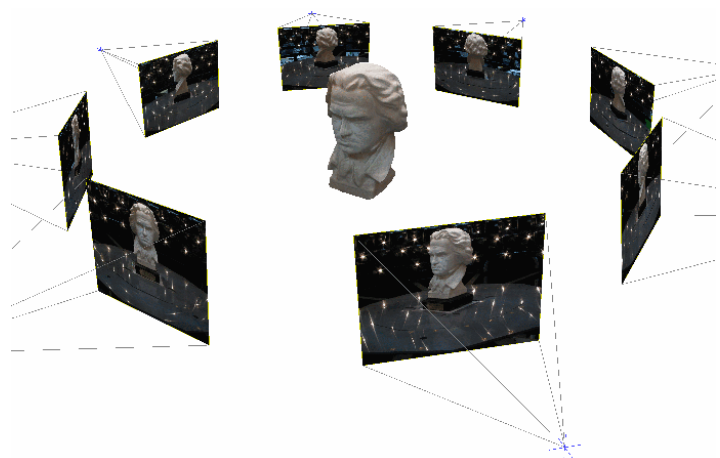
3D Machine Vision

Feature Matching

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

Basics of Feature Matching

1. Correspondence Problem
2. Correlation
3. Interest Points
4. Scale Space
5. Descriptors
6. Optical Flow



Feature Matching

Correspondence Problem

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

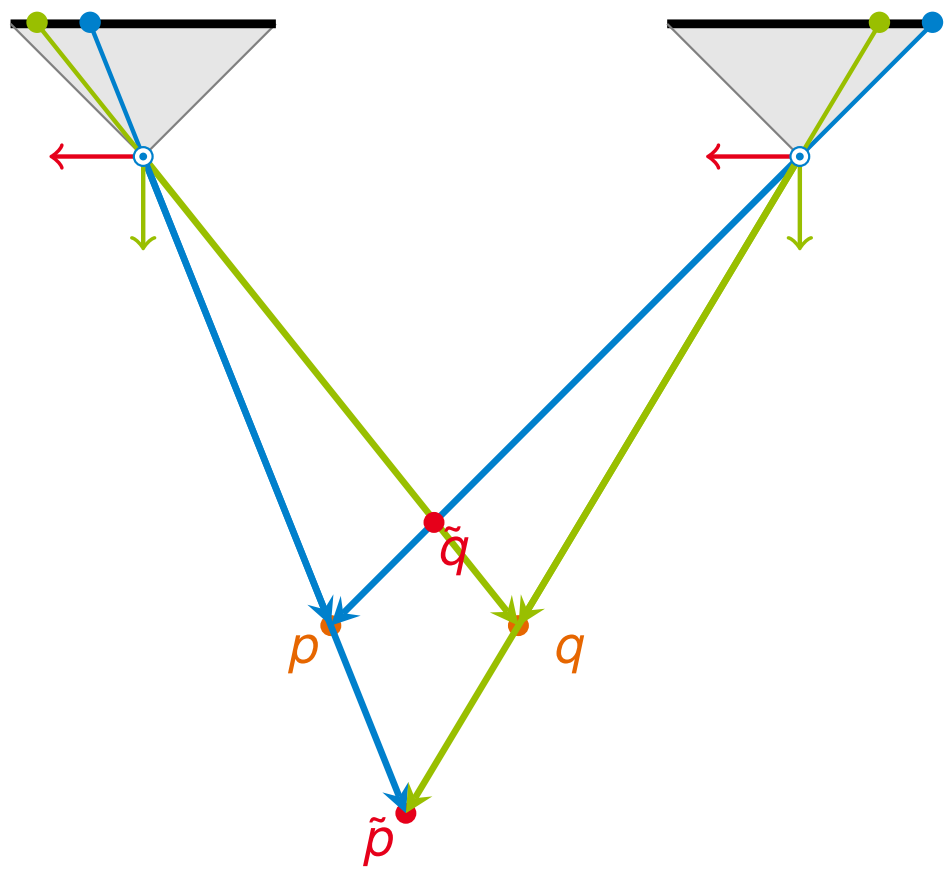
Problem Given an image point in some image, what is the corresponding point in another image, which is the projection of the same 3D point?

Feature Matching

Correspondence Problem - Recap: Stereo System

In a calibrated stereo system the corresponding points between left and right image can be found on parallel lines. The distance between corresponding points is called disparity.

If the image patches of the projected points look all the same the correspondences between left and right image are ambiguous.

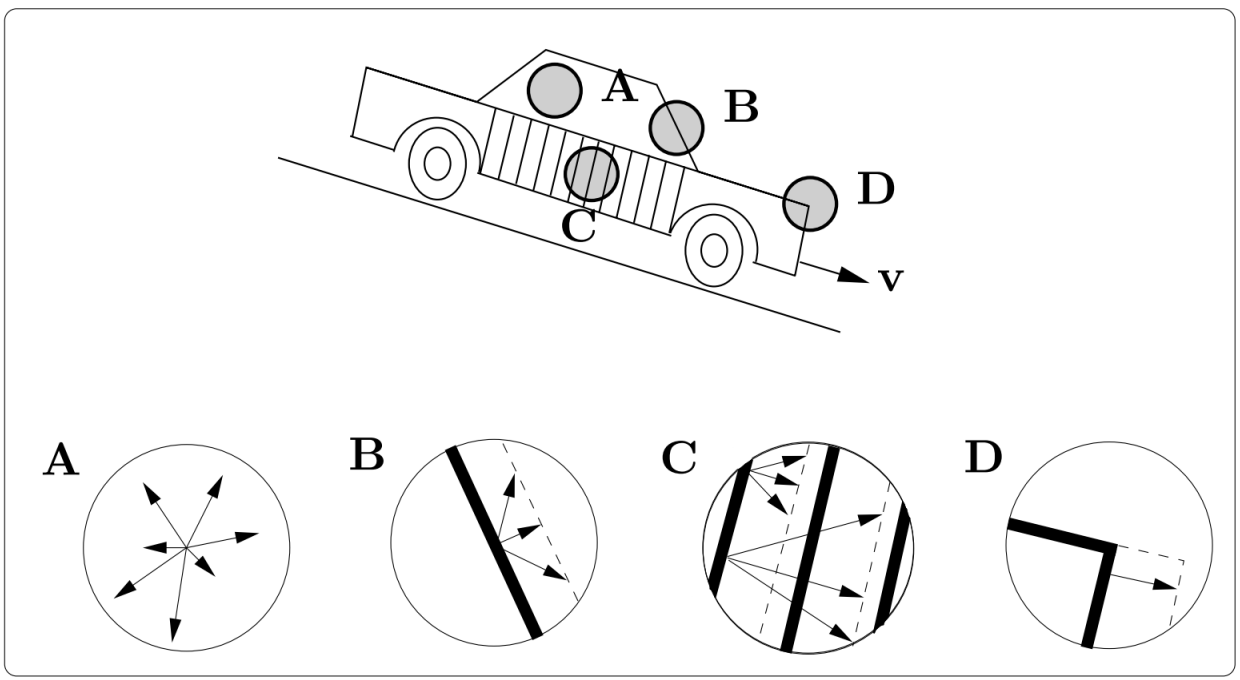


Feature Matching

Correspondence Problem - Aperture Problem

Technical University of Applied Sciences Würzburg-Schweinfurt

Stereoscopic reconstruction from small baselines or a moving camera is also called motion parallax. In this context the correspondence problem is also called aperture problem.

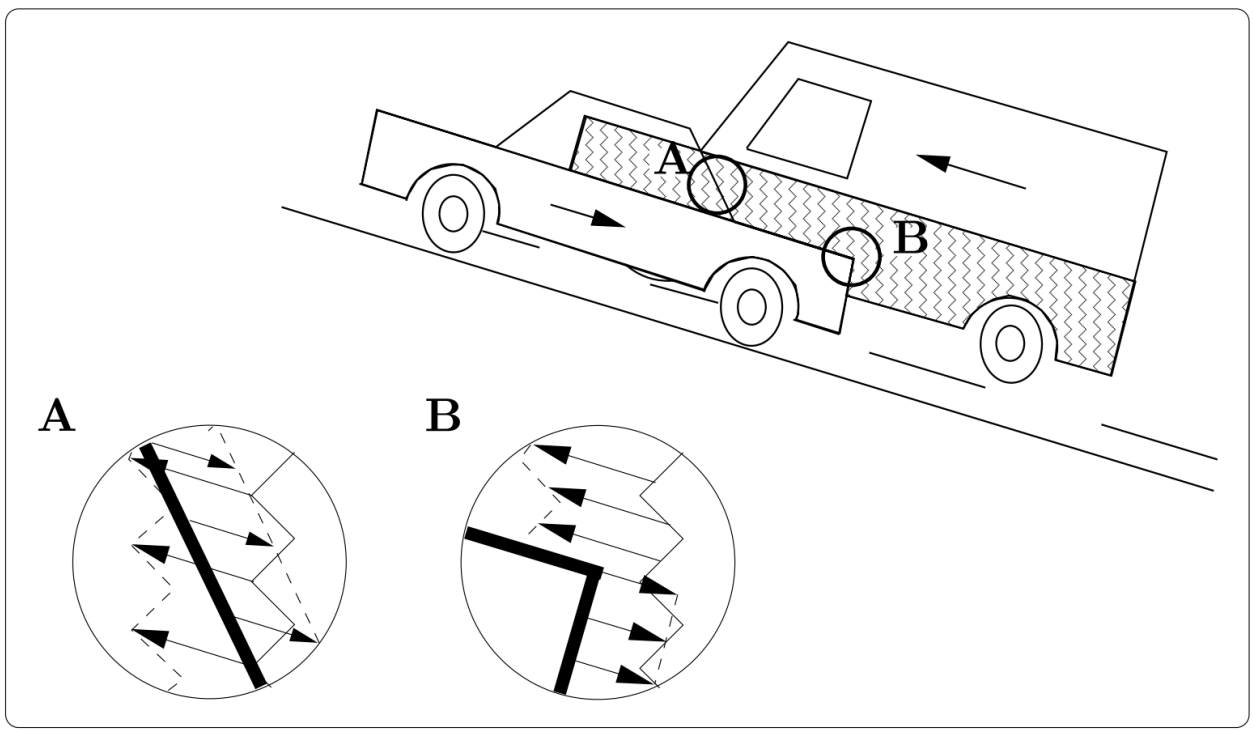


Feature Matching

Correspondence Problem - Aperture Problem

Technical University of Applied Sciences Würzburg-Schweinfurt

Stereoscopic reconstruction from small baselines or a moving camera is also called motion parallax. In this context the correspondence problem is also called aperture problem.



Feature Matching

Correspondence Problem - Challenges

Technical University of Applied Sciences Würzburg-Schweinfurt

Problem Given an image point in some image, what is the corresponding point in another image, which is the projection of the same 3D point?

challenges

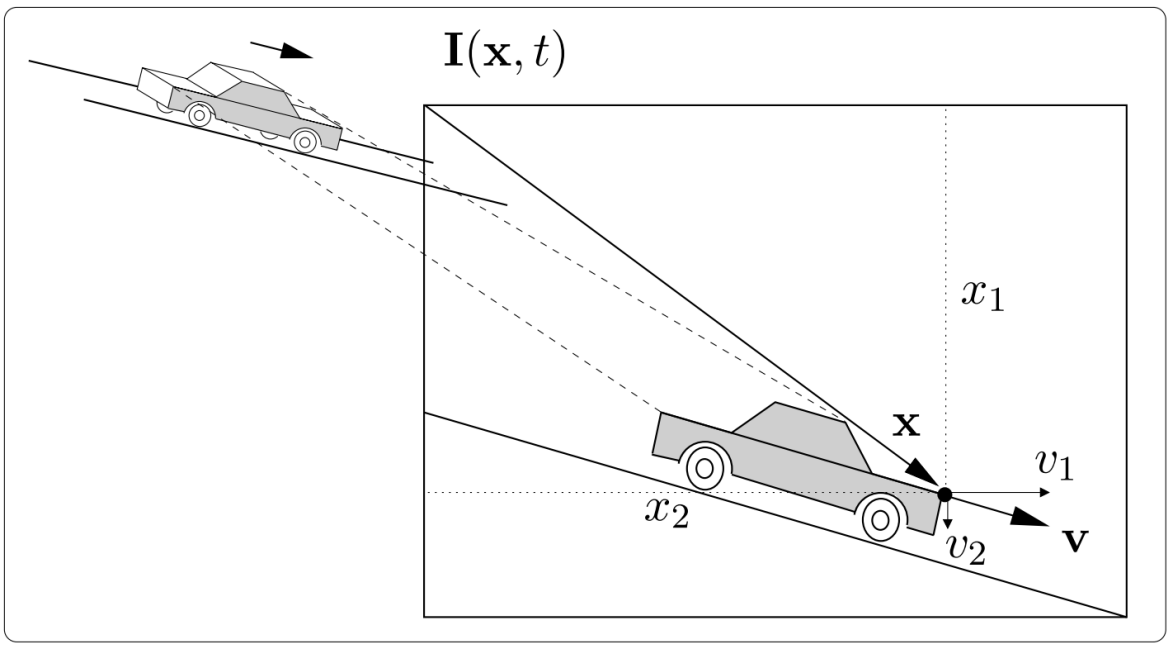
- How to find the right correspondence?
- How to measure the quality of the correspondence?
- How to tackle ambiguities?
- How to deal with illumination changes?
- How to tackle perspective distortions because of strongly different points of view onto the same point?
- How to deal with large scale differences?

Feature Matching

Different Baselines

Technical University of Applied Sciences Würzburg-Schweinfurt

The baseline can be large or small which results in larger or smaller displacements for the same distance. Hence, we can use differential linearized approaches for small disparities and for subpixel measurements. All correspondences form a vector field that is also called motion field.

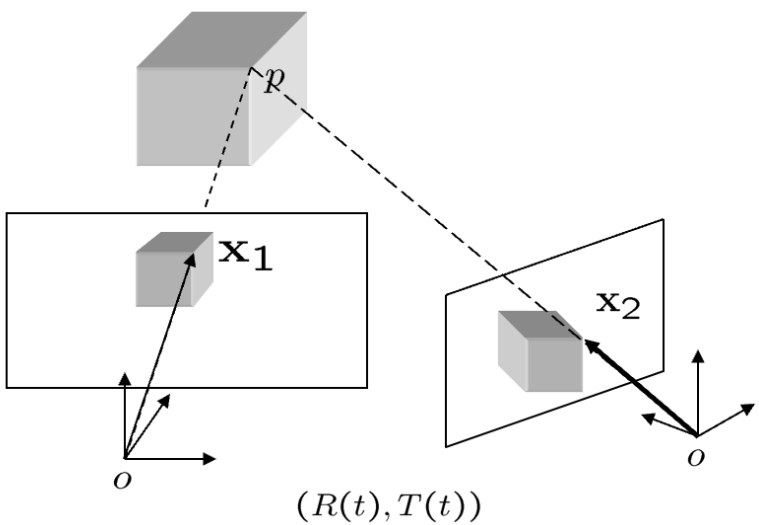


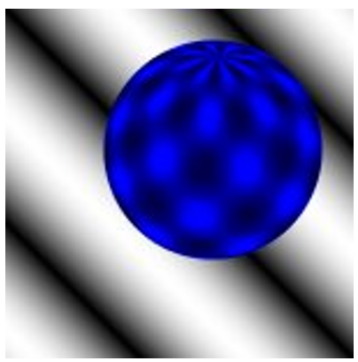
Feature Matching

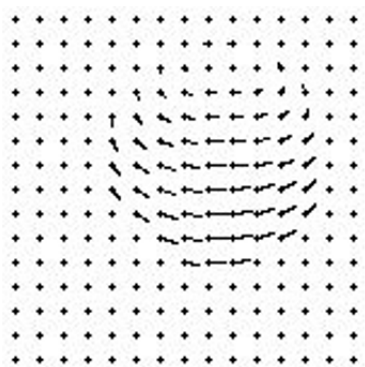
Different Baselines

Technical University of Applied Sciences Würzburg-Schweinfurt

The baseline can be large or small which results in larger or smaller displacements for the same distance. Hence, we can use differential linearized approaches for small disparities and for subpixel measurements. All correspondences form a vector field that is also called motion field.
To measure the motion field optical flow methods are used.







Feature Matching

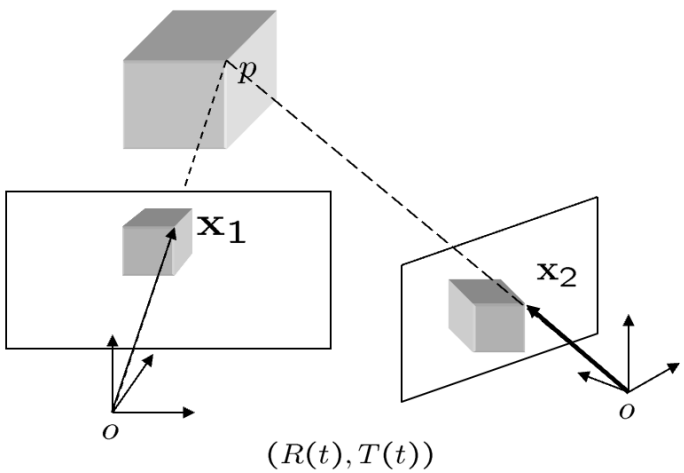
Different Scales

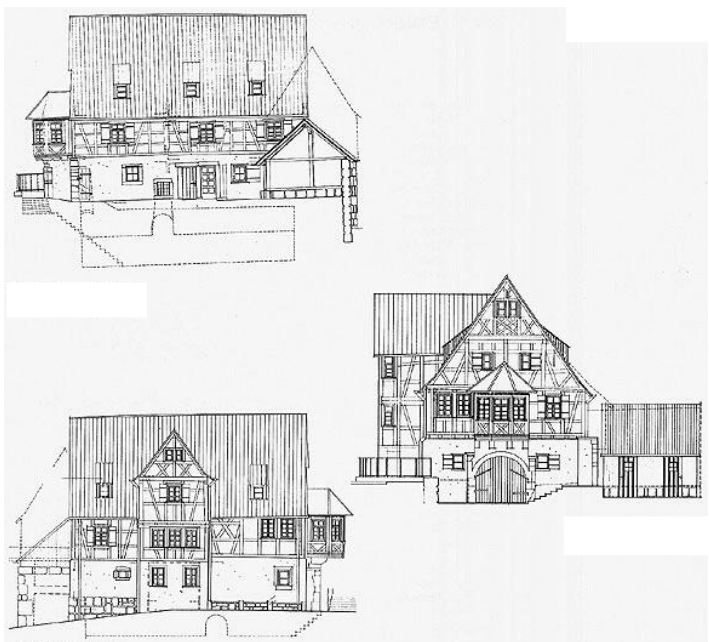
The projected points and its surrounding neighborhood can have same or different scale which results in corresponding ROIs (Region of Interest) of different size.

Feature Matching

Strongly Different Points of View

The point of view can be quite different which results in strong perspective distortions of ROIs or completely different perspectives.





Feature Matching

Ambiguous Image Patches

Technical University of Applied Sciences Würzburg-Schweinfurt

Local image patches can look quite similar and it is hard to distinguish between them. This often happens for repetitive patterns and textures.

Feature Matching

Illumination changes

The same local 3D region can be illuminated quite differently on images from different points of view.

Feature Matching

Correspondence Problem

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

Useful Constraints To find correspondences and reduce the ambiguities the following assumptions can be made:

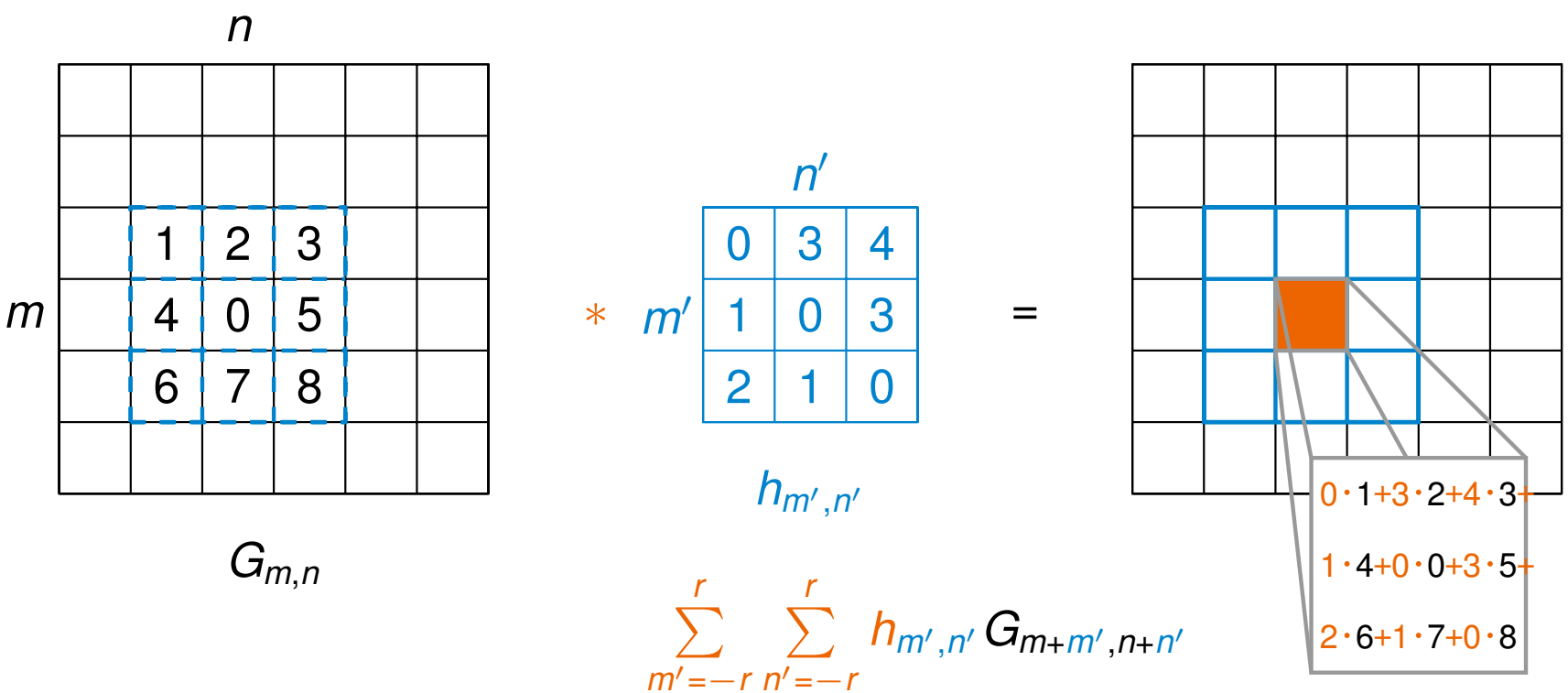
Assumptions

- Brightness constancy assumption
- Rigid body motion assumption: The projected points belong to rigid bodies. Hence, the deformation of the image patches follows from knowledge/assumption about the surface geometry, e.g. planar rigid object parts.
- Small displacement assumption: small disparities/displacements are more probable then larger ones.
- Smoothness assumption: neighboring displacements are similar.

Feature Matching

Recap: Discrete 2D Correlation

A 2D correlation with an arbitrary mask compares image patches with each other. The comparison measure corresponds to the scalar product.



$$\sum_{m'=-r}^{r} \sum_{n'=-r}^{r} h_{m',n'} G_{m+m',n+n'}$$

Feature Matching

Correlation - Autocorrelation

Technical University of Applied Sciences Würzburg-Schweinfurt

In order to correlate gray values at different positions in the image correlation functions are used.

Autocorrelation function: Measurement of the self-similarity of a signal.

expectation value: $c_{GG}(m, n; m', n') = \mathbb{E}\{\mathbf{G}_{m,n}\mathbf{G}_{m',n'}\}$,

correlation: $\hat{c}_{GG}(\Delta m, \Delta n) = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} G_{m,n}G_{m+\Delta m,n+\Delta n}$.

Feature Matching

Correlation - Autocovariance

Mean-free measurement of the self-similarity of a signal. Now the self-similarity is invariant to different brightness values between the image patches.

Mean value $\overline{G} = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} G_{m,n}$

$$\begin{aligned} \hat{s}_{GG}(\Delta m, \Delta n) &= \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (G_{m,n} - \overline{G})(G_{m+\Delta m, n+\Delta n} - \overline{G}) \\ &= \hat{c}_{GG}(\Delta m, \Delta n) - \overline{G}^2 \quad \text{if } \mathbf{G} \text{ is cyclic!} \end{aligned}$$

Feature Matching

Correlation - Normalized Autocovariance

Technical University of Applied Sciences Würzburg-Schweinfurt

Mean-free and variance normalized measurement of the self-similarity of a signal. Now the self-similarity is invariant to different brightness and contrast values between the image patches.

$$\hat{r}_{GG}(\Delta m, \Delta n) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left(\frac{G_{m,n} - \overline{G}}{\sqrt{\tilde{G}}} \right) \left(\frac{G_{m+\Delta m, n+\Delta n} - \overline{G}}{\sqrt{\tilde{G}}} \right)$$

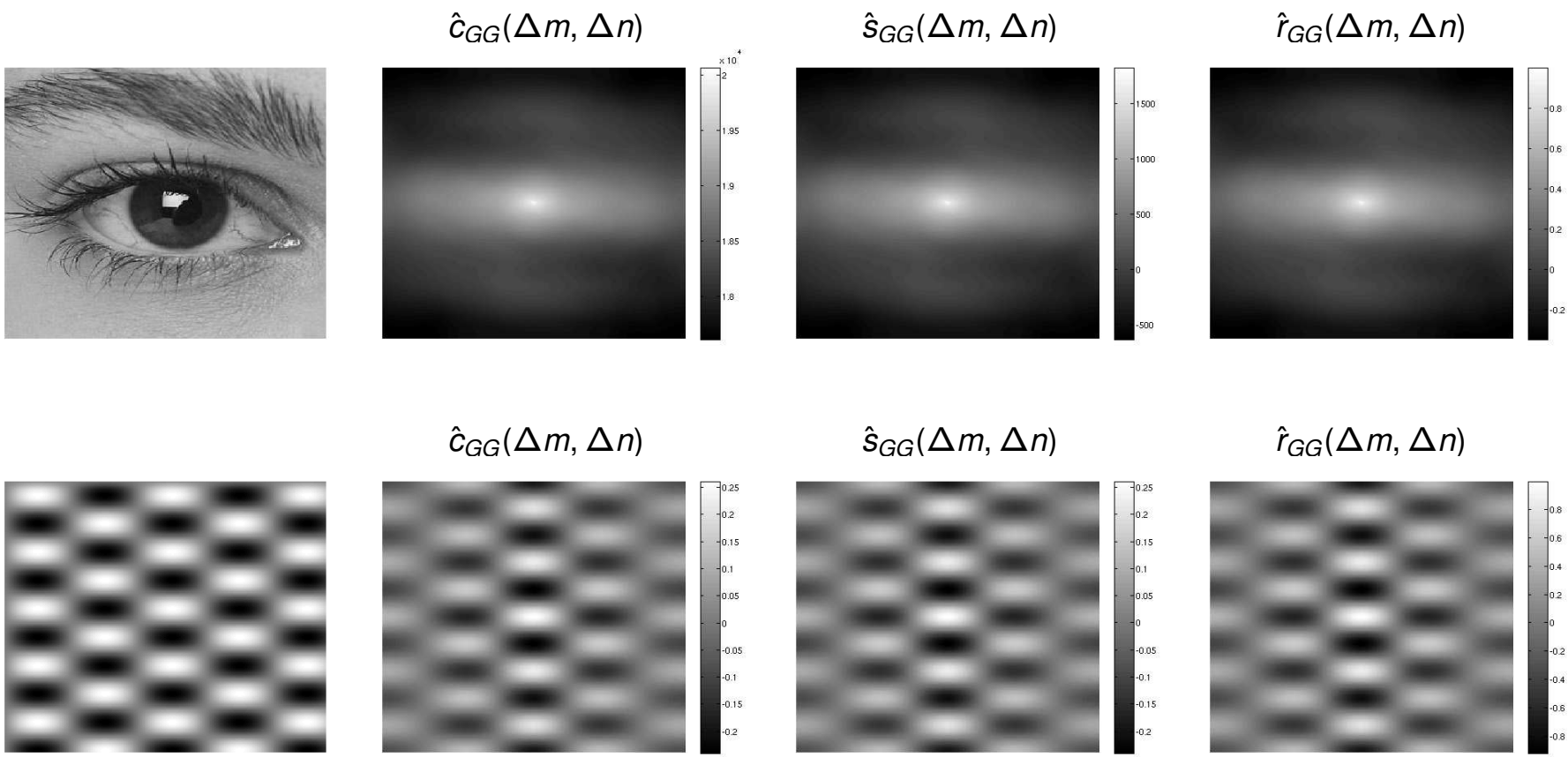
$$= \frac{\hat{c}_{GG}(\Delta m, \Delta n) - \overline{G}^2}{\tilde{G}} \quad \text{if } \mathbf{G} \text{ is cyclic!}$$

With mean value and variance:

$$\overline{G} = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} G_{m,n}\,, \qquad \tilde{G} = \frac{1}{(M-1)(N-1)} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (G_{m,n} - \overline{G})^2\,.$$

Feature Matching

Correlation - Example Autocovariance



Feature Matching

Correlation - Cross Correlation

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

The comparison of two different images at different positions is also called matching or correspondence analysis.

Cross correlation: Measurement of the similarity of two images.

$$\text{Expectation value:} \quad c_{GG'}(m,n;m',n') = \mathbb{E}\{\mathbf{G}_{m,n}\mathbf{G}'_{m',n'}\},$$

$$\text{Cross correlation:} \quad \hat{c}_{GG'}(\Delta m,\Delta n) = \frac{1}{MN}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} G_{m,n}G'_{m+\Delta m,n+\Delta n}.$$

Feature Matching

Correlation - Cross Covariance

Mean-free measurement of similarity between two signals that is invariant to differences in brightness.

$$\hat{s}_{GG'}(\Delta m, \Delta n) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (G_{m,n} - \overline{G})(G'_{m+\Delta m, n+\Delta n} - \overline{G}')$$

$$= \hat{c}_{GG'}(\Delta m, \Delta n) - \overline{G}\,\overline{G}' \quad \text{if } \mathbf{G}, \mathbf{G}' \text{ are cyclic!}$$

Feature Matching

Correlation - Normalized Cross Covariance

Technical University of Applied Sciences Würzburg-Schweinfurt

Mean-free and variance normalized measurement of similarity between two signals that is invariant to differences in brightness and contrast.

$$\hat{r}_{GG'}(\Delta m, \Delta n) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left(\frac{G_{m,n} - \overline{G}}{\sqrt{\tilde{G}}} \right) \left(\frac{G'_{m+\Delta m, n+\Delta n} - \overline{G}'}{\sqrt{\tilde{G}'}} \right)$$

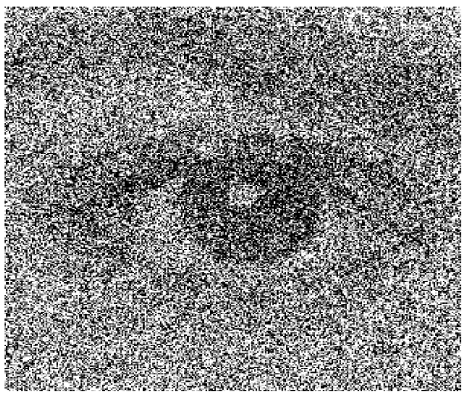
$$= \frac{\hat{c}_{GG'}(\Delta m, \Delta n) - \overline{G}\,\overline{G}'}{\sqrt{\tilde{G}}\sqrt{\tilde{G}'}} \quad \text{if } \mathbf{G}, \mathbf{G}' \text{ is cyclic!}$$

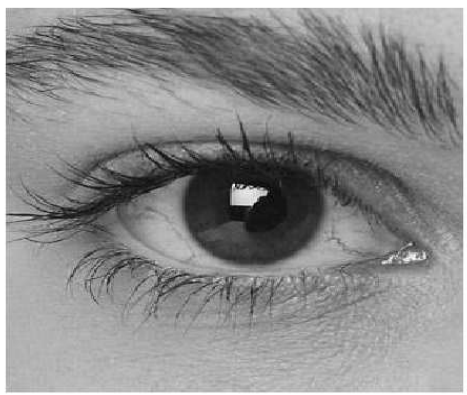
Feature Matching

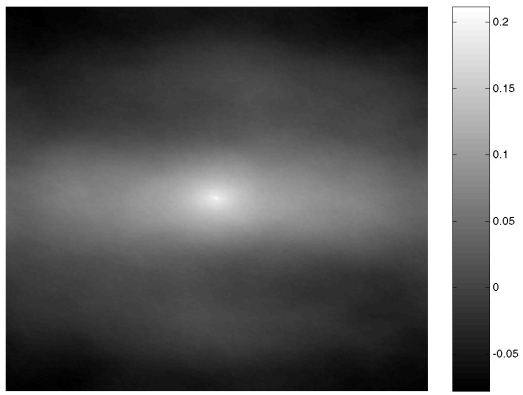
Correlation - Normalized Cross Covariance

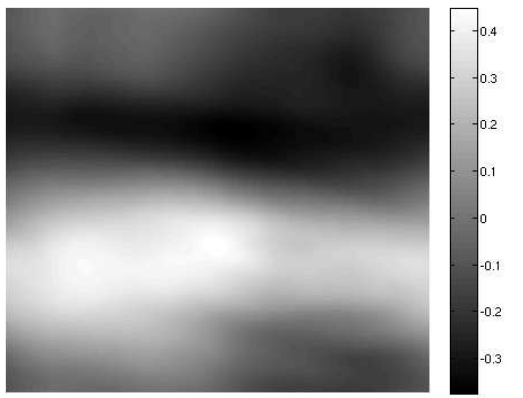
Technical University of Applied Sciences Würzburg-Schweinfurt

$\hat{r}_{GG'}(\Delta m, \Delta n)$









Feature Matching

Correlation - Local Cross Correlation

Technical University of Applied Sciences Würzburg-Schweinfurt

Given knowledge of a prototype pattern or template **T**, the maximum of the local cross correlation function between the pattern **T** of size $K \times L$ and an image **G** of size $M \times N$ yields the location of the most likely occurrence of the pattern in the image. It usually holds: $K \times L \ll M \times N$.

Application: detection of patterns by threshold operation on the cross correlation.

Advantage: Theoretically optimal solution.

Disadvantage: rotation and scaling not covered, high computational cost (one correlation per prototype).

Feature Matching

Local Cross Correlation - Challenges

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

View point changes of the patch

- large translation
- image-plane rotation (patch rotates about z-axis)
- out-of-plane rotation (projected 3D object rotates arbitrarily)
- scale changes (size of projection changes)

Scene changes

- illumination
- clutter
- occlusion
- noise

Feature Matching

Correlation - Local Cross Correlation

Technical University of Applied Sciences Würzburg-Schweinfurt

Local correlations are dependent on location, as local correlation functions are calculated.

Local cross correlation function

$$\hat{c}_{TG}(\Delta m, \Delta n) = \frac{1}{KL} \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} T_{m,n} G_{m+\Delta m, n+\Delta n} \,.$$

Local cross covariance function

$$\begin{aligned} \hat{s}_{TG}(\Delta m, \Delta n) &= \frac{1}{KL} \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} (T_{m,n} - \overline{T})(G_{m+\Delta m, n+\Delta n} - \overline{G}_{\Delta m, \Delta n}) \\ &= \hat{c}_{TG}(\Delta m, \Delta n) - \overline{T}\,\overline{G}_{\Delta m, \Delta n} \,. \end{aligned}$$

Feature Matching

Correlation - Local Cross Correlation

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

Local variance normalized cross covariance function

$$\hat{r}_{TG}(\Delta m, \Delta n) = \frac{\hat{s}_{TG}(\Delta m, \Delta n)}{\sqrt{\tilde{T}\,\tilde{G}_{\Delta m,\Delta n}}} = \frac{\hat{c}_{TG}(\Delta m, \Delta n) - \overline{T}\,\overline{G}_{\Delta m,\Delta n}}{\sqrt{\tilde{T}\,\tilde{G}_{\Delta m,\Delta n}}} .$$

$$\overline{G}_{\Delta m,\Delta n} = \frac{1}{KL}\sum_{m=0}^{K-1}\sum_{n=0}^{L-1} G_{m+\Delta m,n+\Delta n} , \quad \tilde{G}_{\Delta m,\Delta n} = \frac{1}{(K-1)(L-1)}\sum_{m=0}^{K-1}\sum_{n=0}^{L-1}(G_{m+\Delta m,n+\Delta n} - \overline{G}_{\Delta m,\Delta n})^2 .$$

$$\overline{T} = \frac{1}{KL}\sum_{m=0}^{K-1}\sum_{n=0}^{L-1} T_{m,n} , \quad \tilde{T} = \frac{1}{(K-1)(L-1)}\sum_{m=0}^{K-1}\sum_{n=0}^{L-1}(T_{m,n} - \overline{T})^2 .$$

Feature Matching

Correlation - Local Cross Correlation

Alternative correlation measures that are predominantly used for disparity calculation in stereo vision.

SSD Matching (Sum of Squared Differences)

$$\begin{aligned} SSD_{TG}(\Delta m, \Delta n) &= \frac{1}{KL} \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} (T_{m,n} - G_{m+\Delta m, n+\Delta n})^2 \\ &= \overline{G^2}_{\Delta m, \Delta n} - 2\hat{c}_{TG}(\Delta m, \Delta n) + \overline{T^2}\,. \end{aligned}$$

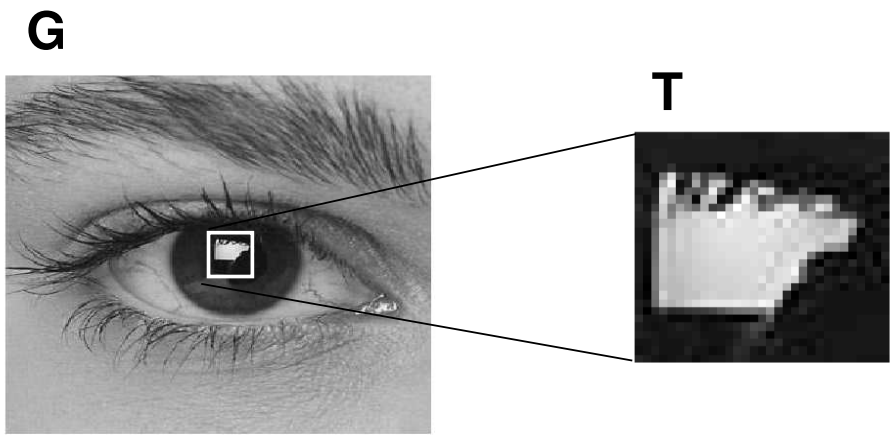
SAD Matching (Sum of Absolute Differences)

$$SAD_{TG}(\Delta m, \Delta n) = \frac{1}{KL} \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} \left| T_{m,n} - G_{m+\Delta m, n+\Delta n} \right|.$$

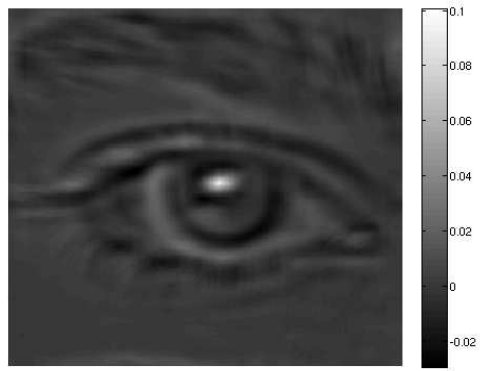
Feature Matching

Correlation - Comparison

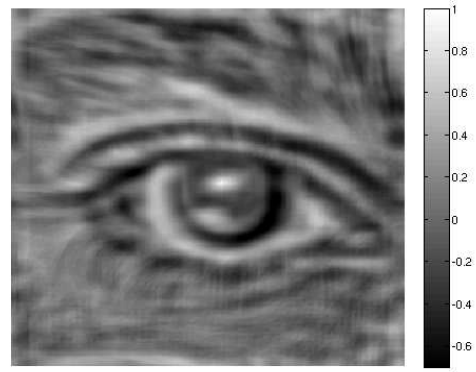
Technical University of Applied Sciences Würzburg-Schweinfurt



$\hat{c}_{TG}(\Delta m, \Delta n)$



$\hat{s}_{TG}(\Delta m, \Delta n)$

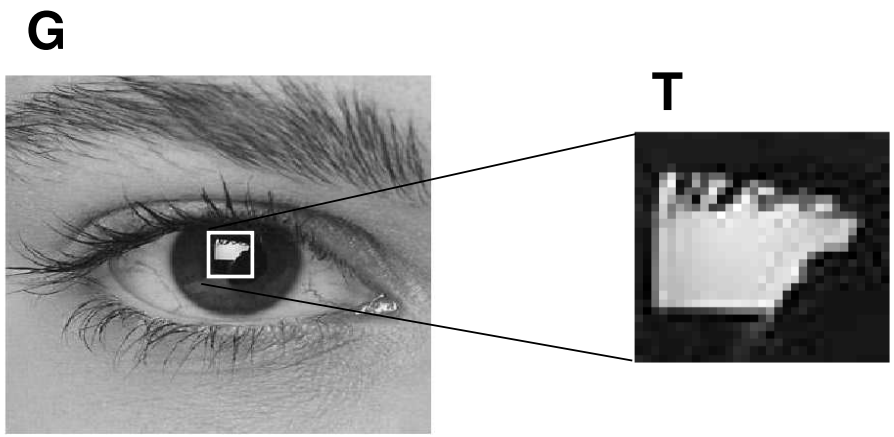


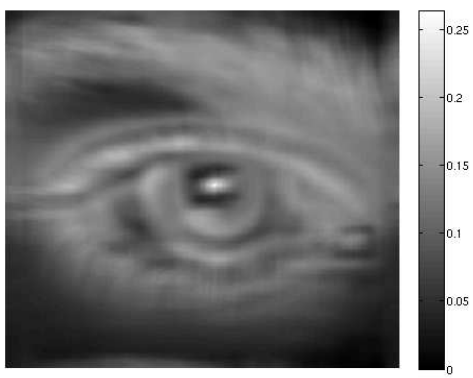
$\hat{r}_{TG}(\Delta m, \Delta n)$

Feature Matching

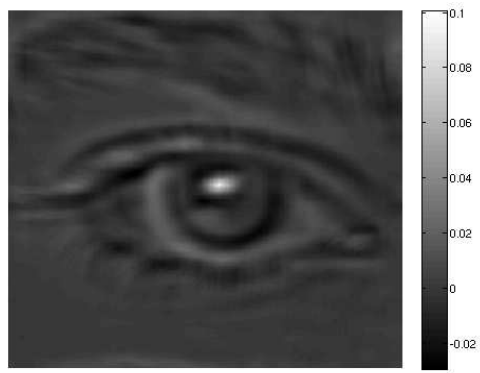
Correlation - Comparison

Technical University of Applied Sciences Würzburg-Schweinfurt

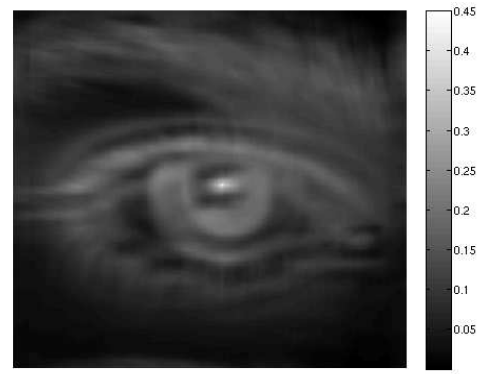




$SSD_{TG}(\Delta m, \Delta n)$
(inverse)



$\hat{s}_{TG}(\Delta m, \Delta n)$



$SAD_{TG}(\Delta m, \Delta n)$
(inverse)

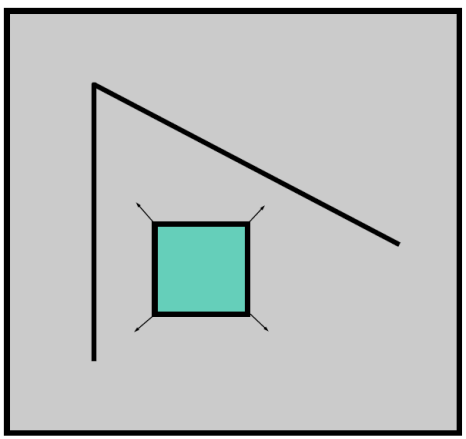
Feature Matching

Interest Points

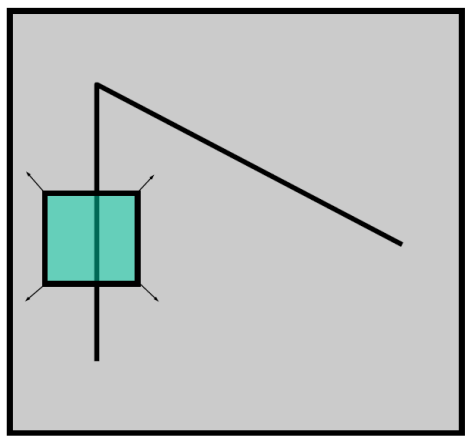
Technical University of Applied Sciences Würzburg-Schweinfurt

To reduce ambiguity we can also first search for so called interest points. These points are easy to detect and do not suffer from the aperture problem.

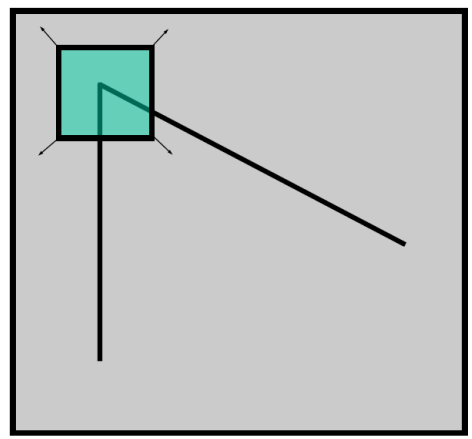
Very simple and robust interest point detectors are corner detectors like the Harris-Corner detector that are based on a local structure analysis.



"flat" region: no change in all directions



"edge": no change along the edge direction



"corner": significant change in all directions

Feature Matching

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

Interest Points - Recap: Structure Tensor

The structure tensor $\mathbf{J}(\mathbf{x})$ is given for a certain neighborhood $\mathcal{N}(\mathbf{x})$ with a certain weight $W(\mathbf{x}-\mathbf{x}')$ around the point $\mathbf{x}$ and defined as follows:

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} J_{11}(\mathbf{x}) & J_{12}(\mathbf{x}) \\ J_{12}(\mathbf{x}) & J_{22}(\mathbf{x}) \end{bmatrix} = \int\limits_{\mathbf{x}' \in \mathcal{N}(\mathbf{x})} W(\mathbf{x}-\mathbf{x}') \left(\nabla(\mathbf{x}') \nabla(\mathbf{x}')^\top \right) d\mathbf{x}' = \mathbf{W} * \left(\nabla \nabla^\top \right).$$

This corresponds in 2D to a symmetric 2×2 matrix, where the individual elements are weighted averages of the derivative combinations $(\partial G(\mathbf{x})/\partial x)^2$, $(\partial G(\mathbf{x})/\partial y)^2$ and $(\partial G(\mathbf{x})/\partial x)(\partial G(\mathbf{x})/\partial y)$:

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} \sum\limits_{\mathbf{x}' \in \mathcal{N}(\mathbf{x})} W(\mathbf{x}-\mathbf{x}') G_x^2(\mathbf{x}') & \sum\limits_{\mathbf{x}' \in \mathcal{N}(\mathbf{x})} W(\mathbf{x}-\mathbf{x}') G_x(\mathbf{x}') G_y(\mathbf{x}') \\ \sum\limits_{\mathbf{x}' \in \mathcal{N}(\mathbf{x})} W(\mathbf{x}-\mathbf{x}') G_y(\mathbf{x}') G_x(\mathbf{x}') & \sum\limits_{\mathbf{x}' \in \mathcal{N}(\mathbf{x})} W(\mathbf{x}-\mathbf{x}') G_y^2(\mathbf{x}') \end{bmatrix}.$$

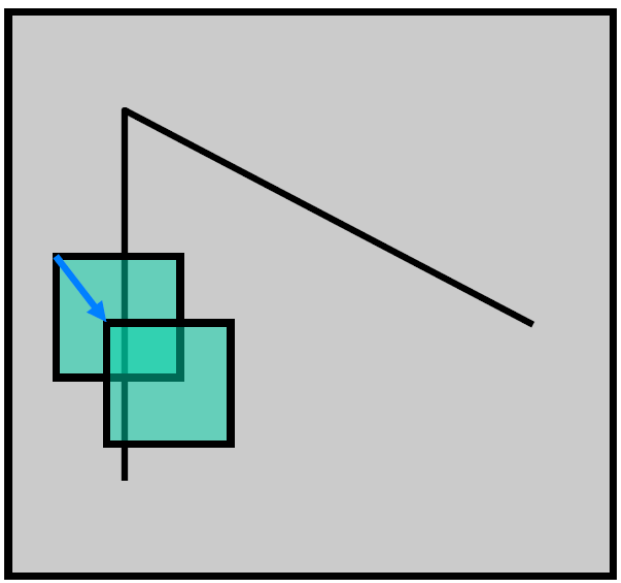
Feature Matching

Relation between SSD and Structure Tensor

Technical University of Applied Sciences Würzburg-Schweinfurt

We can also use SSD-Matching to measure the self-similarity of an image patch:

$$\mathrm{SSD}_{GG}(\Delta m, \Delta n) = \frac{1}{KL} \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} (G_{m+\Delta m, n+\Delta n} - G_{m,n})^2$$



Feature Matching

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

Relation between SSD and Structure Tensor

We can also use SSD-Matching to measure the self-similarity of an image patch:

$$\mathrm{SSD}_{GG}(\Delta m, \Delta n) = \frac{1}{KL} \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} (G_{m+\Delta m, n+\Delta n} - G_{m,n})^2 \,.$$

Approximating the shifted patch $G_{m+\Delta m, n+\Delta n}$ via 1st order Taylor expansion

$$G_{m+\Delta m, n+\Delta n} \approx G_{m,n} + \Delta m \frac{\partial}{\partial m} G_{m,n} + \Delta n \frac{\partial}{\partial n} G_{m,n}$$

leads to

$$\mathrm{SSD}_{GG}(\Delta m, \Delta n) \approx \frac{1}{KL} \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} (\Delta m \frac{\partial}{\partial m} G_{m,n} + \Delta n \frac{\partial}{\partial n} G_{m,n})^2 \,.$$

Feature Matching

Relation between SSD and Structure Tensor

Technical University of Applied Sciences Würzburg-Schweinfurt

Hence, the relation between SSD-Matching and the Structure Tensor is given by:

$$\mathrm{SSD}_{GG}(\Delta m, \Delta n) = \frac{1}{KL} \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} (G_{m+\Delta m, n+\Delta n} - G_{m,n})^2$$

$$\approx$$

$$\frac{1}{KL} \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} \begin{bmatrix} \Delta m & \Delta n \end{bmatrix} \begin{bmatrix} \left(\frac{\partial}{\partial m} G_{m,n}\right)^2 & \left(\frac{\partial}{\partial m} G_{m,n}\right)\left(\frac{\partial}{\partial n} G_{m,n}\right) \\ \left(\frac{\partial}{\partial n} G_{m,n}\right)\left(\frac{\partial}{\partial m} G_{m,n}\right) & \left(\frac{\partial}{\partial n} G_{m,n}\right)^2 \end{bmatrix} \begin{bmatrix} \Delta m \\ \Delta n \end{bmatrix}$$

$$[\Delta m \quad \Delta n] \begin{bmatrix} \frac{1}{KL} \sum_{m,n} \left(\frac{\partial}{\partial m} G_{m,n}\right)^2 & \frac{1}{KL} \sum_{m,n} \left(\frac{\partial}{\partial m} G_{m,n}\right)\left(\frac{\partial}{\partial n} G_{m,n}\right) \\ \frac{1}{KL} \sum_{m,n} \left(\frac{\partial}{\partial n} G_{m,n}\right)\left(\frac{\partial}{\partial m} G_{m,n}\right) & \frac{1}{KL} \sum_{m,n} \left(\frac{\partial}{\partial n} G_{m,n}\right)^2 \end{bmatrix} \begin{bmatrix} \Delta m \\ \Delta n \end{bmatrix}$$

This relation holds for small subpixel displacements $(\Delta m, \Delta n)$.

Feature Matching

Interest Points - Recap: Coherence

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

The coherence is a measure for the isotropy of the gray value structure:

$$c(\mathbf{x}) = \frac{\sqrt{(J_{22} - J_{11})^2 + 4J_{12}^2}}{J_{11} + J_{22}} = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2},$$

and varies between zero and one $c \in [0; 1]$. A structure with an ideal orientation has the value one ($\lambda_2 = 0, \lambda_1 > 0$) and a completely isotropic gray value structure has the value zero ($\lambda_1 = \lambda_2 > 0$).

Thus, the eigenvalues characterize the structure:

- $\lambda_1 = \lambda_2 = 0$: The local environment is constant,
- $\lambda_1 > 0, \lambda_2 = 0$: Ideal orientation, all gradients are identically oriented.
- $\lambda_1 > 0, \lambda_2 > 0$: Gray values change in all directions.
- $\lambda_1 = \lambda_2 > 0$: Gray values change equally in all directions.

Feature Matching

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

Interest Points - Recap: Corner Detectors

From the eigenvalues and the relation to the coherence, a very simple corner detector can be defined: If the smallest eigenvalue exceeds a certain predefined threshold τ, then a corner point is present.

Shi-Tomasi corner detector

- $\min(\lambda_1, \lambda_2) > \tau$: corner point exists.

Another variation of this type of corner point detector was proposed by Harris and Stephens in 1988:

Harris corner detector

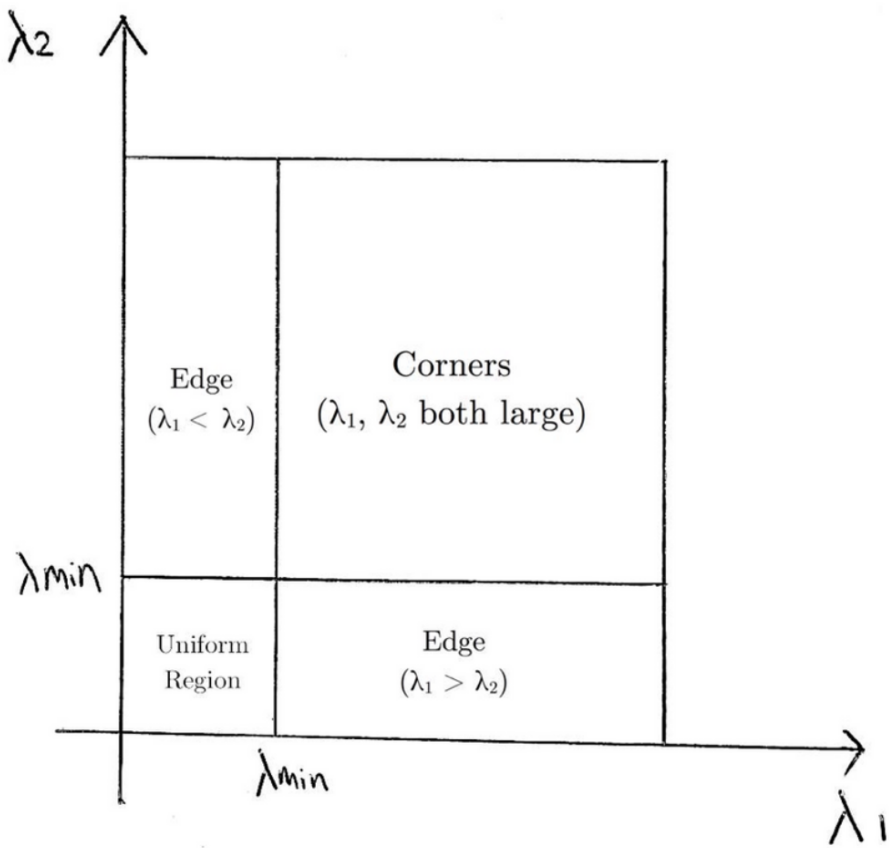
- $det(\mathbf{J}) + k \cdot trace^2(\mathbf{J}) = \lambda_1\lambda_2 - k(\lambda_1 + \lambda_2)^2 > \tau$: corner point exists.

With $k = [0.04; 0.06]$ the sensitivity to the gray level change can be adjusted along a preferred direction.

Feature Matching

Interest Points - Shi-Tomasi Corner Detector

- $\min(\lambda_1, \lambda_2) > \tau$: Corner is detected.

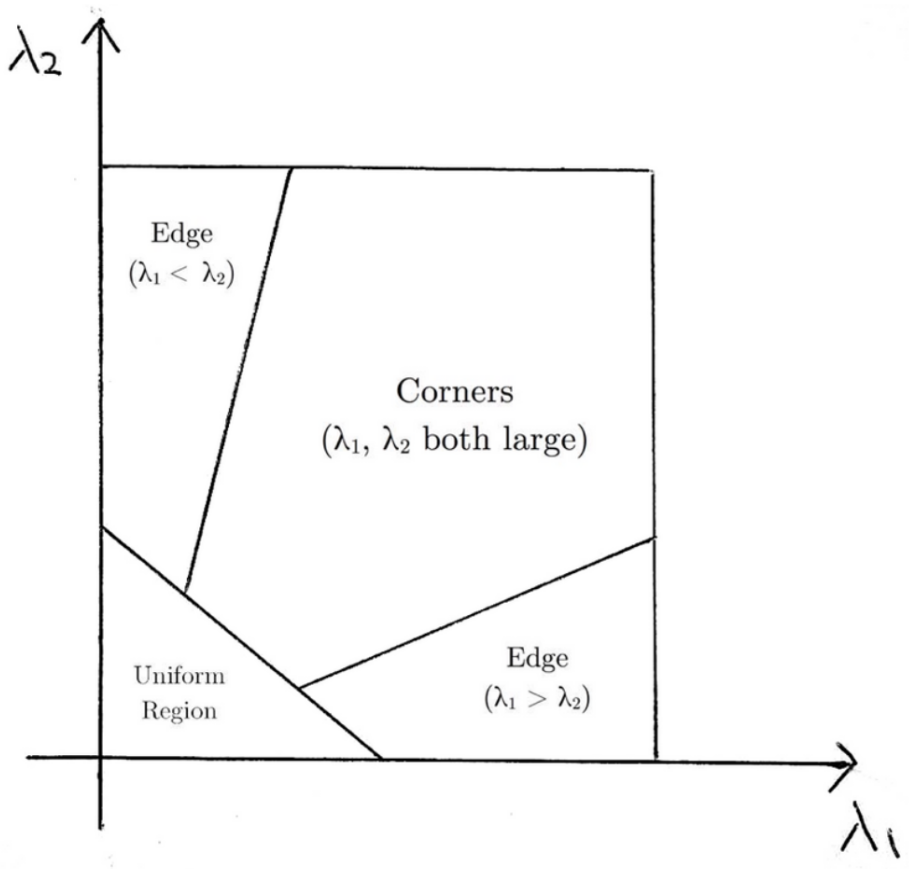


Feature Matching

Interest Points - Harris Corner Detector

Technical University of Applied Sciences Würzburg-Schweinfurt

- $det(\mathbf{J}) + k \cdot trace^2(\mathbf{J}) = \lambda_1\lambda_2 - k(\lambda_1 + \lambda_2)^2 > \tau$: Corner is detected.

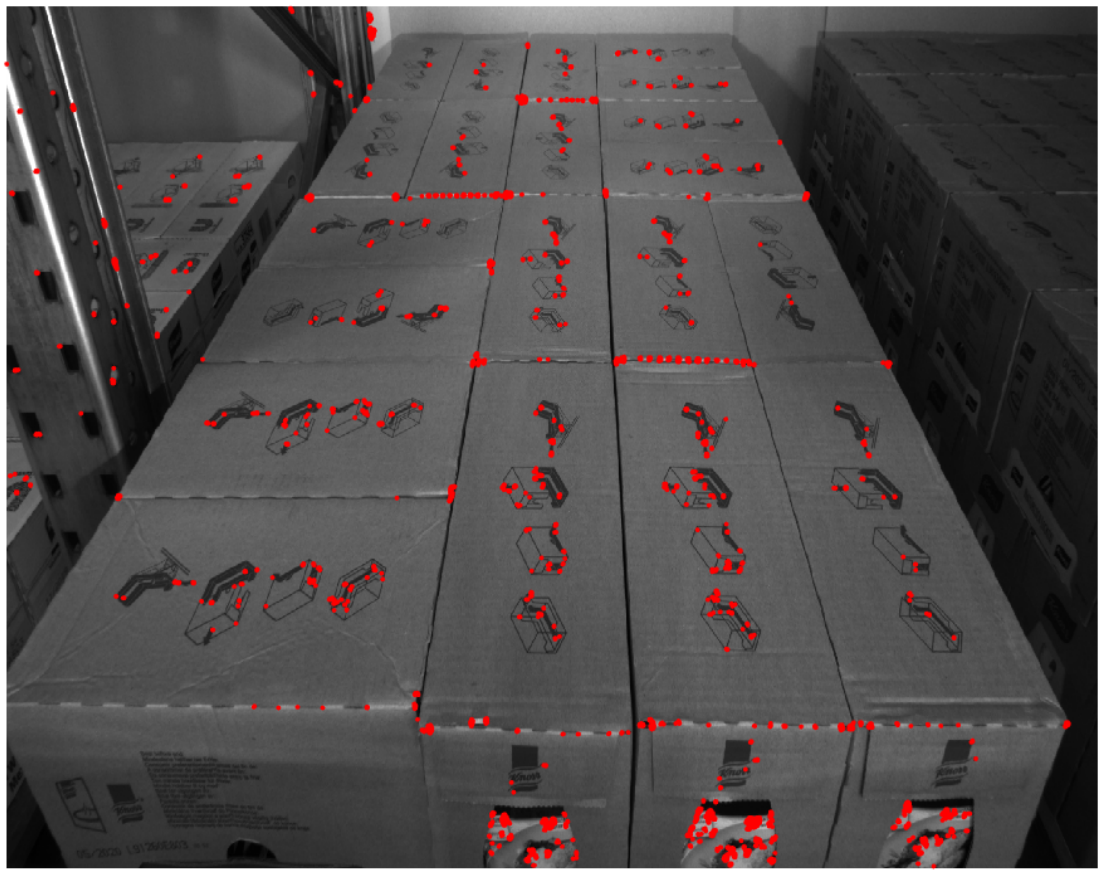


Feature Matching

Interest Points - Harris vs. Shi-Tomasi

Technical University of Applied Sciences Würzburg-Schweinfurt

Harris Corners

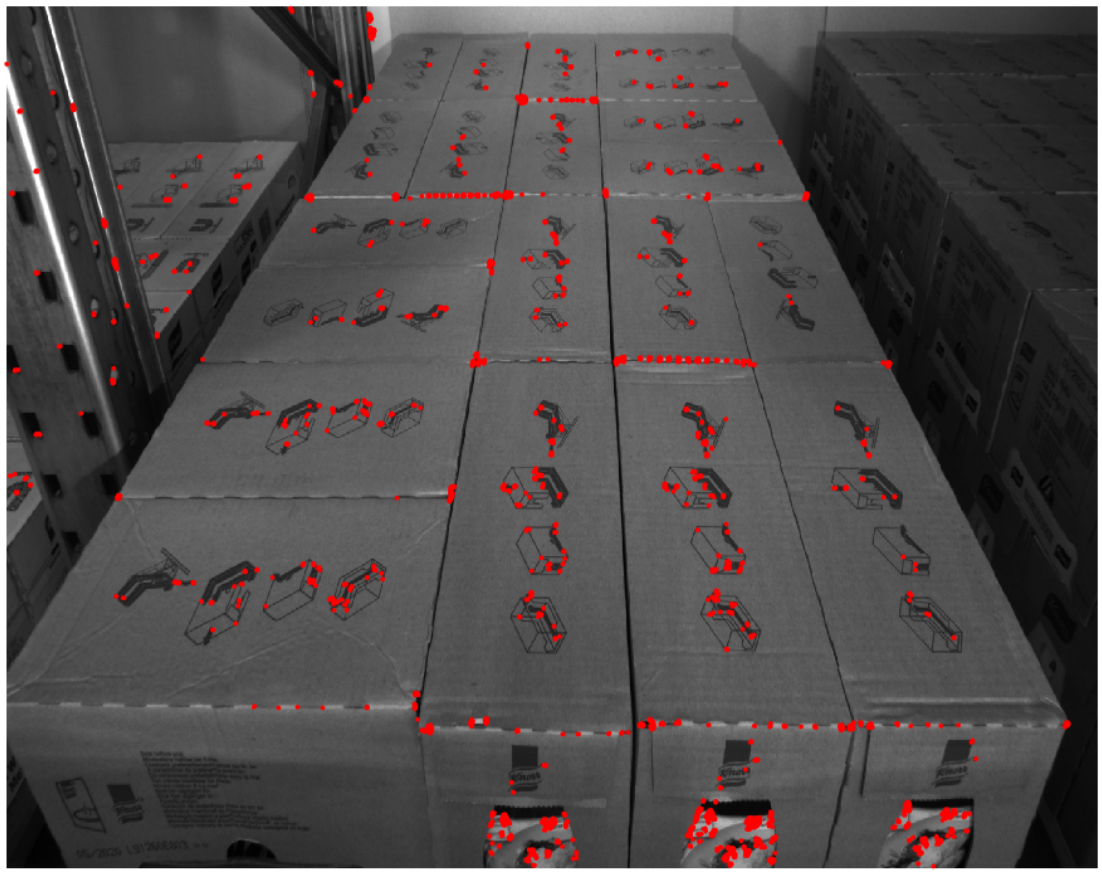


Feature Matching

Interest Points - Harris vs. Shi-Tomasi

Technical University of Applied Sciences Würzburg-Schweinfurt

Shi-Tomasi Corners



Feature Matching

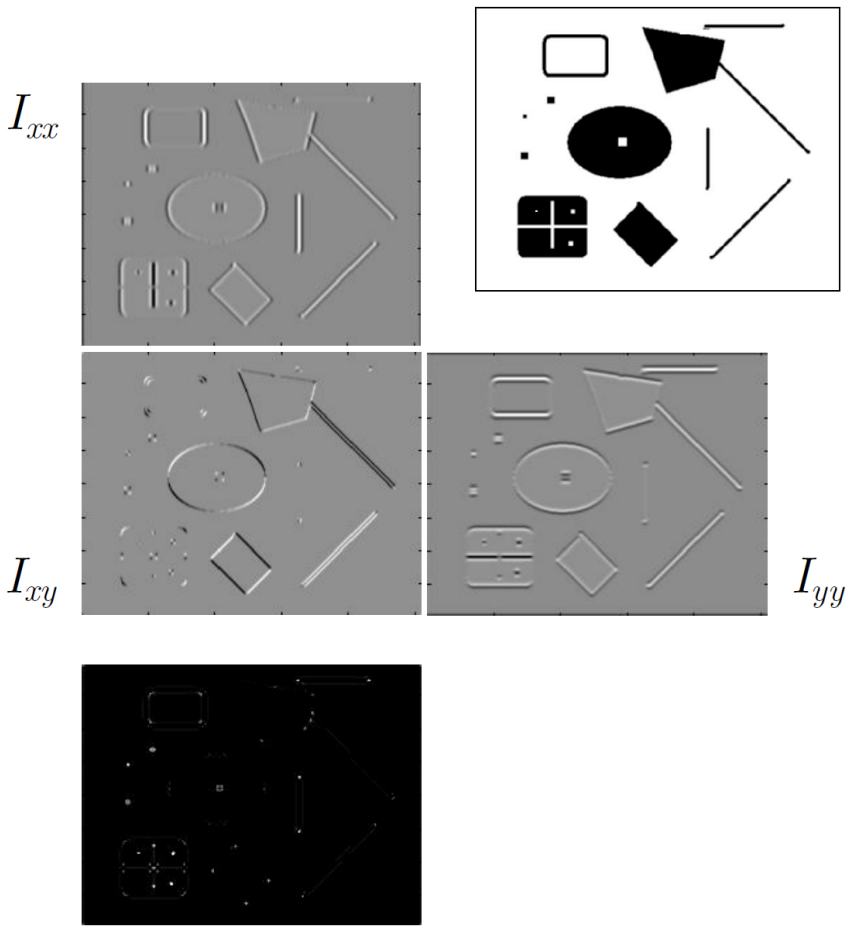
Interest Points - Hessian Determinant

The Hessian matrix is symmetric and holds all second order derivatives:

$$\mathbf{H} = \begin{bmatrix} G_{xx} & G_{xy} \\ G_{yx} & G_{yy} \end{bmatrix}$$

Interest Points can also be found by thresholding the determinant of the Hessian:

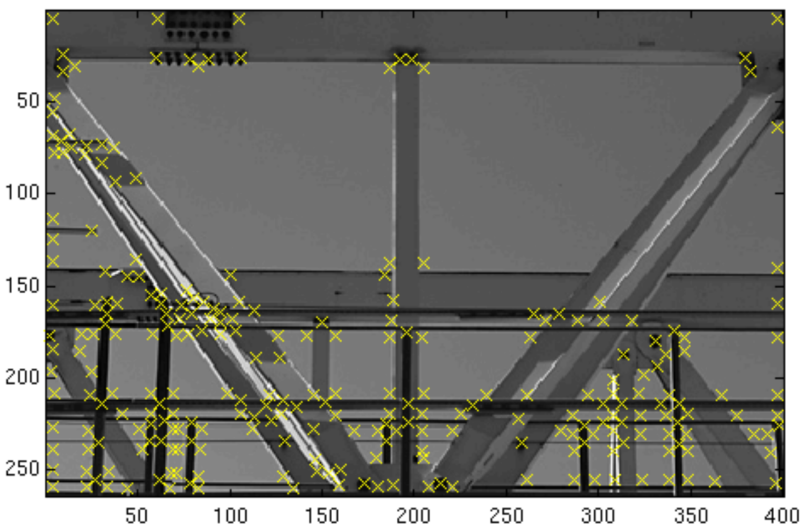
$$det(\mathbf{H}) = G_{xx}G_{yy} - G_{xy}^2 > \tau$$

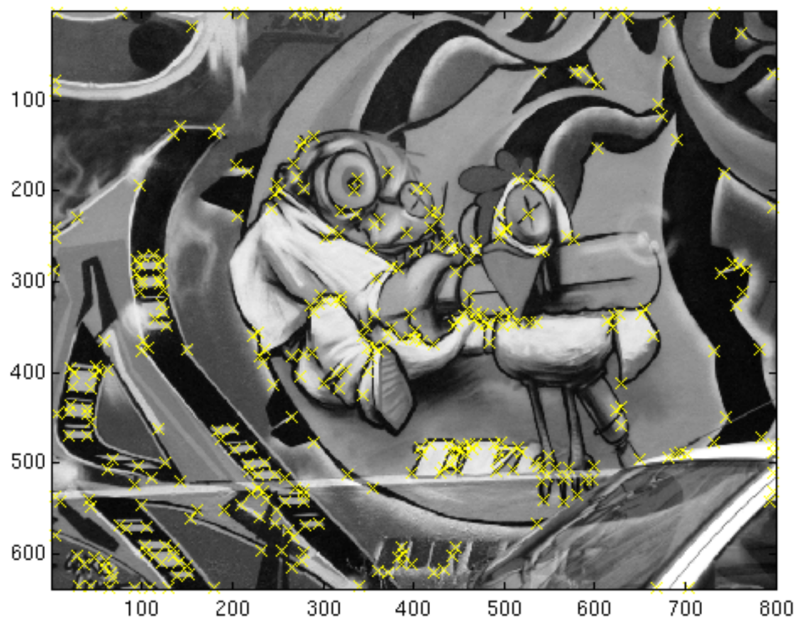


Feature Matching

Interest Points - Hessian Determinant

Technical University of Applied Sciences Würzburg-Schweinfurt





Feature Matching

Interest Points - Corner Detectors

Technical University of Applied Sciences Würzburg-Schweinfurt

Corner detectors are distinctive and invariant to rotations in the plane!

∇I

$\mathbf{R}^T \nabla I$

Feature Matching

Interest Points - Corner Detectors

Technical University of Applied Sciences Würzburg-Schweinfurt

Corner detectors are not invariant to scale, affine and projective transformations!

(1)

(2)

(3)

Feature Matching

Scale Invariant Interest Point Detection

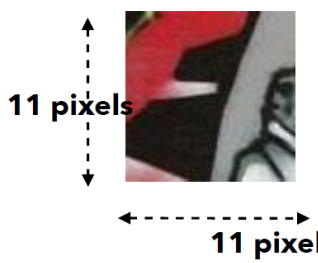
Find some interest points

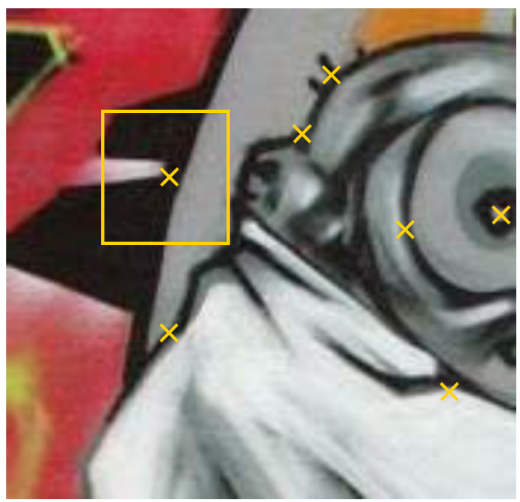
Feature Matching

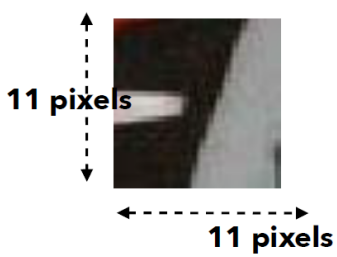
Scale Invariant Interest Point Detection

Technical University of Applied Sciences Würzburg-Schweinfurt

Extract patches



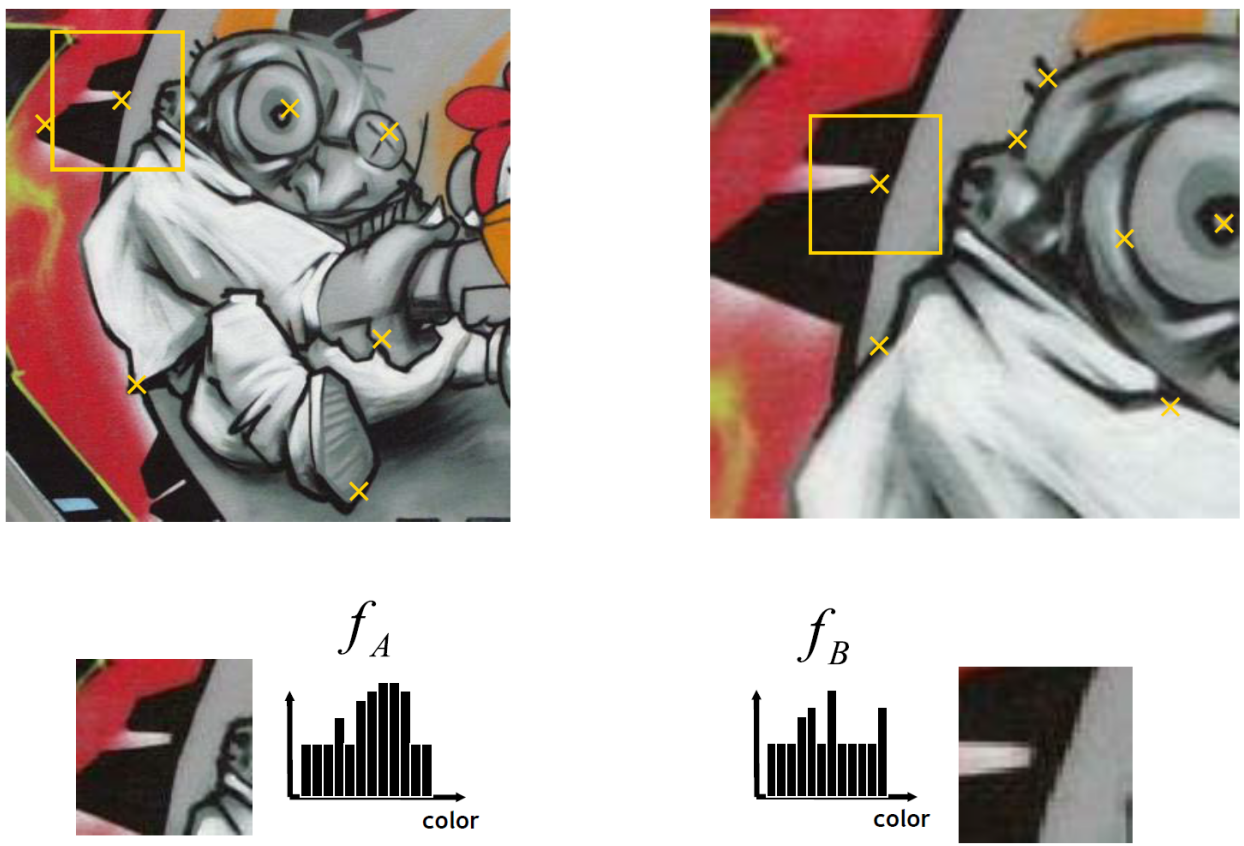




Feature Matching

Scale Invariant Interest Point Detection

Computing feature descriptors, e.g. color histograms

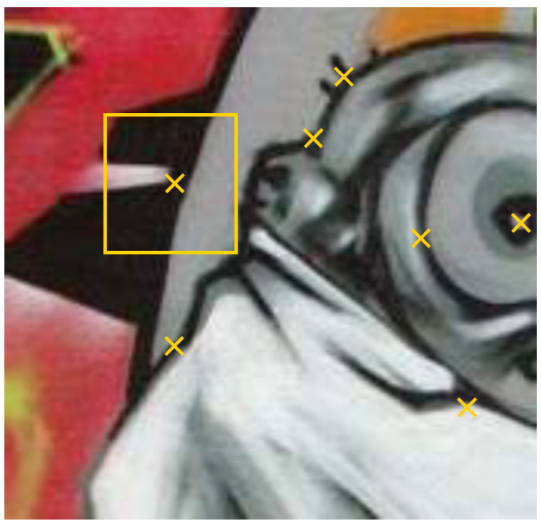


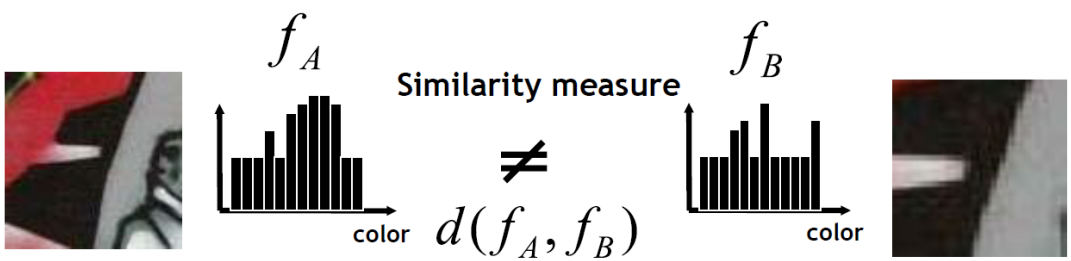
Feature Matching

Scale Invariant Interest Point Detection

Technical University of Applied Sciences Würzburg-Schweinfurt

Compare feature descriptors: The features are not similar due to different patch content



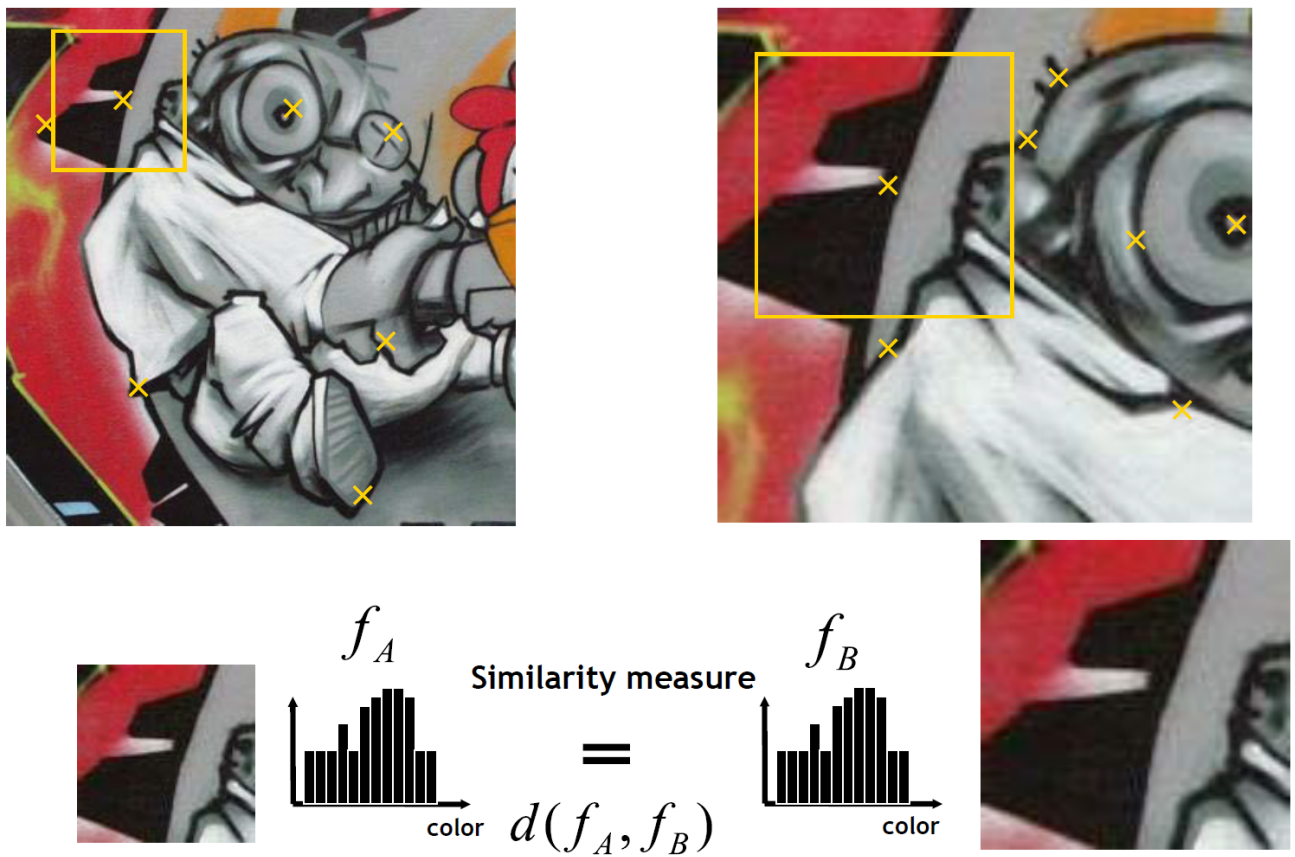


Feature Matching

Scale Invariant Interest Point Detection

Technical University of Applied Sciences Würzburg-Schweinfurt

Compare feature descriptors: The features are similar due to same patch content

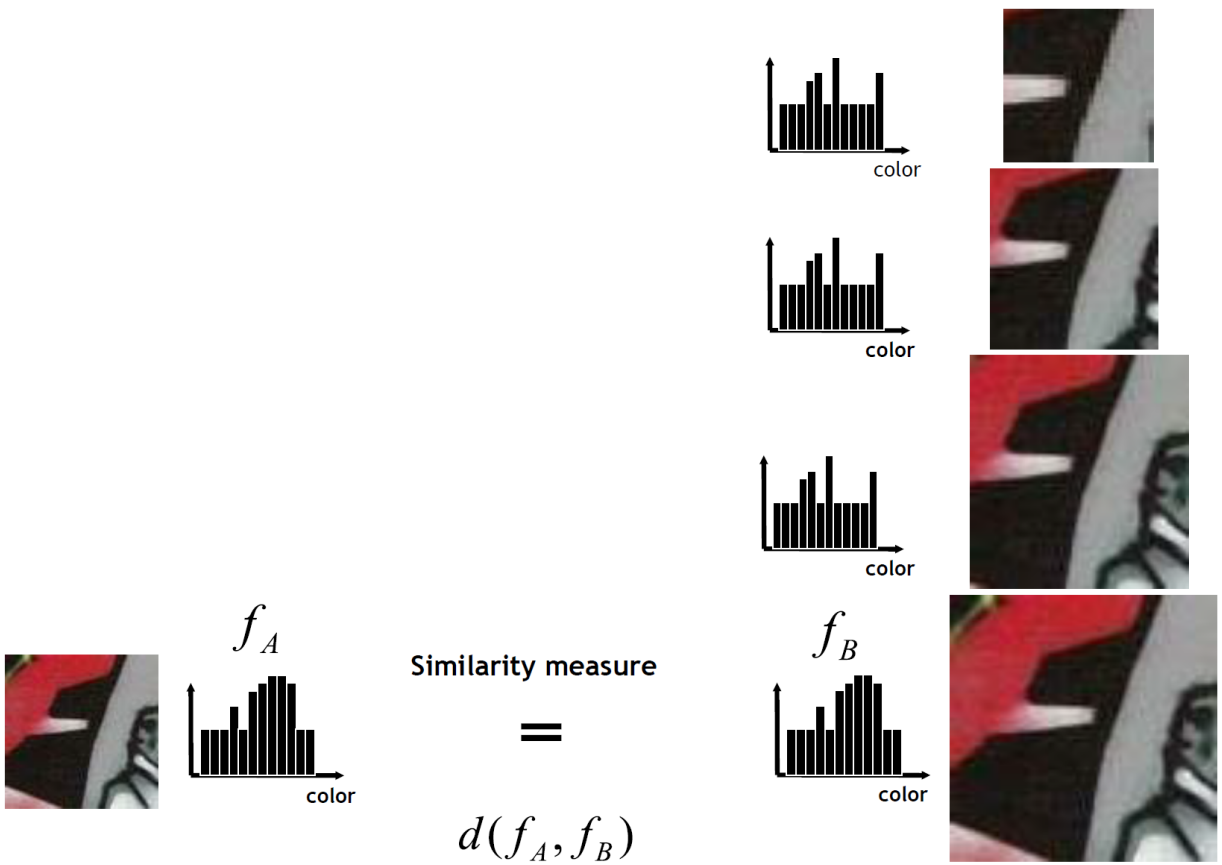


Feature Matching

Scale Invariant Interest Point Detection

Technical University of Applied Sciences Würzburg-Schweinfurt

Inefficient Solution: Compare feature descriptors while varying the patch size

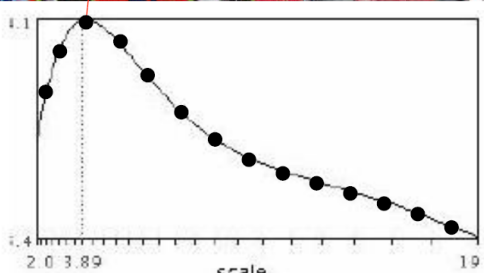


Feature Matching

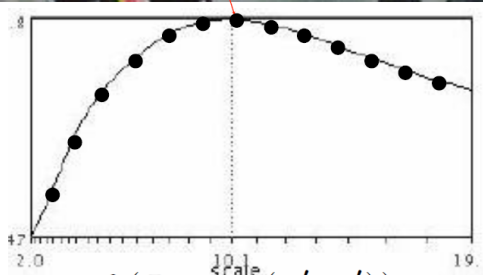
Scale Invariant Interest Point Detection

Technical University of Applied Sciences Würzburg-Schweinfurt

Better Solution: Detector finds location and optimal scale of an interest point.
What is a good measure for the optimal scale?



$f\left(I_{r_1 \ldots r_m}(x, \sigma)\right)$

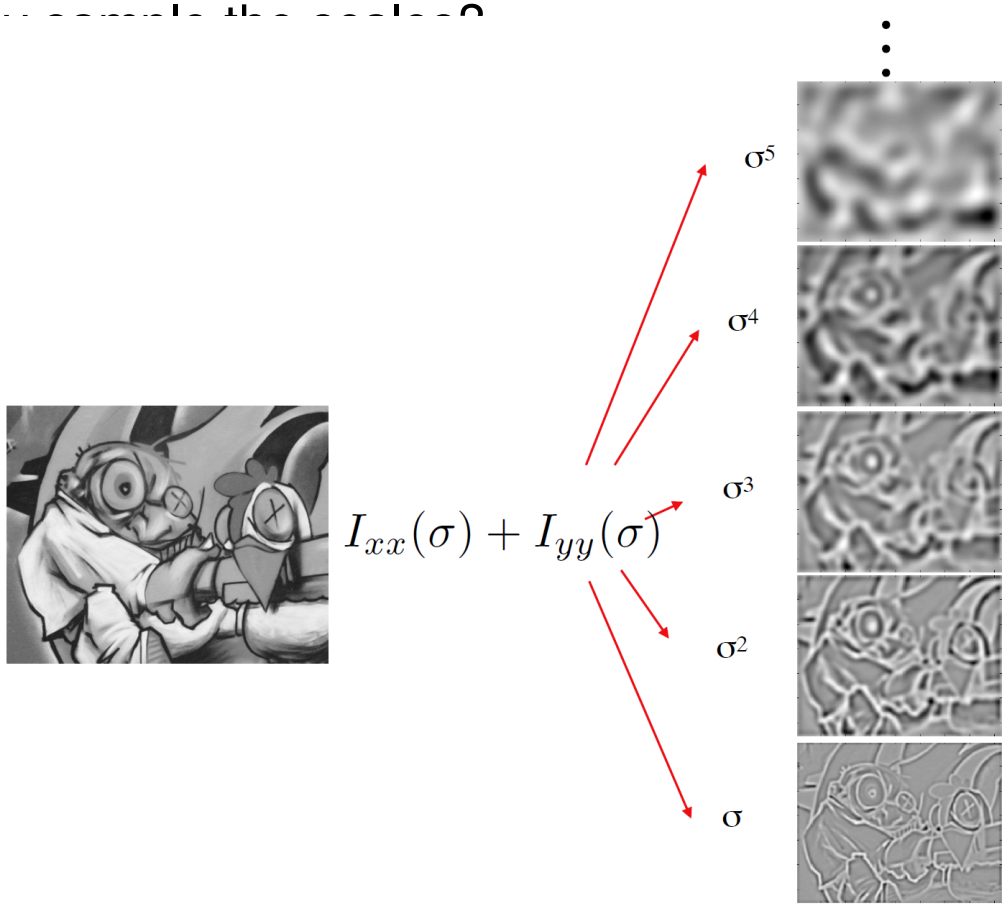


$f\left(I_{r_1 \ldots r_m}(x', \sigma')\right)$

Feature Matching

Scale Invariant Interest Point Detection

A good measure to find the local optimal scale is the Laplacian of Gaussian. How to efficient[illegible]

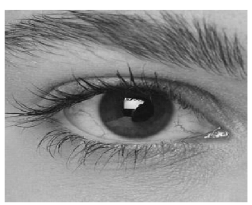


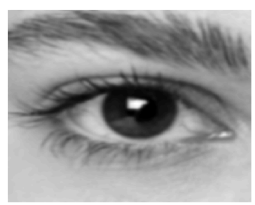
Feature Matching

Scale Space

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

Benefit: The detection of features of an object in an image depends on the size of the object and the available image resolution. For each feature, there is not only one suitable detector (e.g. the Laplace operator for edge detection), but also a suitable image resolution.





Definition: A data structure consisting of a sequence of images with different resolutions is called a scale space. The smaller the image resolution, the smaller the wavenumber range that can be represented. A scale corresponds to a particular wavenumber range.

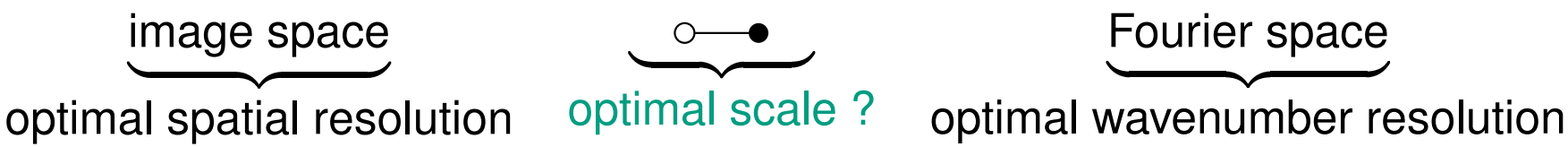
Feature Matching

Scale Space

The representations of an image in spatial domain or in wavenumber domain represent two extremes:

image space: The positional accuracy is equal to the image resolution, but no wavenumber information is available.

Fourier space: The wavenumber resolution is equal to the image resolution, but there is no position information left.



Feature Matching

Scale Space - Grid Structures

Technical University of Applied Sciences Würzburg-Schweinfurt

Fine structures, i.e. high wavenumber ranges, can only be displayed with a high local image resolution. Coarse structures, i.e. low wavenumber ranges, require only a low local image resolution to be displayed.

Idea: Therefore, it makes sense to scan different wavenumber ranges also only with the resolution that is needed for a complete display. (Caution: Consider sampling theorem!)

wavenumber range full resolution

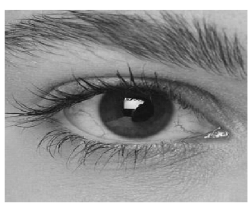
wavenumber range 1/16 resolution

Feature Matching

Scale Space - Gaussian Pyramid

The Gaussian pyramid is a series of recursively low-pass filtered images where the cutoff wavelength is halved by halving the image resolution at each recursion step. The total memory of the pyramid only increases by 1/3 compared to the original image due to the recursive halving. The computational cost only increases by 1/3.

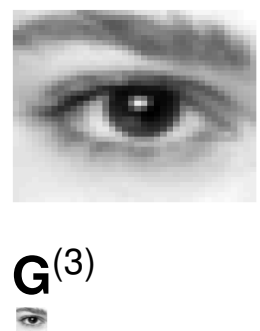
$$\mathbf{G}^{(0)} = \mathbf{G}\,, \quad \mathbf{G}^{(q+1)} = \mathcal{B}_{\downarrow 2}\mathbf{G}^{(q)}\,. \qquad \mathcal{B}_{\downarrow 2}: \text{Binomial filter \& downsampling}$$



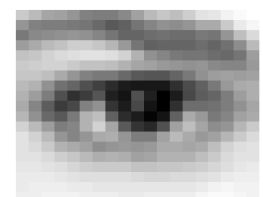
$\mathbf{G}^{(0)}$

$\mathbf{G}^{(1)}$

$\mathbf{G}^{(2)}$



$\mathbf{G}^{(3)}$



$\mathbf{G}^{(4)}$

Feature Matching

Scale Space - Laplacian Pyramid

Technical University of Applied Sciences Würzburg-Schweinfurt

The Laplacian pyramid is a series of bandpass filtered images in which the central wavenumber $\mathbf{k}_0$ is halved from scale to scale. By limiting the wavenumber resolution, good spatial resolution is preserved. The wavenumber resolution is quite coarse and there is no directional decomposition (All wavenumbers within the band are included).

The Laplacian pyramid can be constructed directly from the Gaussian pyramid:

$$\mathbf{L}^{(P)} = \mathbf{G}^{(P)}, \quad \mathbf{L}^{(p)} = \mathbf{G}^{(p)} -_{\uparrow 2} \mathbf{G}^{(p+1)}. \qquad \uparrow 2 : \text{Upsampling}$$

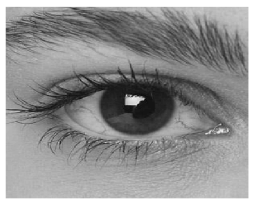
The original image can be completely reconstructed from the Laplacian pyramid:

$$\mathbf{G}^{(P)} = \mathbf{L}^{(P)}, \quad \mathbf{G}^{(p-1)} = \mathbf{L}^{(p-1)} +_{\uparrow 2} \mathbf{G}^{(p)}.$$

Feature Matching

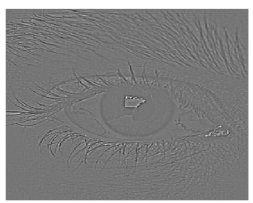
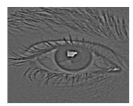
Scale Space - Laplacian Pyramid

$\mathbf{G}^{(0)}$ $\mathbf{G}^{(1)}$ $\mathbf{G}^{(2)}$ $\mathbf{G}^{(3)}$ $\mathbf{G}^{(4)}$





From Gaussian pyramid $\mathbf{G}^{(p)}$ to Laplacian pyramid $\mathbf{L}^{(p)}$.

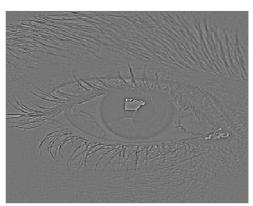
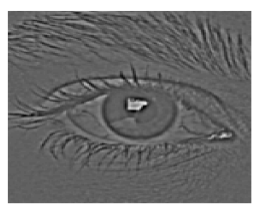
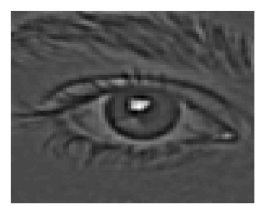

$\mathbf{L}^{(0)} = \mathbf{G}^{(0)} -_{\uparrow 2} \mathbf{G}^{(1)}$

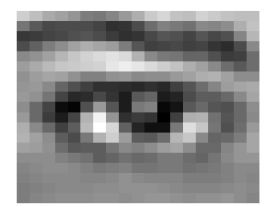
$\mathbf{L}^{(1)} = \mathbf{G}^{(1)} -_{\uparrow 2} \mathbf{G}^{(2)}$

$\mathbf{L}^{(2)} = \mathbf{G}^{(2)} -_{\uparrow 2} \mathbf{G}^{(3)}$

$\mathbf{L}^{(3)} = \mathbf{G}^{(3)} -_{\uparrow 2} \mathbf{G}^{(4)}$

$\mathbf{L}^{(4)} = \mathbf{G}^{(4)}$

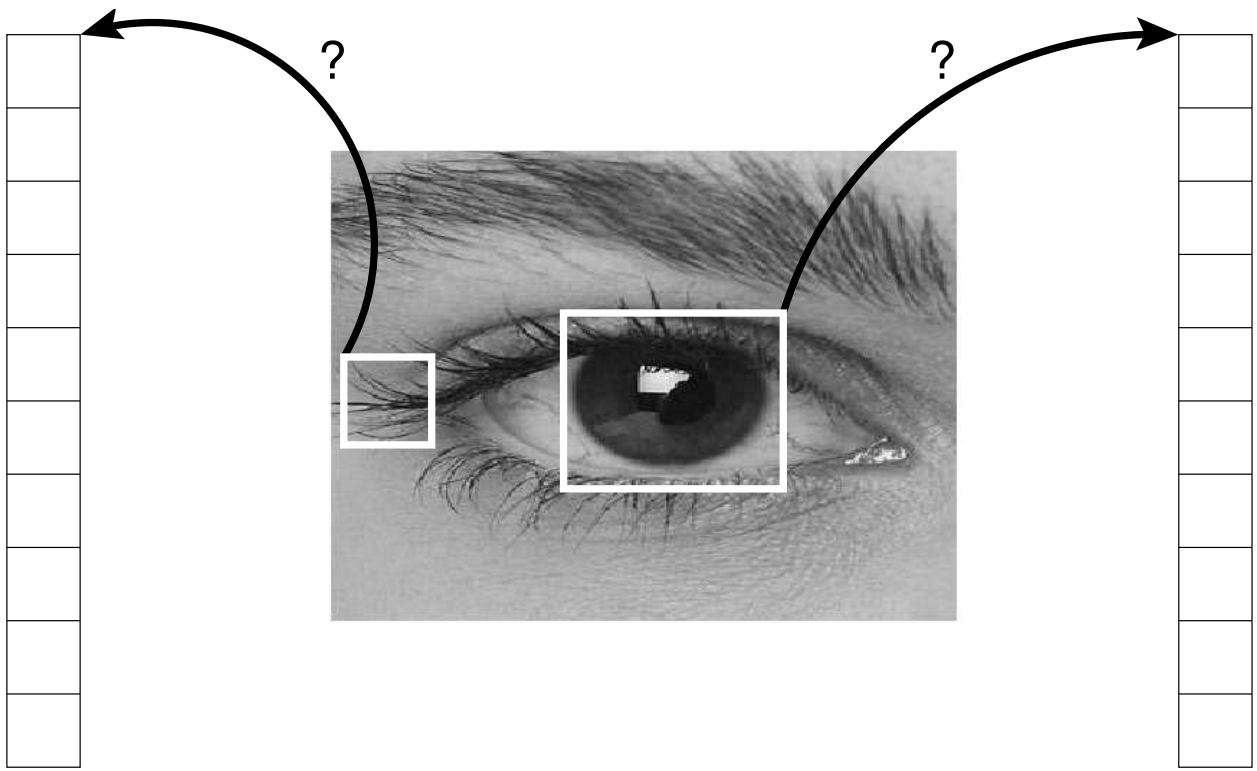






Deskriptors

Idea

Question: How is a patch of the image or an object in the image (ROI) described and how is this description represented?



Deskriptors

Definition - Aim & Usefulness

Definition: A descriptor represents certain properties of an image section or object in the form of a feature vector.

Aim: Depending on the task, these features should be as discriminative (distinguishable) as possible for a certain object or a certain object class and as invariant as possible (uninfluenceable) against fluctuations in the characteristics of an object.

Benefit: The more discriminative and invariant a descriptor is, the easier a segmentation, detection or classification task can be solved.

Optimization: The choice of the relationship between the complexity of a descriptor and the nonlinearity of a detector (2-class problem) or classifier (multiclass problem) depends on the discriminativity and the invariance properties of an object resp. of an object class.

Deskriptors

Properties - Invariance & Discriminativity

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

There are two criteria that are crucial for evaluating a descriptor:

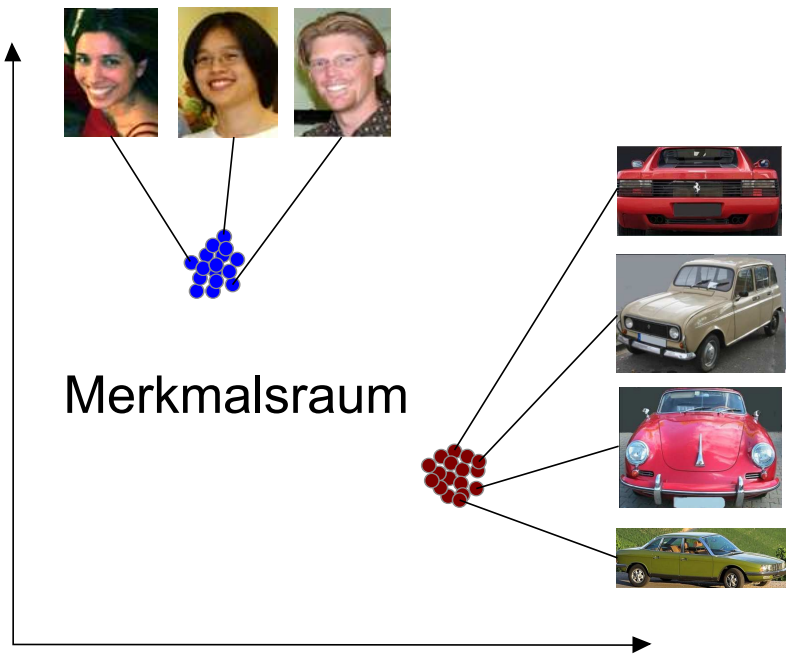
Discriminativity: The fewer feature vectors are in a neighborhood of the feature vector of a given descriptor and the larger the area with no adjacent feature vector in the feature space is, the more discriminative (distinguishable) a descriptor is.

Invariance: The less variations of object properties affect the variables of the feature vector of the descriptor, the more invariant it is given over such variations.

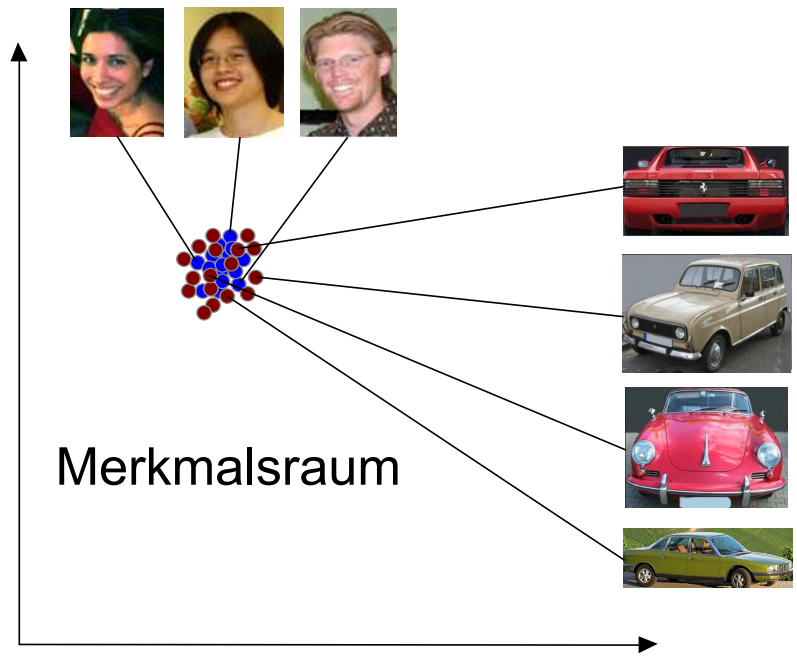
Datasets: The choice of an appropriate data set to assess the goodness of a descriptor is not trivial and should include all occurring variances of an object with a sufficiently large number of combinations (e.g. color and scale).

Deskriptors

Quality of Features - Discriminativity



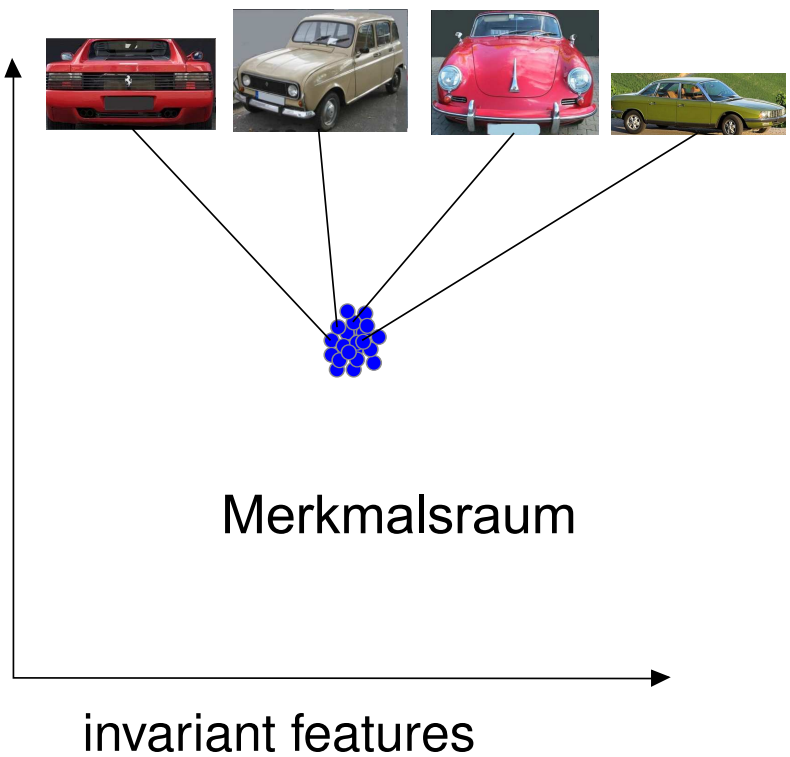
discriminative, distinguishable features

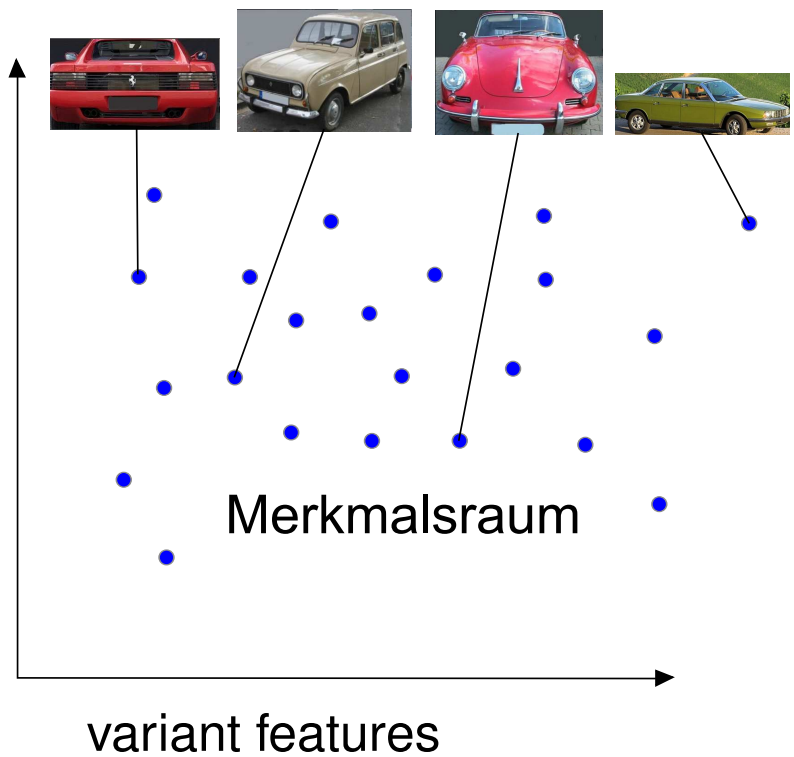


similar, ambiguous features

Deskriptors

Quality of Features - Invariance





Deskriptors

Variations

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

There are a number of variations from which a descriptor should be as independent as possible:

- illumination variations (brightness, contrast, light source movement).
- object variations (see properties)
- background variations (see properties)
- size variations (distance or object?)
- view variations (rotation, perspective distortion)

Deskriptors

Features to be used

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

There is a set of features and their properties that a descriptor can use to describe objects or image patches:

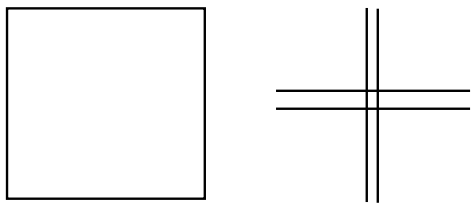
- color properties
- contour properties (2D)
- shape properties (3D)
- structure properties
- texture properties
- motion properties

The properties are basically divided into properties describing the the spatial content and statistical properties, such as the frequency of a certain feature.

Descriptors

Histograms - HoG

Another possibility to describe the structure of an image region is offered by histograms of structural features. The simplest descriptor of this kind is the so-called Histogram of Gradients (HoG). For this purpose, the image section to be described is divided into contiguous cells and for each cell a histogram weighted with the gradient strength is generated. The histogram is generated for each cell over a certain number of bins with different edge orientations. Each of these histograms can be additionally normalized by a vector norm. All histograms concatenated in a vector then result in the HoG descriptor.



Histogram is identical for both structures

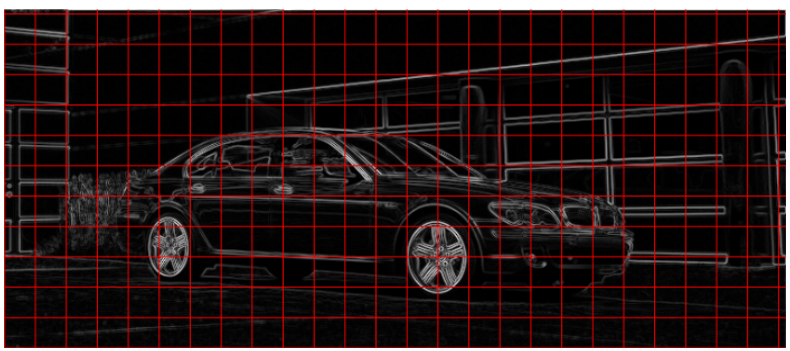
Descriptors

Histograms - HoG

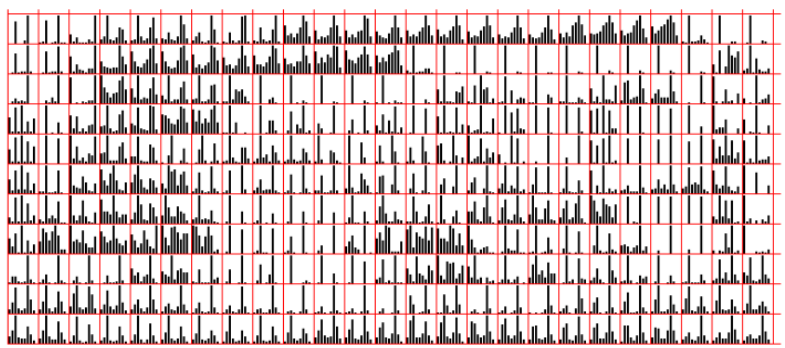
Technical University of Applied Sciences Würzburg-Schweinfurt

Advantage HoGs are to some extent invariant to size scaling. and brightness variations.

Disadvantage The local distribution is not completely lost, but is reduced to the resolution of the cells.



Cells

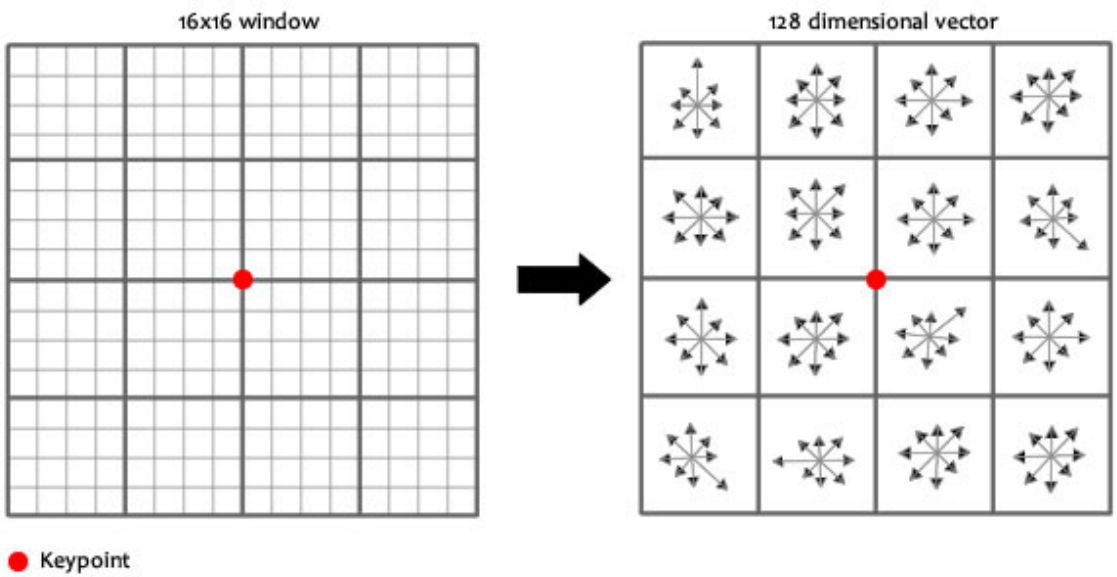


Histograms of Gradients

Descriptors

SIFT - Scale-Invariant Feature Transform

- Divide a 16 $\times$ 16 patch in 4 $\times$ 4 grid of cells
- Compute an orientation histogram for each cell
- 16 cells $\times$ 8 orientations = 128 dimensional descriptor

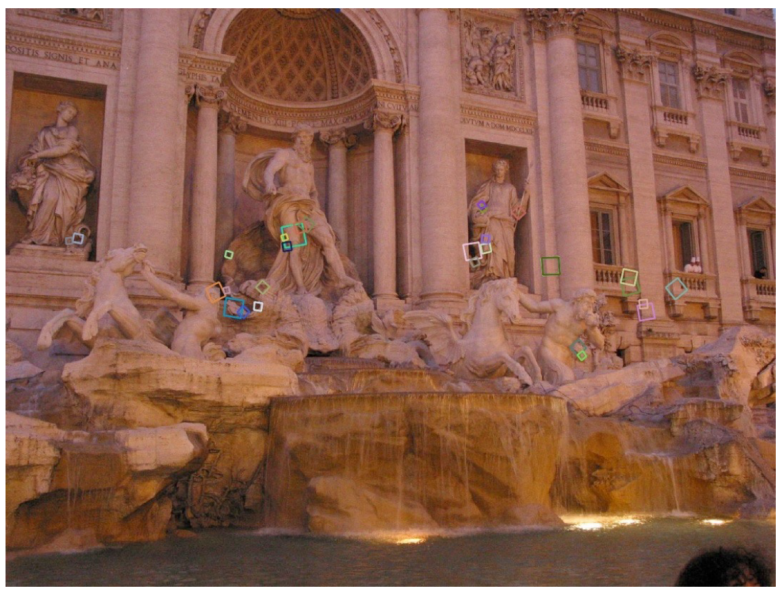


Descriptors

SIFT - Scale-Invariant Feature Transform

Technical University of Applied Sciences Würzburg-Schweinfurt

- can handle changes in viewpoint up to 60 degrees out of plane rotation
- can handle significant changes in illumination
- fast and efficient - can run in realtime

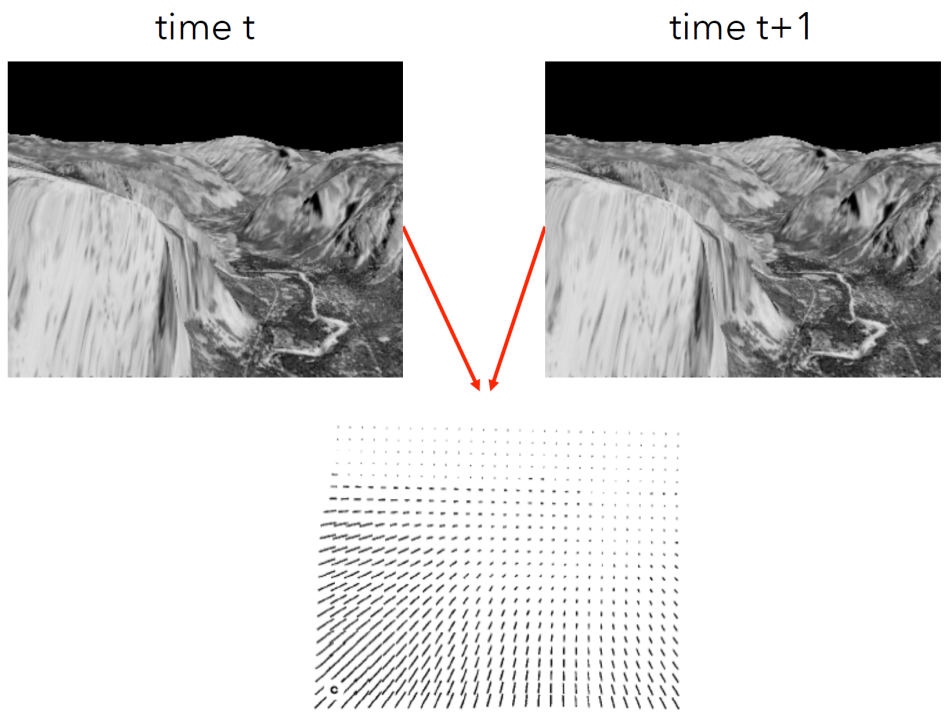


Feature Matching

Optical Flow

Technical University of Applied Sciences Würzburg-Schweinfurt

Optical Flow methods try to estimate the motion field between two images taken at two adjacent instances in time and/or from two slightly different views. The flow field either results from moving objects or from the movement of the camera or from a superposition of both.

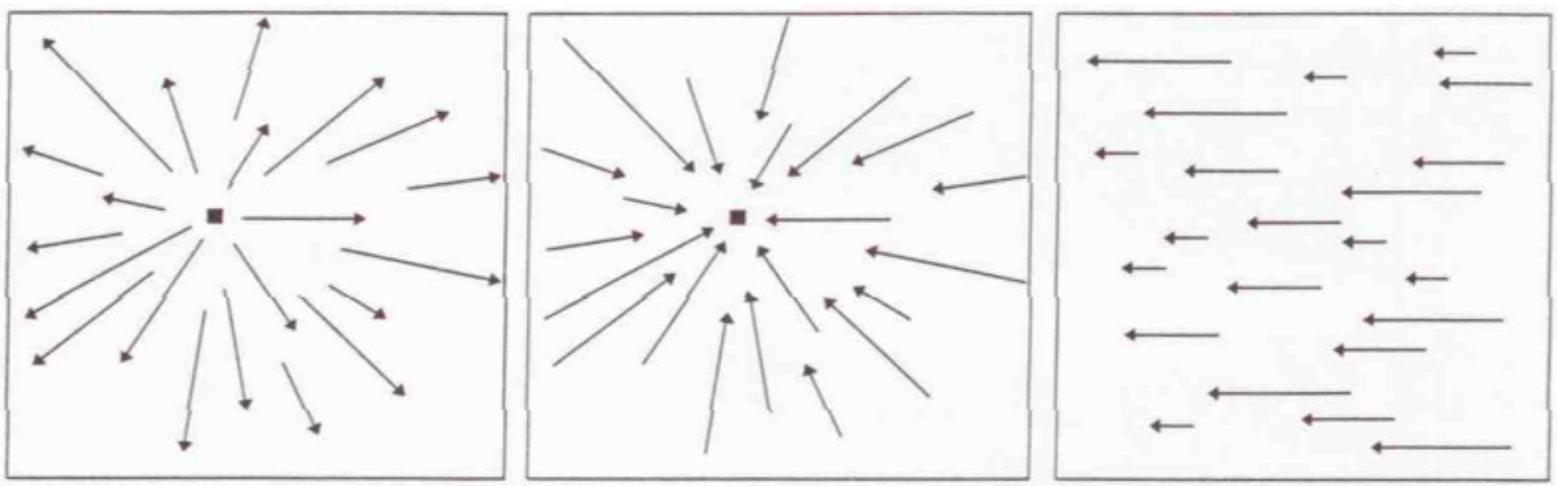


Feature Matching

Ego Motion Flow Fields

Technical University of Applied Sciences Würzburg-Schweinfurt

If a camera moves and we have a small baseline between adjacent camera frames the following typical flow fields appear. If the camera moves forward/backward then we get expanding/contracting flows with a certain focus of expansion/contraction. Pure camera rotation results in a translational flow.



Feature Matching

Optical Flow - Lucas-Kanade Method

Technical University of Applied Sciences Würzburg-Schweinfurt

Let's derive the Lucas-Kanade Method!

This is a classical very fast flow technique for small up to medium displacements invented by B. D. Lucas and T. Kanade and published in 1981.

The title of the paper is: *An iterative image registration technique with an application to stereo vision*.

It was published on the *International Joint Conferences on Artificial Intelligence* (IJCAI), pp. 674-679, 1981.

Feature Matching

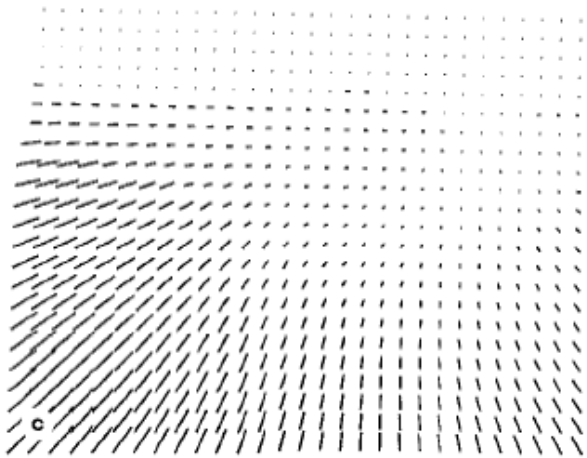
Optical Flow - Task

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

Measurements We can measure the image brightness $G(x, y, t)$ at each pixel location $\mathbf{x} = [x, y]^\top$ at different points in time t from one camera.

Estimates We want to estimate the 2D velocity field for each pixel $\mathbf{v}(\mathbf{x}) = [u(x, y), v(x, y)]^\top$.

Method We cannot measure the velocity field directly but the optical flow which corresponds to the velocity field in a lot of cases.



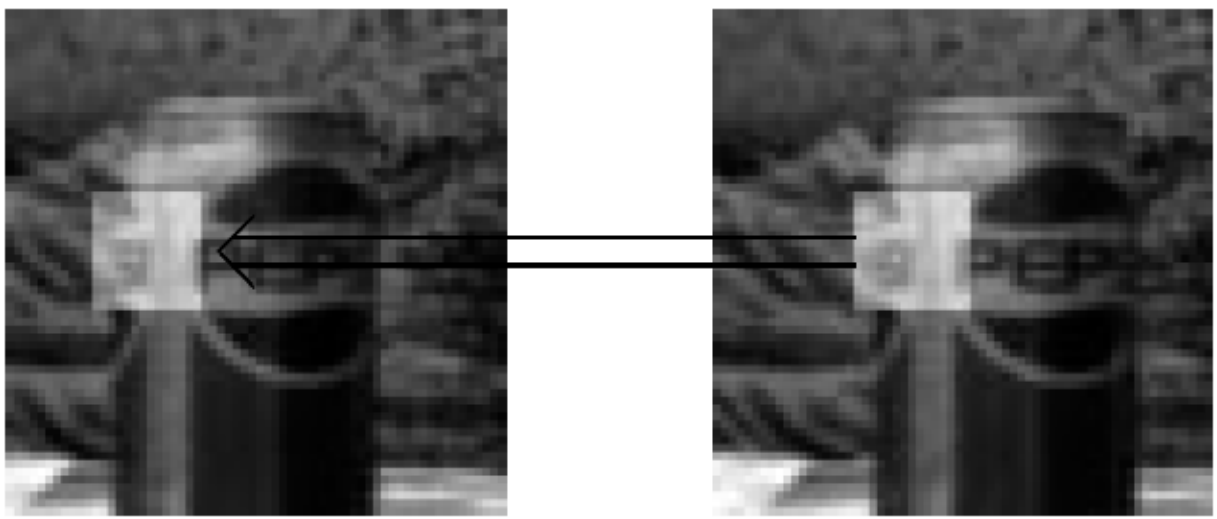
Feature Matching

Optical Flow - Assumptions

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

1. Assumption Gray value measurements in a small region remain the same although their location may change.

Brightness Constancy $G(x + u(x, y), y + v(x, y), t + 1) = G(x, y, t)$.



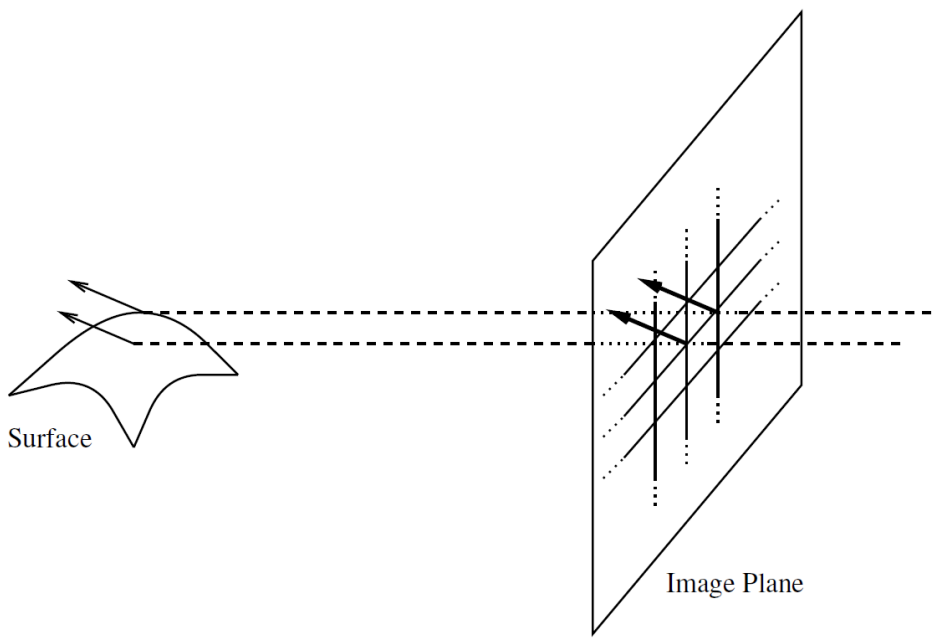
Feature Matching

Optical Flow - Assumptions

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

2. Assumption Neighboring points in the scene typically belong to the same surface and hence typically have similar 3D motions.

Spatial Coherence Since they also project to nearby points in the image, we expect neighboring 2D points to have similar velocity: $\mathbf{v}(\mathbf{x}) = \mathbf{v}(\mathbf{x}')$.



Feature Matching

Optical Flow - Optimization Problem

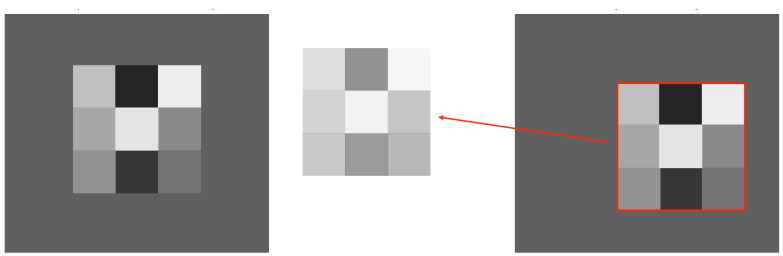
Technical University of Applied Sciences Würzburg-Schweinfurt

Minimize Brightness Differences We can combine both assumptions, brightness constancy and spatial coherence, in one objective.

SSD Matching Using a sliding window approach (correlation), we can measure the SSD for a number of discrete velocity vectors.

Advantage Works for an arbitrary number of also large displacements.

Disadvantages It is computationally expensive and we cannot achieve subpixel accuracy.



Feature Matching

Optical Flow - Differential Approach

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

Linearize Brighness Constancy Assumption (1st order Taylor expansion)

$$\begin{aligned} G(x+u, y+v, t+1) &\approx G(x,y,t) + uG_x(x,y) + vG_y(x,y) + G_t(x,y) \\ G(x+u, y+v, t+1) - G(x,y,t) &\approx uG_x(x,y) + vG_y(x,y) + G_t(x,y) \end{aligned}$$

Include approximation in SSD Matching leads to a convex objective

$$\text{argmin}_{u,v} \sum_{x,y \in \mathcal{N}} \left(uG_x(x,y) + vG_y(x,y) + G_t(x,y) \right)^2 .$$

Solving this minimization is called the Lucas-Kanade Method.

Feature Matching

Optical Flow - Brightness Constancy

Technical University of Applied Sciences Würzburg-Schweinfurt

The linearized brightness constancy assumption leads to the so-called optical flow constraint equation:

$$G(x+u, y+v, t+1) - G(x, y, t) \approx uG_x + vG_y + G_t \stackrel{!}{=} 0\,,$$

that assumes the gray value difference between corresponding pixels to be zero.

In vector notation this constraint reads

$$\nabla \mathbf{G}^\top \mathbf{v} + G_t = 0\,,$$

with $\nabla \mathbf{G} = [G_x, G_y]^\top$ and $\mathbf{v} = [u, v]^\top$.

Feature Matching

Optical Flow - Constraint

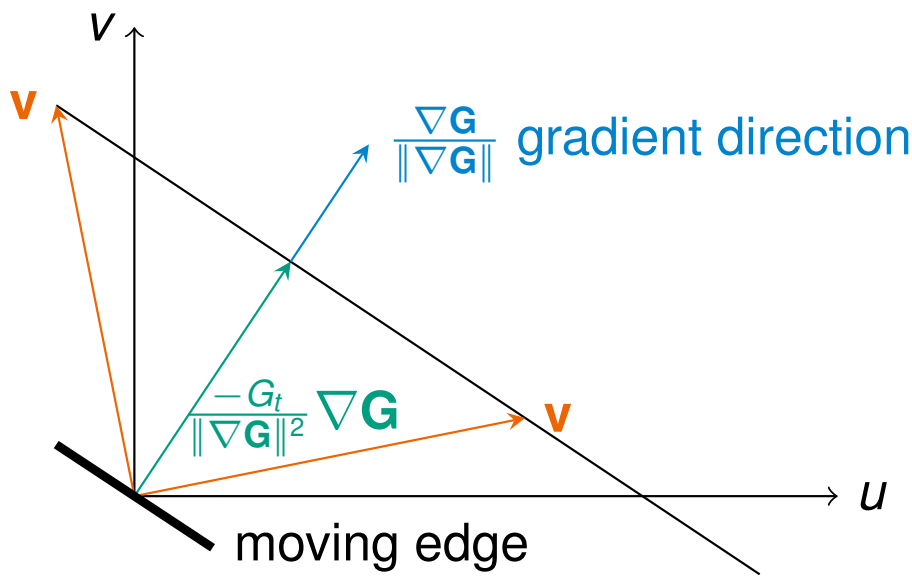
Technical University of Applied Sciences Würzburg-Schweinfurt

For each single pixel we get a constraint line for the flow vector **v**:

$$\left(\frac{\nabla \mathbf{G}}{\|\nabla \mathbf{G}\|}\right)^{\top} \mathbf{v} - \left(-\frac{G_t}{\|\nabla \mathbf{G}\|}\right) = 0\,.$$

Here, rewritten in Hesse normal form with the normal flow vector $-\frac{G_t}{\|\nabla \mathbf{G}\|^2}\nabla \mathbf{G}$

perpendicular to the moving edge.
The flow component parallel
to the edge cannot be found
from one constraint.

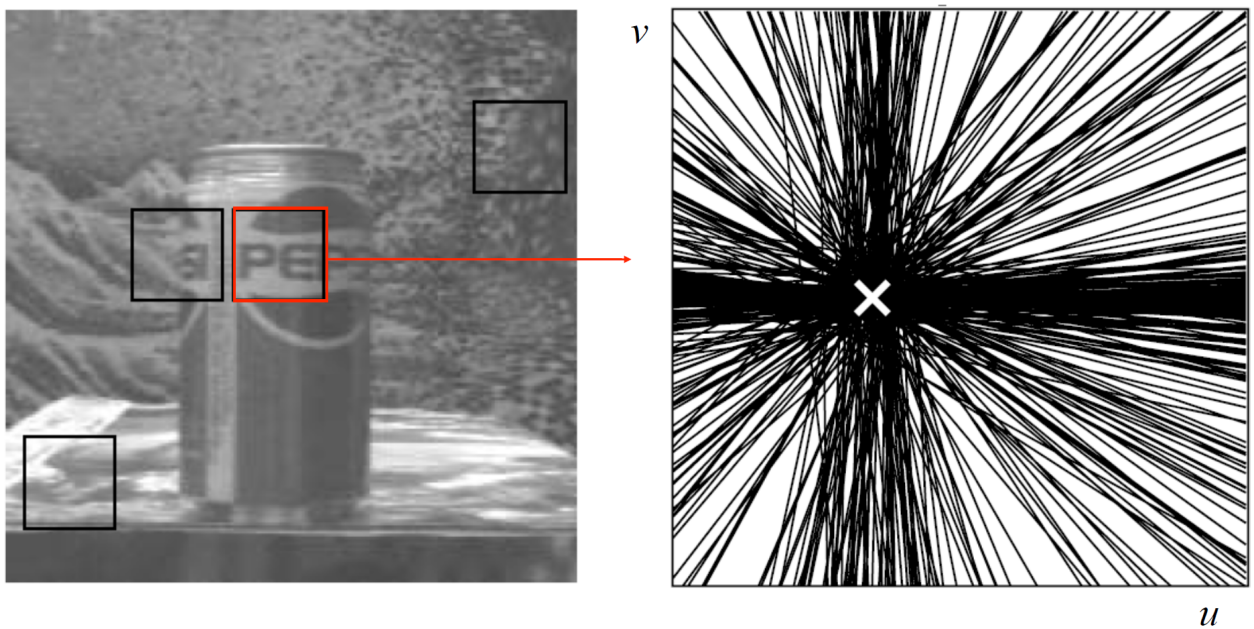


Feature Matching

Optical Flow - Multiple Constraints

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

If we assume all *N* pixels of an image patch moving with the same constant velocity **v** these pixels define *N* line constraints for this particular velocity.



Feature Matching

Optical Flow - Overdetermined EQS

thws
Technical University
of Applied Sciences
Würzburg-Schweinfurt

All N neighboring pixels $\mathbf{x}'$ define N line constraints. This leads to an overdetermined linear equation system with N linear equations for the two unknowns $\mathbf{v} = [u, v]^\top$ that define the optical flow vector of an image patch centered at $\mathbf{x}$:

$$\begin{bmatrix} \nabla \mathbf{G}^\top(\mathbf{x}_1) \\ \vdots \\ \nabla \mathbf{G}^\top(\mathbf{x}') \\ \vdots \\ \nabla \mathbf{G}^\top(\mathbf{x}_N) \end{bmatrix} \mathbf{v} = \begin{bmatrix} -G_t(\mathbf{x}_1) \\ \vdots \\ -G_t(\mathbf{x}') \\ \vdots \\ -G_t(\mathbf{x}_N) \end{bmatrix}, \qquad \nabla \mathbf{v} = \mathbf{g},$$

which can be solved with linear regression techniques, like the least squares method. The solution minimizes the linearized SSD objective given on slide 77.

Feature Matching

Optical Flow - Lucas-Kanade Method

Technical University of Applied Sciences Würzburg-Schweinfurt

The Lucas-Kanade least squares solution that minimizes the linearized SSD objective reads:

$$
\begin{aligned}
\mathbf{v}(\mathbf{x}) &= \begin{bmatrix} \sum_{\mathbf{x}'} G_x^2(\mathbf{x}') & \sum_{\mathbf{x}'} G_x(\mathbf{x}')G_y(\mathbf{x}') \\ \sum_{\mathbf{x}'} G_y(\mathbf{x}')G_x(\mathbf{x}') & \sum_{\mathbf{x}'} G_y^2(\mathbf{x}') \end{bmatrix}^{-1} \begin{bmatrix} \sum_{\mathbf{x}'} -G_x(\mathbf{x}')G_t(\mathbf{x}') \\ -G_y(\mathbf{x}')G_t(\mathbf{x}') \end{bmatrix}, \\
&= -\left(\sum_{\mathbf{x}'} \nabla\mathbf{G}\nabla\mathbf{G}^\top \right)^{-1} \left(\sum_{\mathbf{x}'} G_t \nabla\mathbf{G} \right), \\
&= \left(\nabla^\top\nabla\right)^{-1}\nabla^\top\mathbf{g} = \nabla^+\mathbf{g}.
\end{aligned}
$$

The solution includes the inverse of the structure tensor and the pseudo inverse ∇^+ of the gradient matrix ∇. Hence, we can use the structure tensor to define interest points and then solve for their motion.

The structure tensor is a very important tool for computer vision!

Feature Matching

Optical Flow - Pyramidal Approach

Technical University of Applied Sciences Würzburg-Schweinfurt

Coarse-to-fine strategy: Apply Lucas-Kanade on multiple scales!

- Build a Gaussian pyramid
- Start with the lowest resolution (motion is small!)
- Use this motion to pre-warp the next finer scale (see backward warping from Machine Vision I course)
- Only compute motion increments (motion is small!)