DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF COPENHAGEN

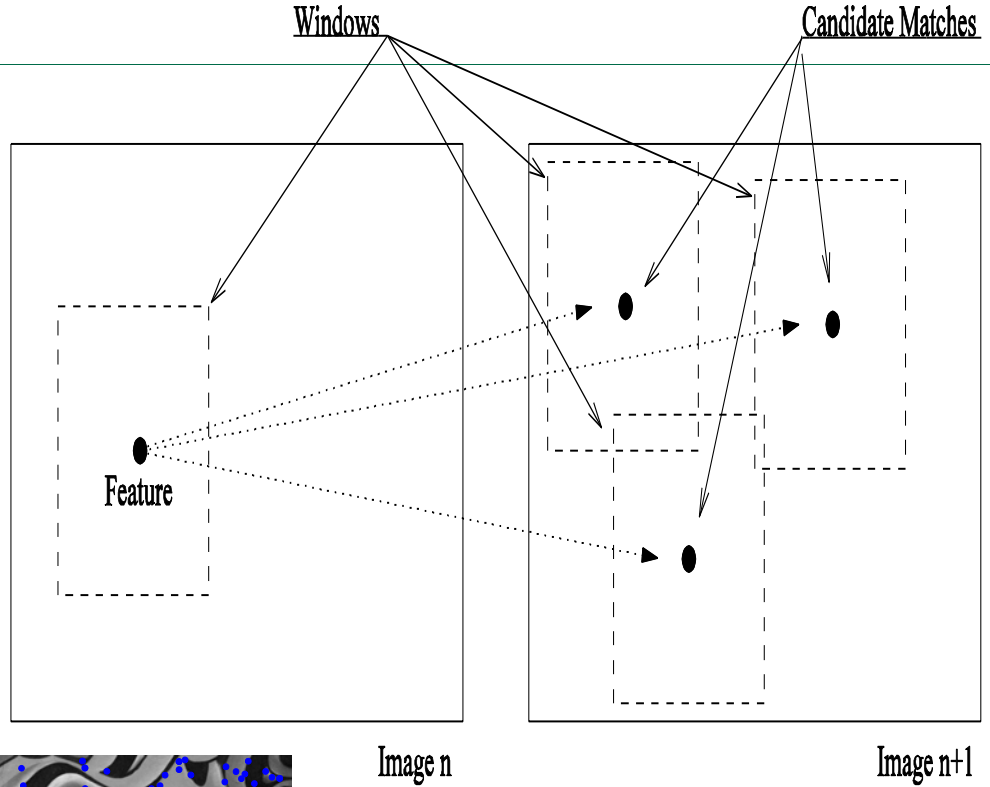
Features III:
Advanced Descriptors

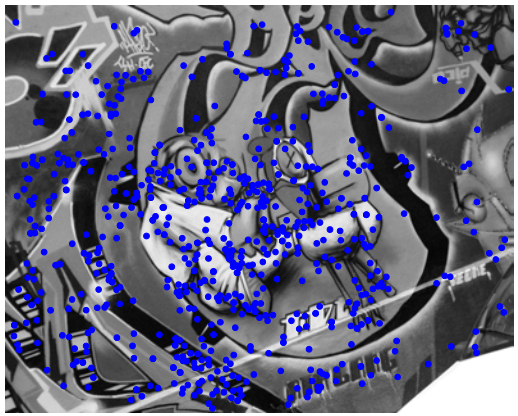
Kim Steenstrup Pedersen

Plan for today

- Again we consider the problem of matching two or more images
- Today we look at advanced descriptors
- We end by looking at a comparative study of state of the art descriptors

Matching Strategy Illustrated





Salient Feature Matching Strategy

For a pair of images:

1. Extract salient points, via **detectors**
e.g. Harris corners, DoG (blobs), MSER

2. Compute Feature **descriptors**,
eg. Raw patch (correlation), SIFT, DAISY

3. Match features by pairing similar descriptors

Aggregate solution to multiple images, if needed.

Matching patches

Raw pixel descriptor: Use pixels in patch and compare with Normalized Cross Correlation

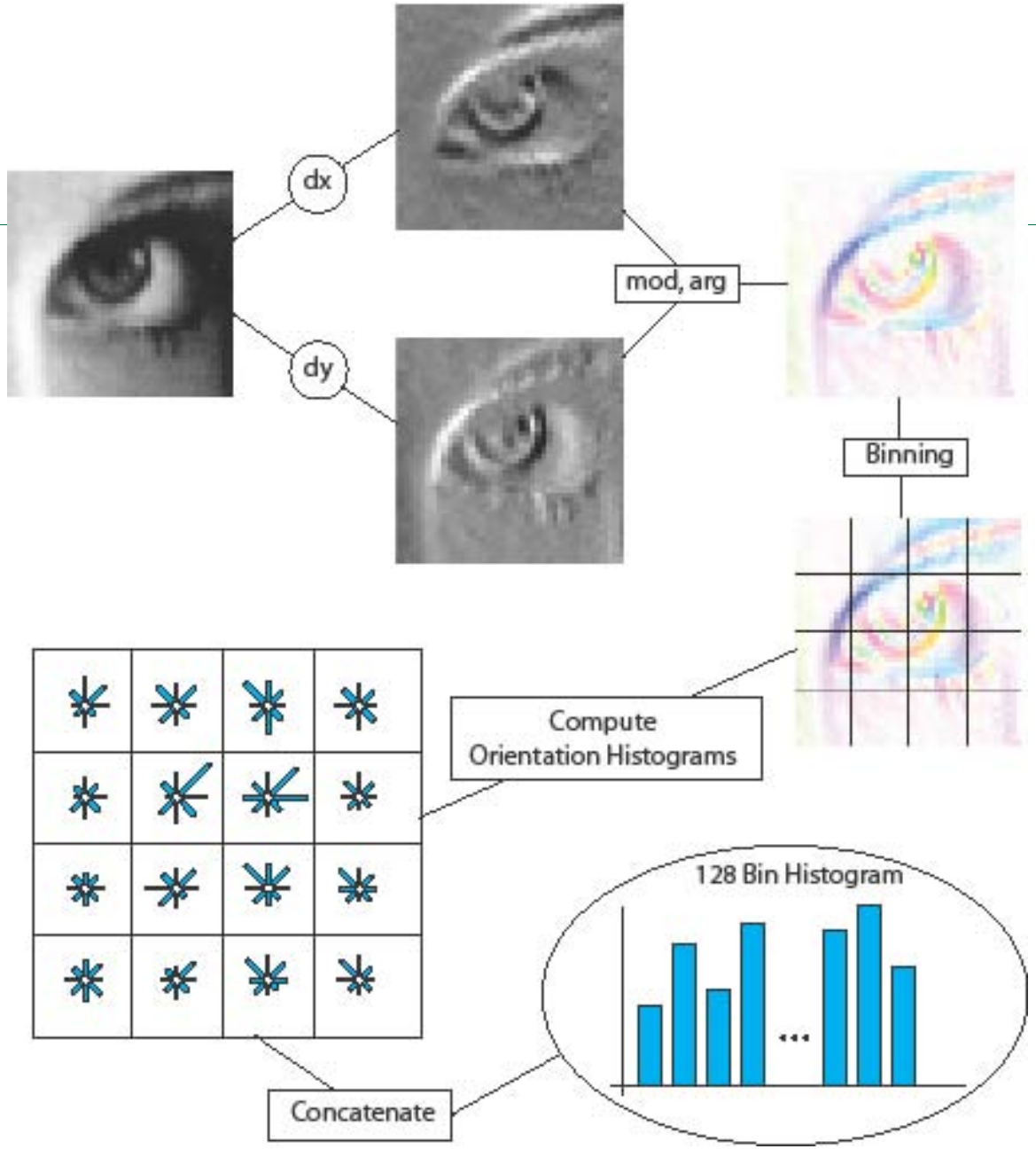
Open problems for raw pixel descriptor

- What patch size should we use?
  - Use the detection scale and resample so both patches have equal size in pixels
- Is this approach robust to scale changes in the scene?
  - Yes, if we do the resampling
- Is this approach robust to rotation and translation in the scene?
  - No, this will lead to large dissimilarities
- Is it robust to perspective distortions?
  - No, this will lead to large dissimilarities

Scale Invariant Feature Transform (SIFT)

- SIFT is a very popular descriptor
  (google.scholar says 21528 citations Nov. 2013).
- Scale invariance:
  - This is obtained by using the DoG blob detector which is multi-scale. Descriptor build at these interest points in scale-space.
  - After detection we have an interest point at $(\tilde{x},\tilde{y},\tilde{\sigma})$
- Rotational invariance:
  - Estimate an orientation of the interest point and build the descriptor relative to this.
- Translational invariance:
  - To some extend by construction of the descriptor (more on this)
- Illumination invariance:
  - By construction of the descriptor (more on this)

SIFT


dx
dy
mod, arg
Binning
Compute
Orientation Histograms
128 Bin Histogram
Concatenate

SIFT: Rotational invariance by orientation assignment

- At detection scale compute image gradients

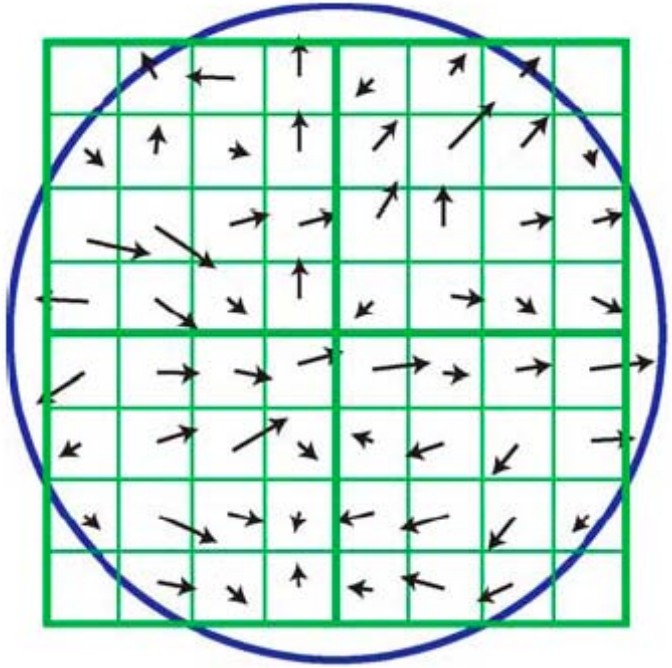
$$\nabla L(x,y,\tilde{\sigma}) = \left(L_x, L_y\right)^T$$ for all points in the image

- Gradient orientation and magnitude images

$$\theta(x,y,\tilde{\sigma}) = \tan^{-1}\left(\frac{L_y}{L_x}\right)$$

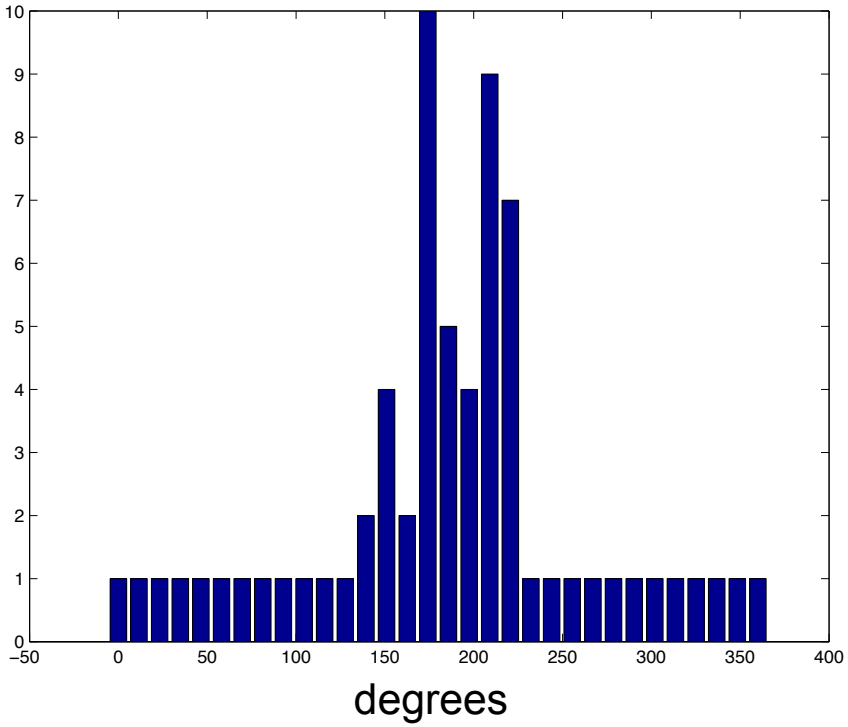
$$m(x,y,\tilde{\sigma}) = \sqrt{L_x^2 + L_y^2}$$

- Build a 32 bin orientation histogram for neighborhood around $(\tilde{x},\tilde{y},\tilde{\sigma})$
  - Every point weighted with *m* and a Gaussian window $G(x-\tilde{x}, y-\tilde{y}, 1.5\tilde{\sigma})$

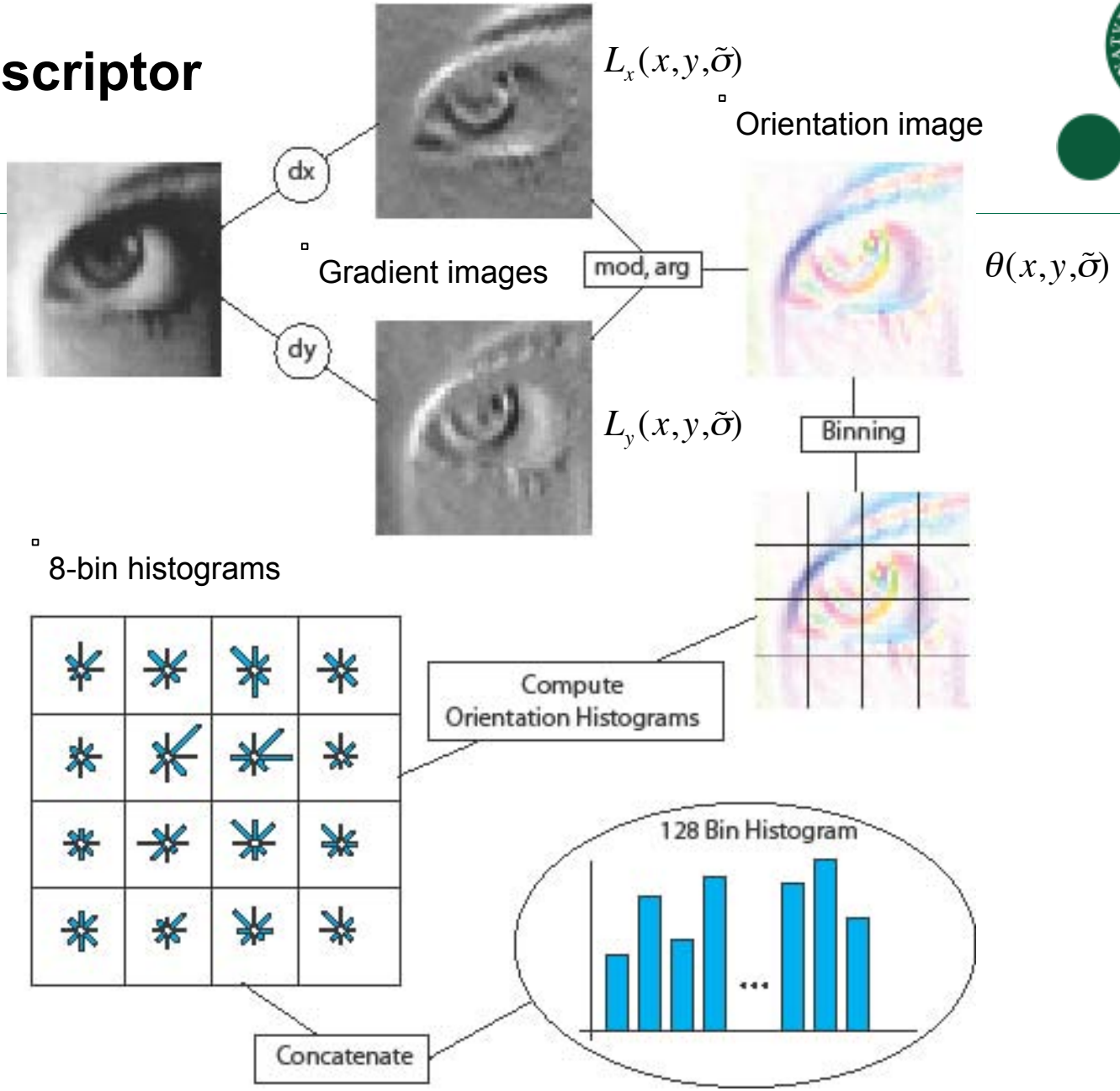


SIFT: Rotational invariance by orientation assignment

- Find highest peak and its orientation
- Create an interest point descriptor using this orientation
- For all other peaks larger than 80% of the highest peak – also create a descriptor using these orientations



SIFT descriptor



The SIFT descriptor: Details

- 8-bin orientation histograms:
  - When adding to bin, every data point is weighted with *m* and a Gaussian aperture window $G(x - \tilde{x}, y - \tilde{y}, 1.5\tilde{\sigma})$
  - Adding a data point to a bin also add a little to the neighboring bins (linear interpolation)
  - Pixels on the other side of a histogram grid border contributes a little to the histogram (linear interpolation)

- Feature vector:
  - Concatenate 8-bin histograms from the 4x4 grid into one vector with dimensionality 4x4x8 = 128.

The SIFT descriptor: Details

- Normalization of feature vector:
  - Normalize the feature vector: $\tilde{F} = F / \|F\| \Rightarrow \left\|\tilde{F}\right\| = 1$
  - Non-linear illumination changes (e.g. shadows) may cause high gradient magnitudes locally.
  - Therefore reduce all bin values larger than 0.2 down to 0.2.
  - Renormalize (normalize again): $\tilde{F} = F / \|F\| \Rightarrow \left\|\tilde{F}\right\| = 1$

SIFT matching:

- SIFT features are compared with Euclidean distance (L2-norm):

$$d_2(F_1, F_2) = \sqrt{\sum_{i=1}^{128} (F_1(i) - F_2(i))^2}$$

- Matching SIFT features:
  - A match if this is true

$$\frac{\text{Best}}{\text{2nd Best}} \leq 0.8$$

  - Best refers to the distance for the pair of features with smallest distance.
  - 2nd Best refers to the distance for the pair of features with second smallest distance.

The SIFT descriptor invariance's

- Scale invariance:
  - From the (DoG) detector and further processing done at detection scale
- Rotational invariance:
  - From the orientation assignment procedure
- Approximate translational invariance:
  - From the grid of histograms. Can handle translations up to 4 pixels (within a grid cell).
- Affine illumination invariance:
  - From the choice of gradients additive brightness invariance is obtained. From normalization we obtain invariance to multiplicative contrast change.
  - Reduction of peaks gives some robustness to non-linear changes such as shadows.

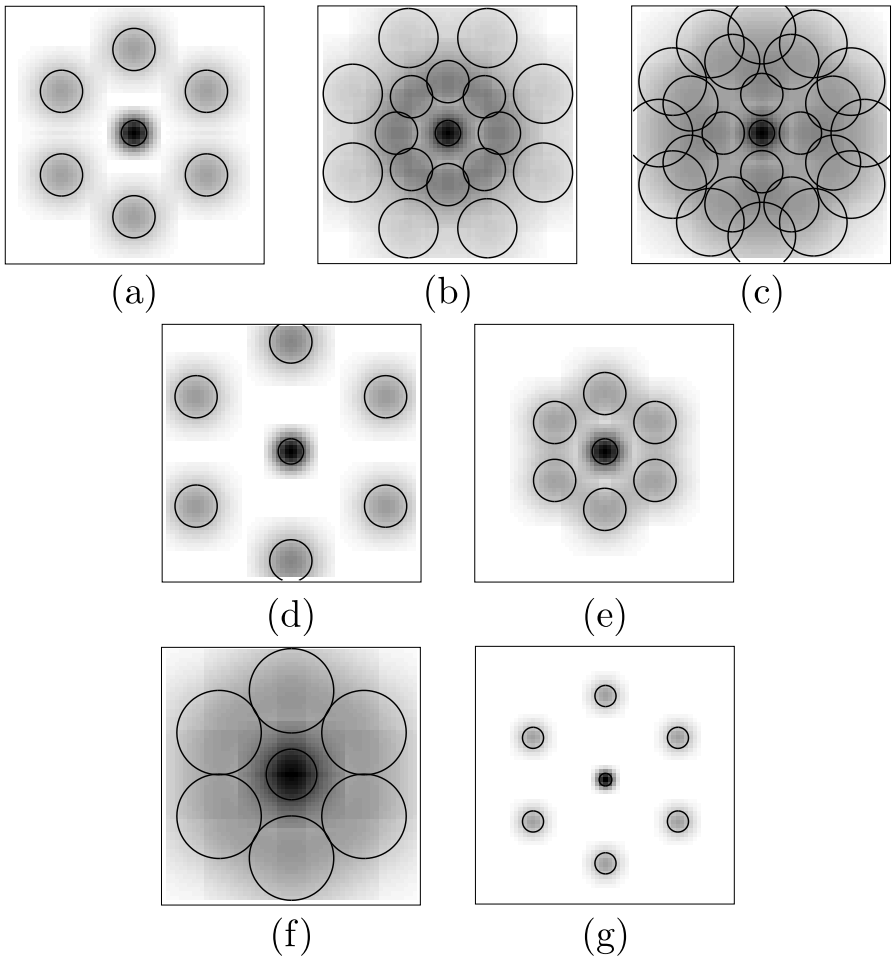
Open problems for SIFT descriptors

- Perspective distortion
- Non-linear illumination such as cast shadows, changes to light color, and material reflection properties
- It is fairly high dimensional (128 dim.) and redundant

Variations on SIFT

- There are many variations of the SIFT descriptor:
  - GLOH
  - SURF
  - DAISY
  - PCA SIFT
  - Opponent SIFT
  - Gaussian opponent SIFT
  - CSIFT
  - ORB
  - BRISK
  - FREAK
  - …

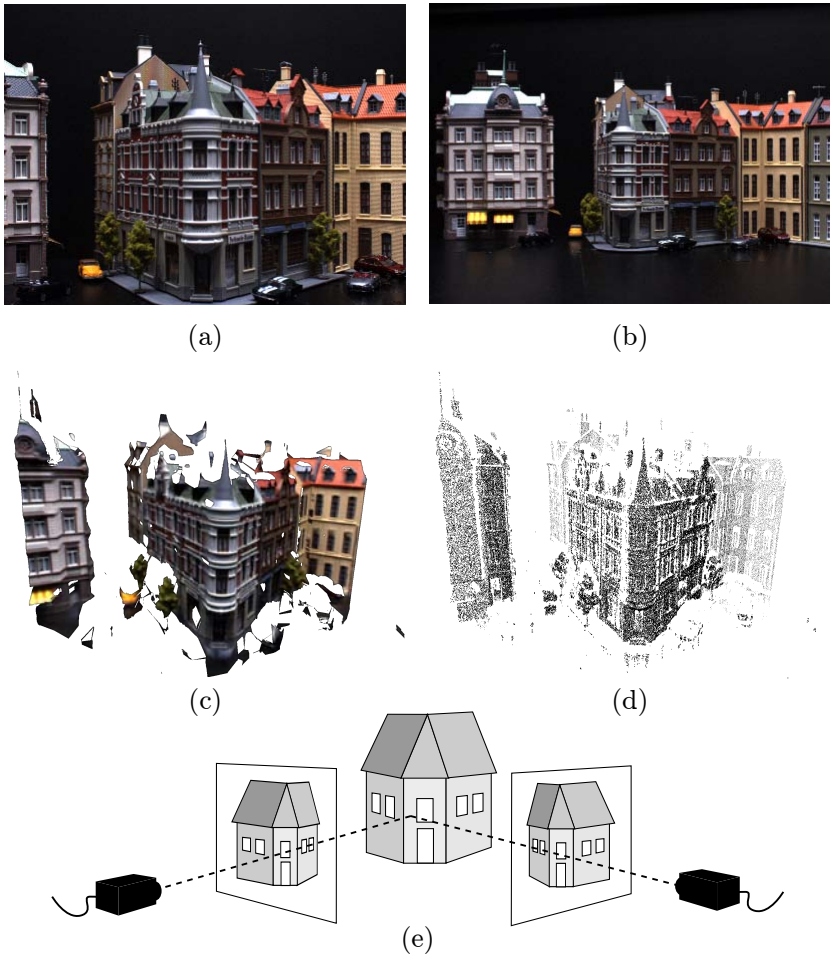
DAISY – a common SIFT variation: Locations and spread of histograms

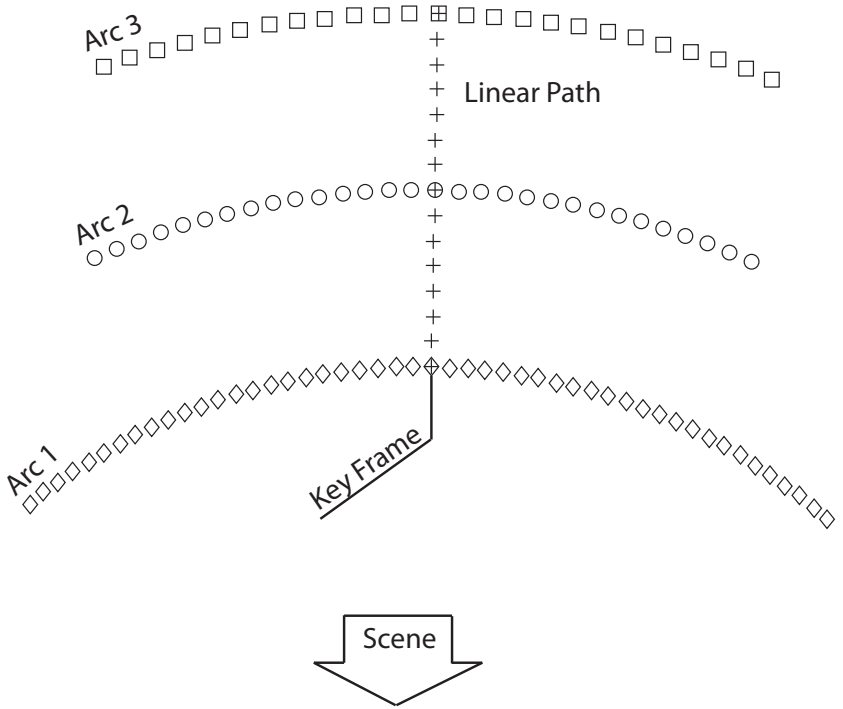


What is the best detector and descriptor combination?

Based on joint work with Anders Dahl and Henrik Aanæs from DTU

Recall the DTU robot data set





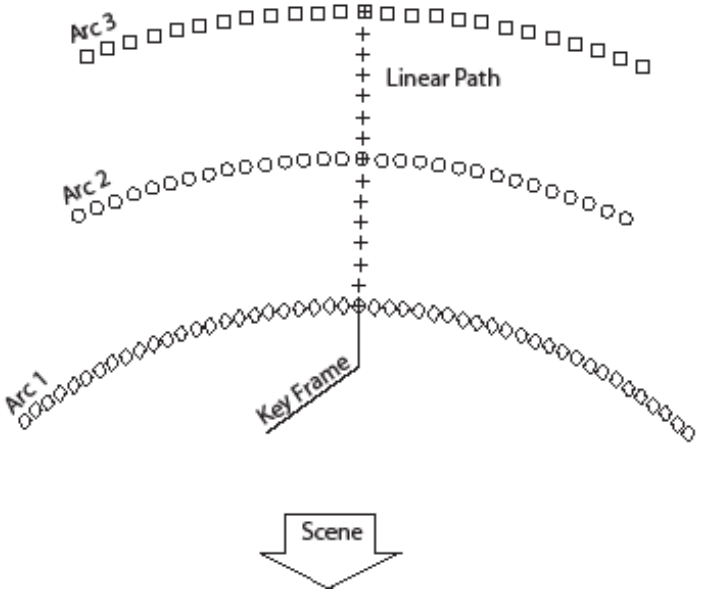
Detector - Descriptor Combination

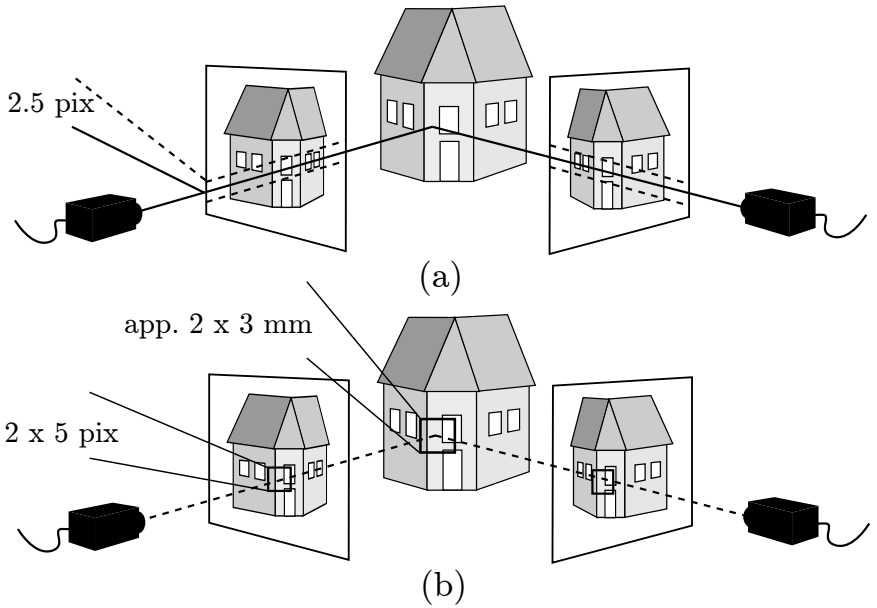
Detector	#Detect.	Affine Invariant
Harris	1	no
Harris Laplace	2	no
Harris Affine	3	yes
Hessian Laplace	4	no
Hessian Affine	5	yes
MSER	6	yes
DoG	7	no
Multiscale Harris	8	no
Multiscale Harris nl.	9	no
Lindeberg corner	10	no
Lindeberg corner nl.	11	no

Descriptor	#Desc.	Color	#Dim.
Raw patch (NCC)	1	no	$34 \times 34 \times 3$
SIFT gray	2	no	128
SIFT RGB bin	3	yes	128
SIFT RGB	4	yes	384
Opponent SIFT	5	yes	384
CSIFT	6	yes	384
Gaussian opponent SIFT	7	yes	384
Hist. eq. SIFT	8	no	128
Hist. eq. SIFT RGB bin	9	yes	128
Hist. eq. SIFT RGB	10	yes	384
DAISY 1-6-6 s	11	no	52
DAISY 1-6-6 l	12	no	104
DAISY 1-8-8-8 s	13	no	100
DAISY 1-8-8-8 l	14	no	200

Evaluation Criteria

- Extract features from the Key frame and another frame.
- Matching criterion: Same as before except no scale
- Compute:
  r = Best / 2nd Best < T
- Compute ROC and AUC.
- Use AUC as performance measure (average over scenes).

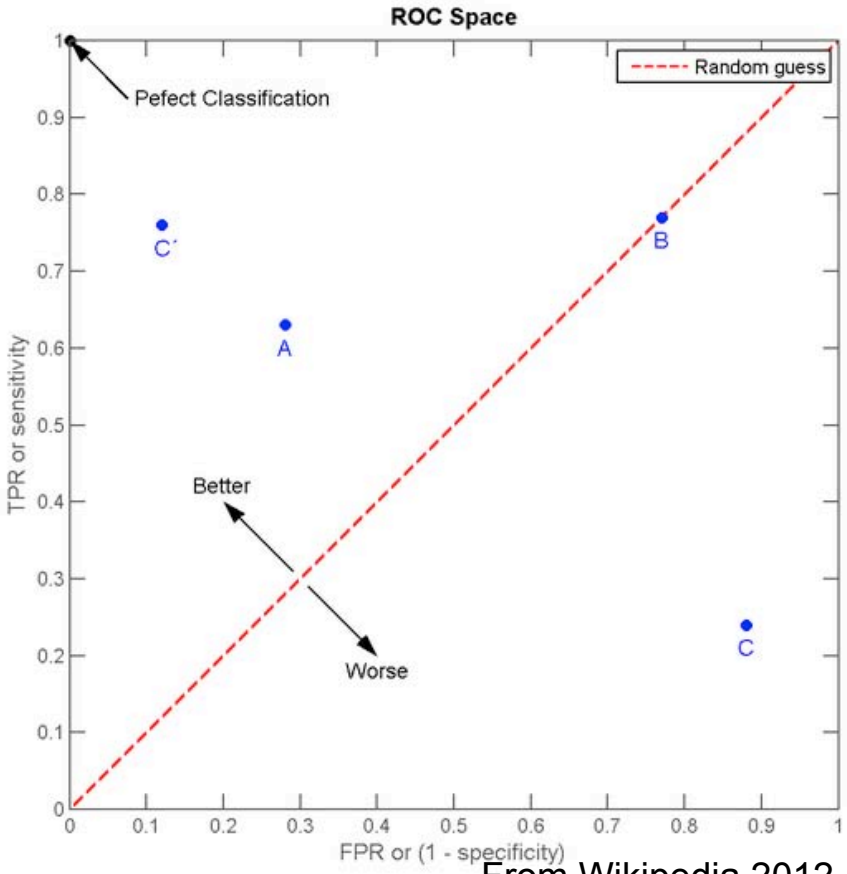




Aside: ROC and AUC?

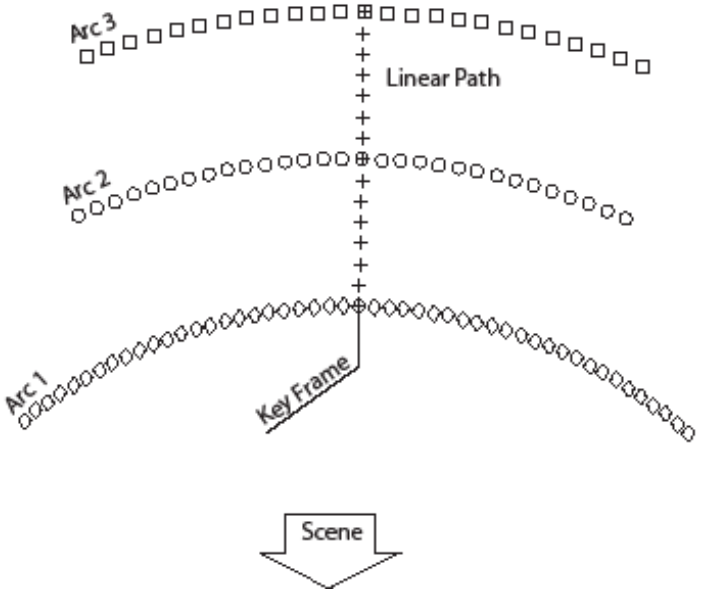
- Receiver operating characteristic (ROC):
  - TPR = TP / P (Recall)
  - FPR = FP / N
- Area under the ROC curve (AUC):
  - AUC close to 1 is good
  - The probability of a correct match

		actual value		
		p	***n***	**total**
prediction outcome	***p'***	True Positive	False Positive	P'
	n'	False Negative	True Negative	N'
	total	P	N	

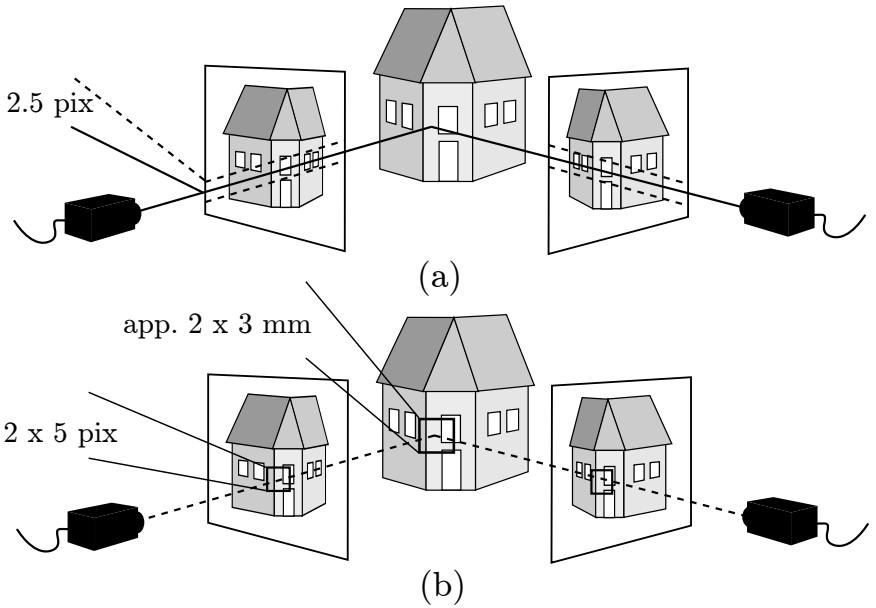


From Wikipedia 2012

Evaluation Criteria



- Extract features from the Key frame and another frame.
- Matching criterion: Same as before except no scale
- Compute:
  r = Best / 2nd Best
- Compute ROC and AUC.
- Use AUC as performance measure (average over scenes).



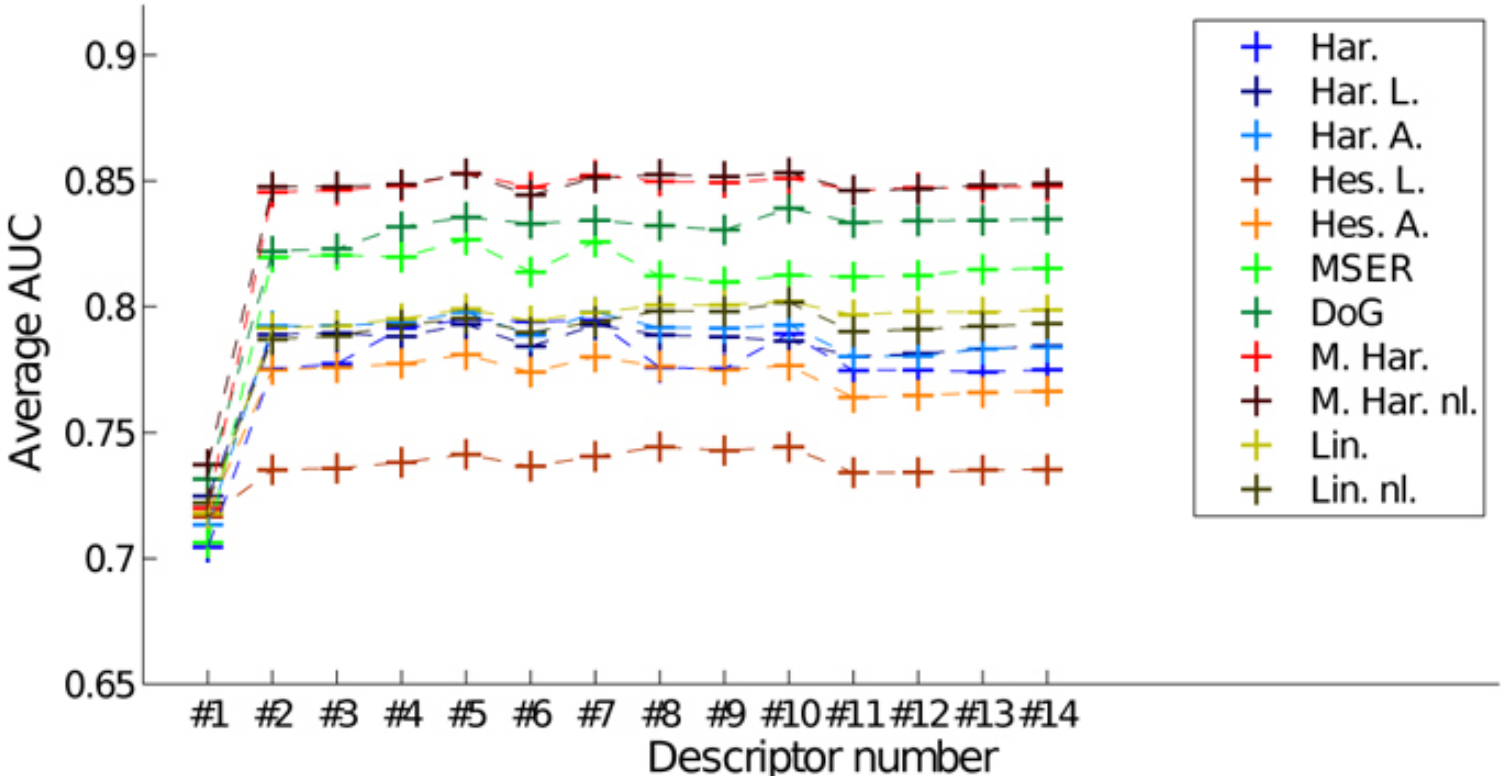
Results at a Glance

Table 11 Overall average performance of descriptors and detectors for all experiments. A difference of approximately 0.001 is statistical significant on a 5% level whereas a difference of approximately 0.002 is significant on a 0.05% level.

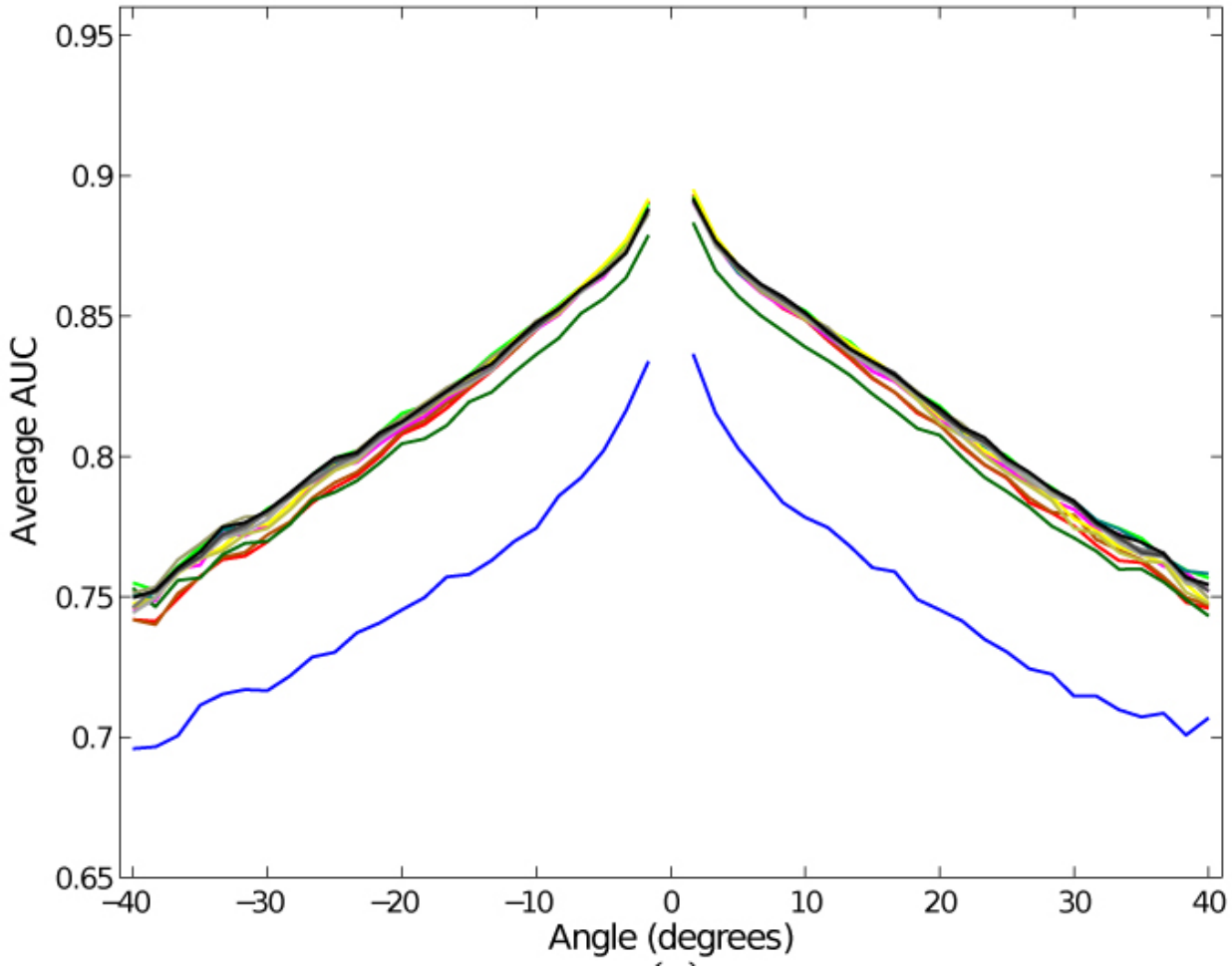
	Har	HarL	HarA	HesL	HesA	MSER	DoG	MHar	MHarNL	Lin	LinNL
Raw patch	0.704	0.725	0.713	0.716	0.718	0.706	0.731	0.720	0.737	0.718	0.722
SIFT gray	0.775	0.789	0.792	0.735	0.775	0.820	0.822	0.846	0.848	0.791	0.787
SIFT RGB bin	0.777	0.790	0.793	0.736	0.776	0.821	0.823	0.846	0.848	0.792	0.788
SIFT RGB	0.791	0.788	0.793	0.738	0.777	0.820	0.832	0.848	0.849	0.795	0.793
Opponent SIFT	0.795	0.793	0.798	0.741	0.781	0.827	0.836	**0.853**	**0.853**	0.799	0.795
CSIFT	0.794	0.784	0.789	0.737	0.774	0.814	0.833	0.848	0.844	0.794	0.790
Gaussian opponent SIFT	0.795	0.793	0.798	0.741	0.780	0.826	0.834	0.852	0.851	0.798	0.794
Hist. eq. SIFT	0.776	0.789	0.792	0.744	0.776	0.812	0.832	0.850	0.852	0.801	0.798
Hist. eq. SIFT RGB bin	0.775	0.788	0.791	0.743	0.775	0.810	0.831	0.849	0.852	0.801	0.798
Hist. eq. SIFT RGB	0.789	0.786	0.793	0.744	0.777	0.812	0.839	0.851	**0.853**	0.802	0.802
DAISY 1-6-6 s	0.775	0.780	0.780	0.734	0.764	0.812	0.833	0.846	0.846	0.797	0.790
DAISY 1-6-6 l	0.775	0.781	0.780	0.734	0.765	0.812	0.834	0.847	0.847	0.798	0.791
DAISY 1-8-8-8 s	0.774	0.783	0.783	0.735	0.766	0.815	0.834	0.847	0.848	0.798	0.792
DAISY 1-8-8-8 l	0.775	0.784	0.784	0.735	0.766	0.815	0.835	0.848	0.849	0.799	0.793

Results at a Glance, Detectors

NB: Not Recall Rate



Results at a Glance, Descriptors



Outline of Findings

- Multiscale Harris is the top performer.
- With detectors the devil is in the details.
- All SIFT based descriptors are almost created equal.
- Light is very challenging.
- On statistical significance…

Quote on Statistical Significance

"A question often overlooked by the computer vision community when comparing results on a given dataset is whether the difference in performance of two methods is statistically significant."

M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman
"The PASCAL Visual Object Classes (VOC) Challenge"
International Journal of Computer Vision, Volume 88, Number 2,
page 303--338, 2010

Summary

- Interest point descriptors:
  - SIFT and variants
- Comparison of interest point descriptors

Literature

Reading material:

- Lowe IJCV 2004 (Sec. 5 – 7.1)
- Dahl-Aanæs-Pedersen 3DIMPVT 2011

Additional material:

  - Larsen et al ECCV 2012

Lets have a break and then work with the assignment

Mandatory assignment 1: Feature extraction

- Build and experiment with interest point detectors
- Find matching points between two images