22} image

After averaging

Example: Structure tensor



Before averaging

T_{12} image

After averaging

Example: Structure tensor in 2D

- In the general 2D case, we obtain

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

- 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 $\lambda_2 = 0$

Structure tensor in 2D, i0D

- If the local signal is constant (i0D), then $\nabla I = 0$
- Consequently: $\mathbf{T} = 0$
- Consequently: $\lambda_1 = \lambda_2 = 0$
- The idea of optimal orientation becomes less relevant the closer λ_1 gets to 0

Structure tensor in 2D, i2D

- If the local signal is i2D, ∇I 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: $\lambda_2 > 0$ if *f* not i1D
- 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 $\mathbf{T} = T_{11} + T_{22} = \lambda_1 + \lambda_2$
  - det $\mathbf{T} = T_{11}T_{22} - T_{12}^2 = \lambda_1\lambda_2$
- 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:

$$\mathbf{T} = \lambda_1 \, \hat{\mathrm{e}}_1 \, \hat{\mathrm{e}}_1^T + \lambda_2 \, \hat{\mathrm{e}}_2 \, \hat{\mathrm{e}}_2^T$$

$$= (\lambda_1 - \lambda_2) \, \hat{\mathrm{e}}_1 \, \hat{\mathrm{e}}_1^T + \lambda_2 \, (\hat{\mathrm{e}}_1 \, \hat{\mathrm{e}}_1^T + \hat{\mathrm{e}}_2 \, \hat{\mathrm{e}}_2^T)$$

$$= (\lambda_1 - \lambda_2) \, \hat{\mathrm{e}}_1 \, \hat{\mathrm{e}}_1^T + \lambda_2 \, \mathbf{I}$$

$\lambda_1 \geq \lambda_2$

Double angle representation

- With this result at hand:

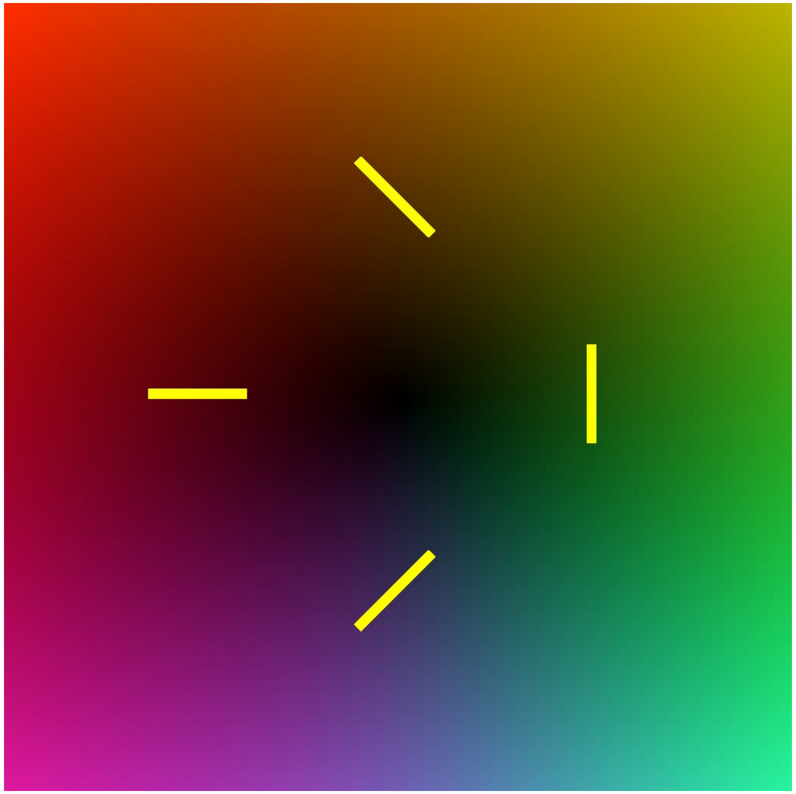
$$\begin{aligned}\mathbf{z} &= \begin{pmatrix} T_{11} - T_{22} \\ 2T_{12} \end{pmatrix} \\ &= (\lambda_1 - \lambda_2) \begin{pmatrix} \cos^2\alpha - \sin^2\alpha \\ 2\cos\alpha\sin\alpha \end{pmatrix} \\ &= (\lambda_1 - \lambda_2) \begin{pmatrix} \cos 2\alpha \\ \sin 2\alpha \end{pmatrix}\end{aligned}$$

Remember:
$\lambda_1 \geq \lambda_2$

z cannot distinguish between i0D and i2D

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

Color coding of the double angle representation



LINKÖPING UNIVERSITY

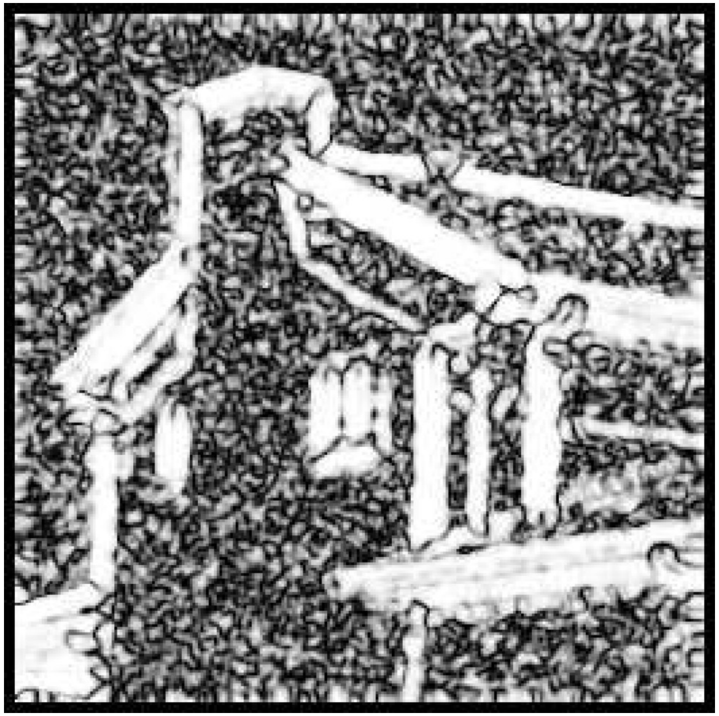
Example

trace of T

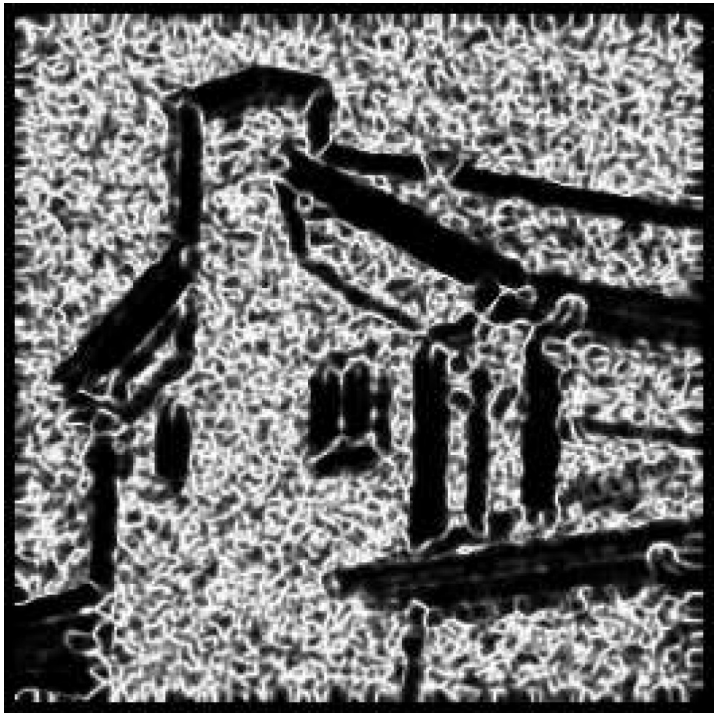
determinant of T

LINKÖPING UNIVERSITY

Example



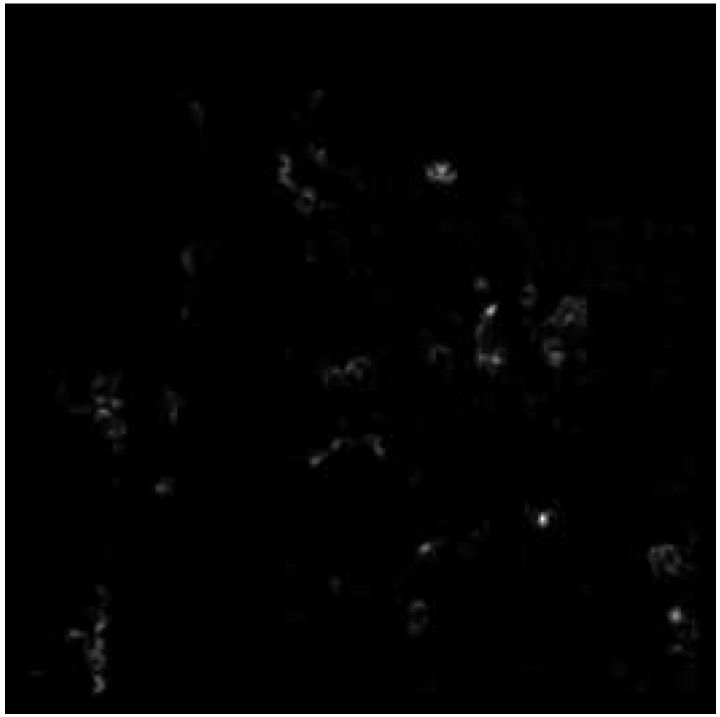
c_1



c_2

Example

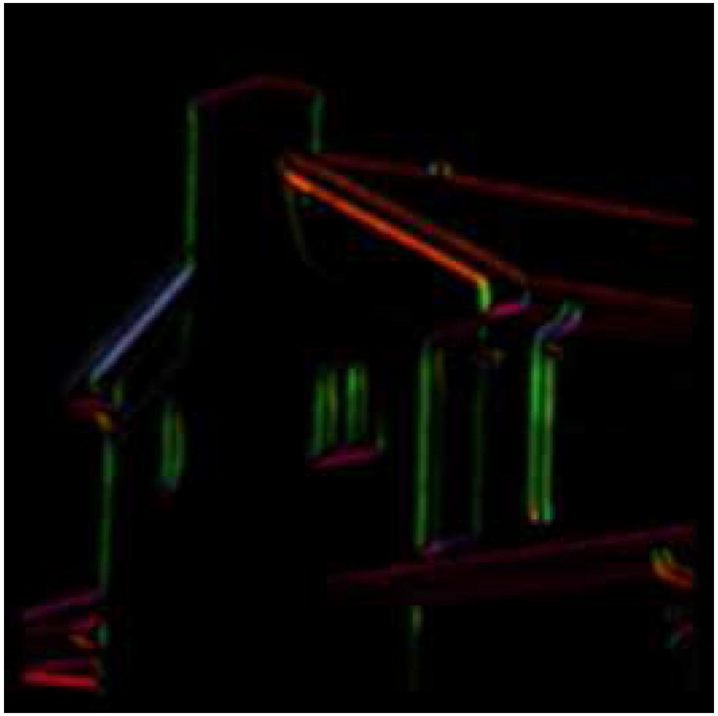
λ_1



λ_2

LINKÖPING UNIVERSITY

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 $c_2 = 1$.

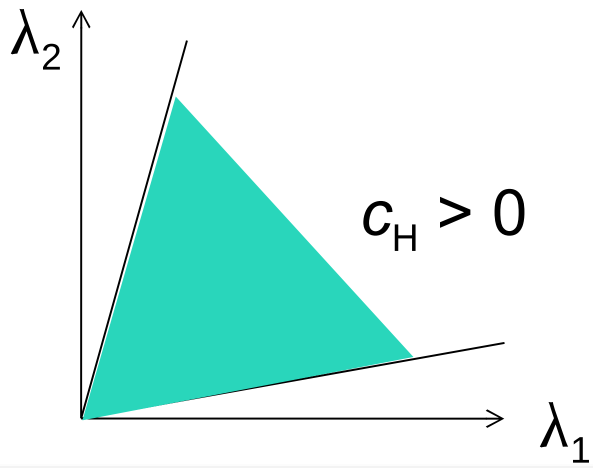
LINKÖPING UNIVERSITY

Harris measure

- The Harris-Stephens detector is based on c_H, defined as

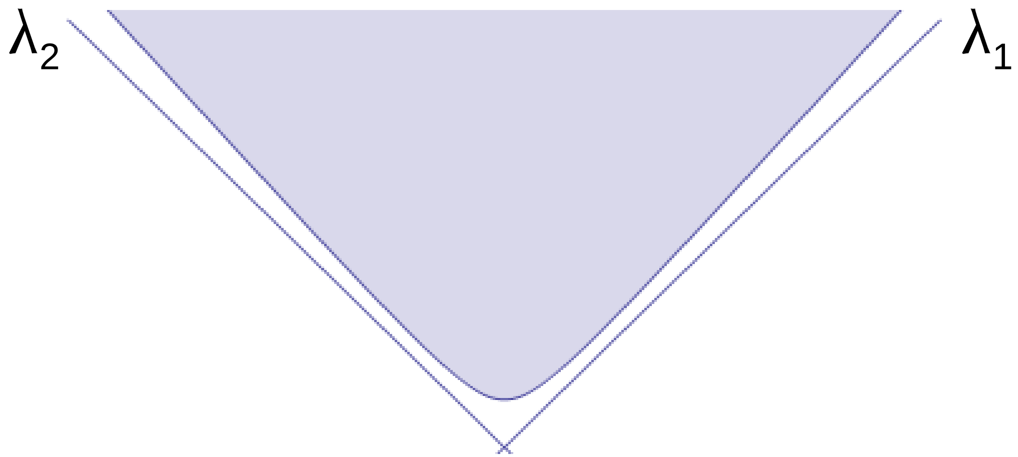
$$\begin{aligned} c_H &= \det \mathbf{T} - \kappa(\mathrm{trace}\mathbf{T})^2, \qquad \kappa \approx 0.05 \\ &= \lambda_1\lambda_2 - \kappa(\lambda_1 + \lambda_2)^2 \end{aligned}$$

Different values for κ have been proposed in the literature!



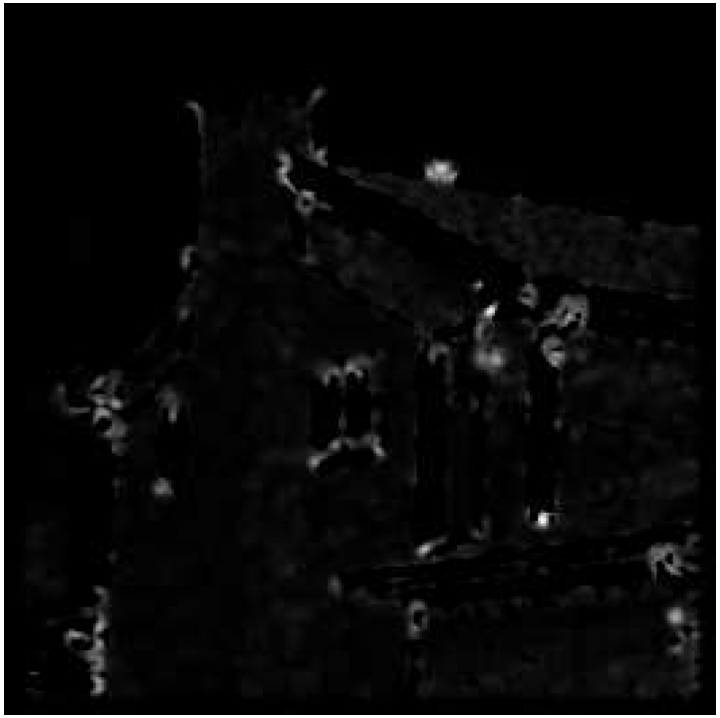
Harris measure

- 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

Origina
l



Harris