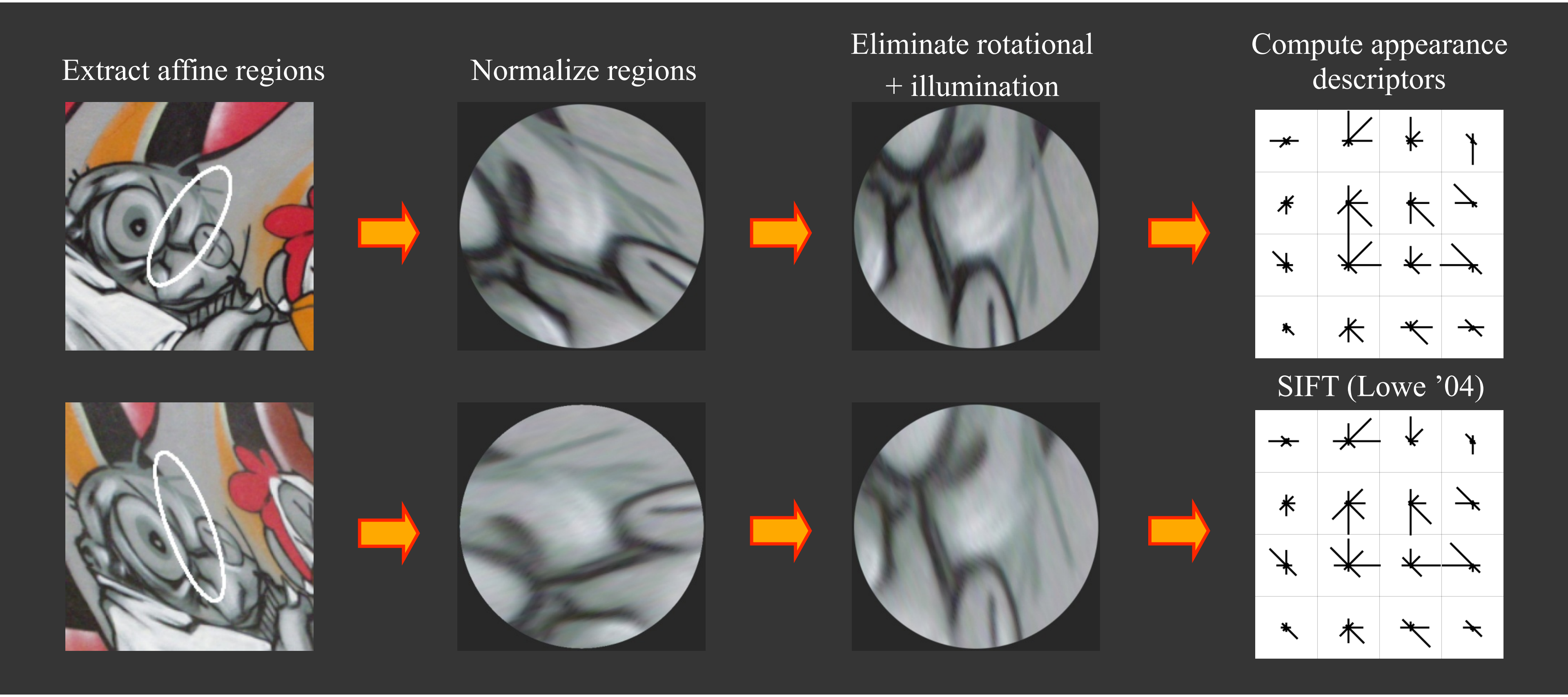
Descriptors

Cordelia Schmid
Tinne Tuytelaars
Krystian Mikolajczyk

Overview – descriptors

- **Introduction**
- Modern descriptors
- Comparison and evaluation

Descriptors



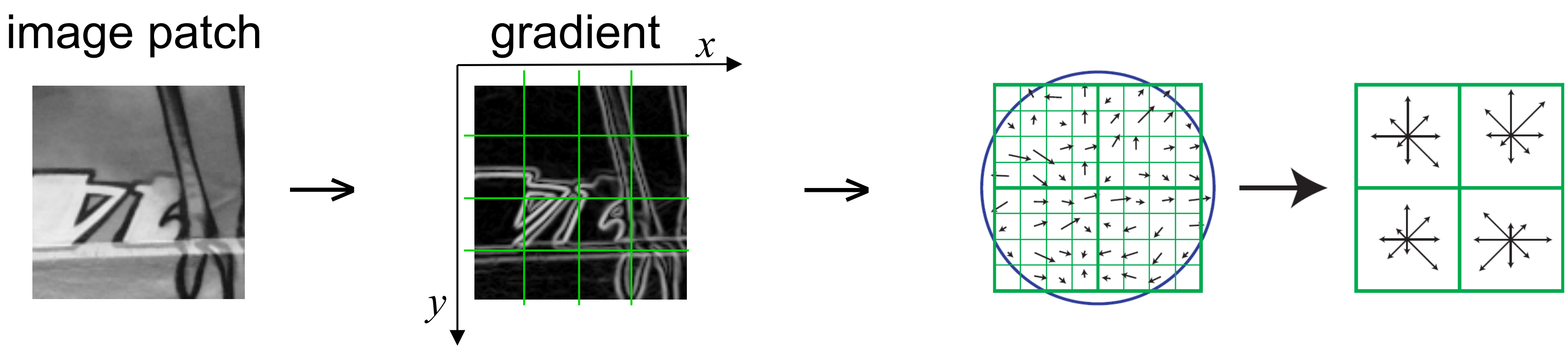
Descriptors - history

- Normalized cross-correlation (NCC) [~ 60s]
- Gaussian derivative-based descriptors
  - Differential invariants [Koenderink and van Doorn'87]
  - Steerable filters [Freeman and Adelson'91]
- Moment invariants [Van Gool et al.'96]
- SIFT [Lowe'99]
- Shape context [Belongie et al.'02]
- Gradient PCA [Ke and Sukthankar'04]

- SURF descriptor [Bay et al.'08]
- DAISY descriptor [Tola et al.'08, Windler et al'09]
- …….

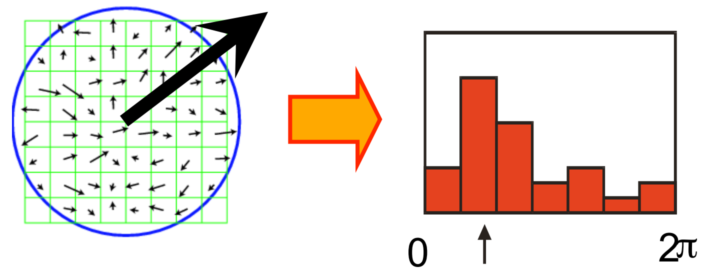
SIFT descriptor [Lowe'99]

- Spatial binning and binning of the gradient orientation
- 4x4 spatial grid, 8 orientations of the gradient, dim 128
- Soft-assignment to spatial bins
- Normalization of the descriptor to norm one (robust to illumination)
- Comparison with Euclidean distance

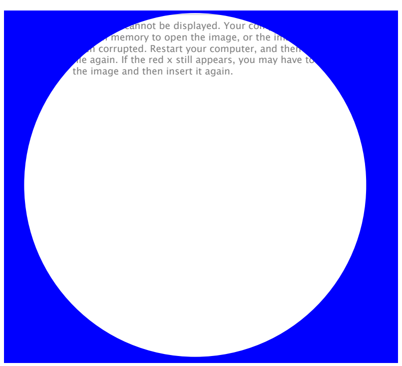


Local descriptors - rotation invariance

- Estimation of the dominant orientation
  - extract gradient orientation
  - histogram over gradient orientation
  - peak in this histogram



- Rotate patch in dominant direction

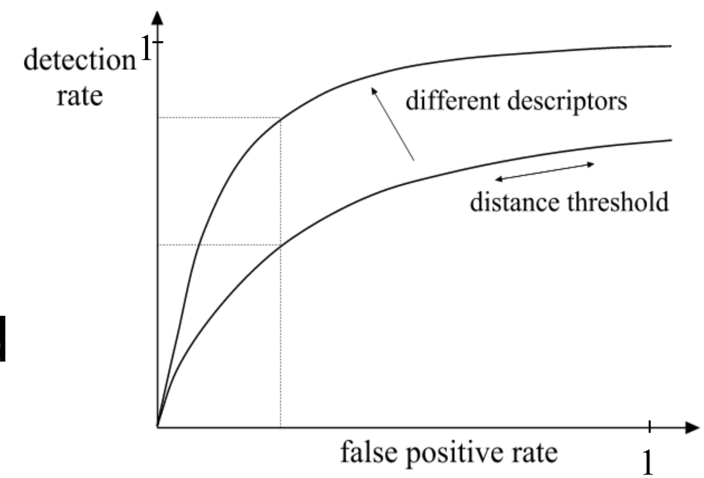


- Plus: invariance
- Minus: less discriminant, additional noise

Evaluation

- Descriptors should be
  - Distinctive (! importance of the measurement region)
  - Robust to changes on viewing conditions as well as to errors of the detector
  - Compactness, speed of computation

- Detection rate (recall)
  - #correct matches / #correspondences
- False positive rate
  - #false matches / #all matches
- Variation of the distance threshold
  - distance (d1, d2) < threshold

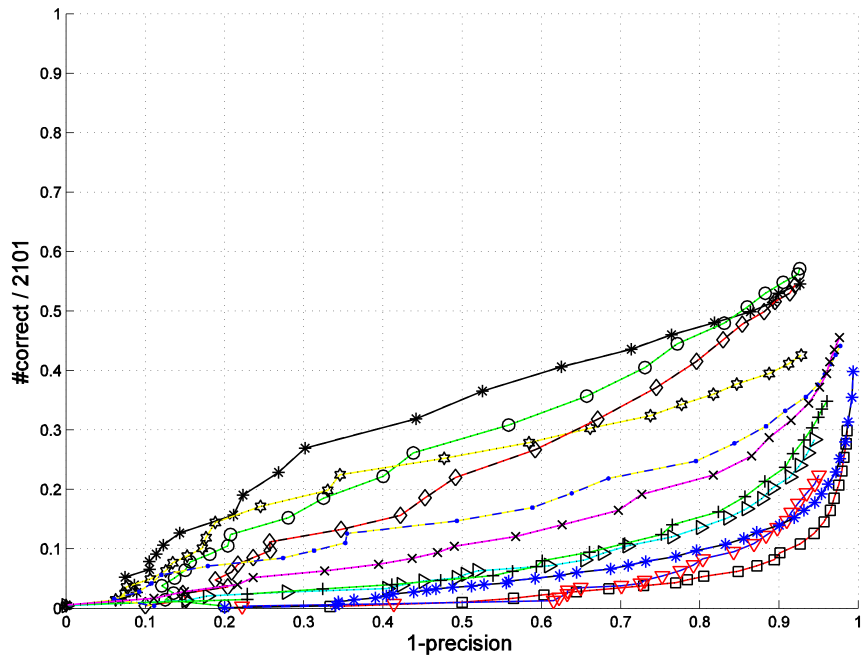


[K. Mikolajczyk & C. Schmid, PAMI'05]

Viewpoint change (60 degrees)

sift
esift
gradient pca
shape context
cross correlation
har-aff esift
steerable filters
gradient moments
complex filters

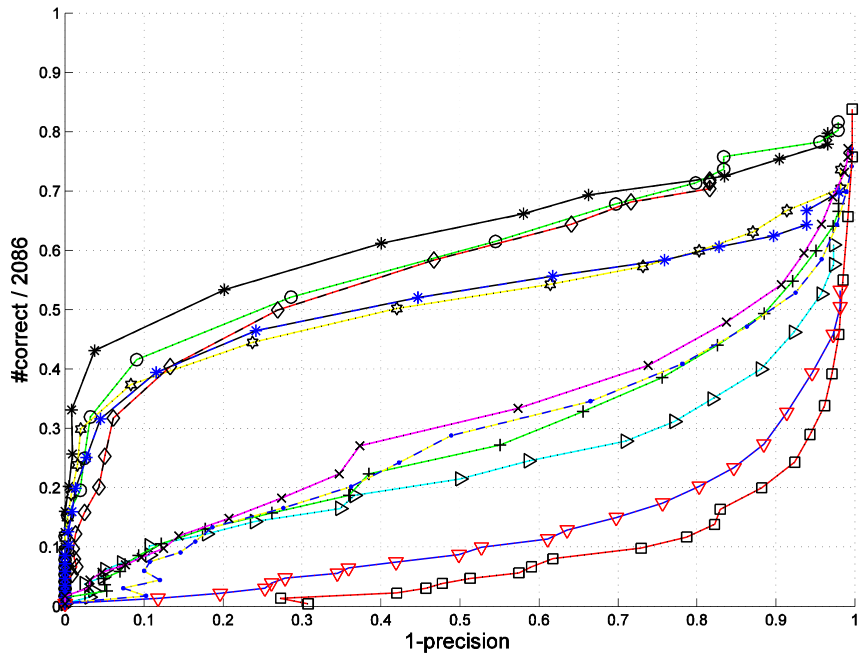
Detector:
Hessian-Affine



Scale change (factor 2.8)

sift
esift
gradient pca
shape context
cross correlation
har-aff esift
steerable filters
gradient moments
complex filters

Detector:
Hessian-Affine



Evaluation - conclusion

- SIFT based descriptors perform best
- Significant difference between SIFT and low dimension descriptors as well as cross-correlation
- Robust region descriptors better than point-wise descriptors
- Performance of the descriptor is relatively independent of the detector

Recent extensions to SIFT

- Color SIFT [Sande et al. 2010]

- Normalizing SIFT with square root transformation

Retrieval Method	SIFT		RootSIFT	
	Ox5k	Ox105k	Ox5k	Ox105k
Philbin *et al.* [23]: tf-idf ranking	0.636	0.515	0.683	0.581
Philbin *et al.* [23]: tf-idf with spatial reranking	0.672	0.581	0.720	0.642

[Arandjelovic et Zisserman'12]

Descriptors – dense extraction

- Many conclusions for descriptors applied to sparse detectors also hold for dense extraction
  - For example normalizing SIFT with the square root improves in image retrieval and classification

- Image retrieval: sparse versus dense

SIFT + Fisher vector for retrieval on the Holidays dataset

Hessian-Affine	MAP = 0.54
Dense	MAP = 0.62

Overview – descriptors

- Introduction
- **Modern descriptors**
- Comparison and evaluation

Overview

- Introduction
- **Modern descriptors**
- Comparison and evaluation

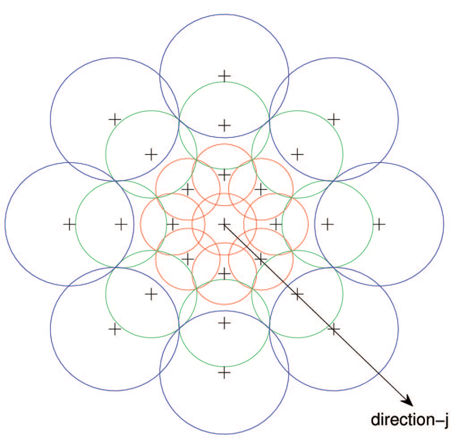
Modern descriptors

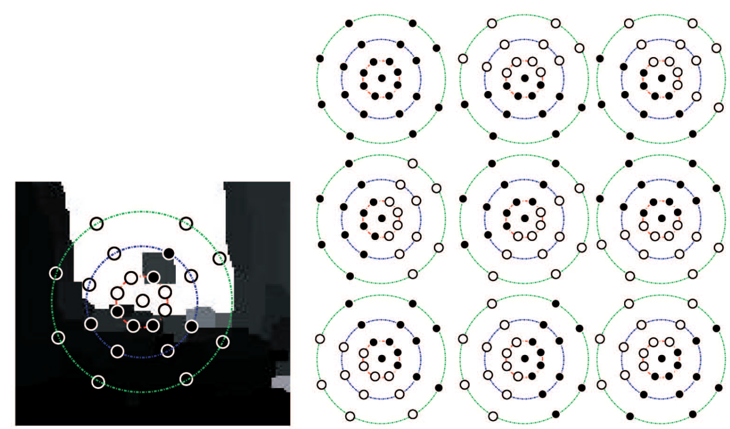
- Efficient descriptors
- Compact binary descriptors
- More robust descriptors
- Learned descriptors

DAISY

- Optimized for dense sampling
- Log-polar grid
- Gaussian smoothing
- Dealing with occlusions

Citations: 150 (2012)





(a) (b)

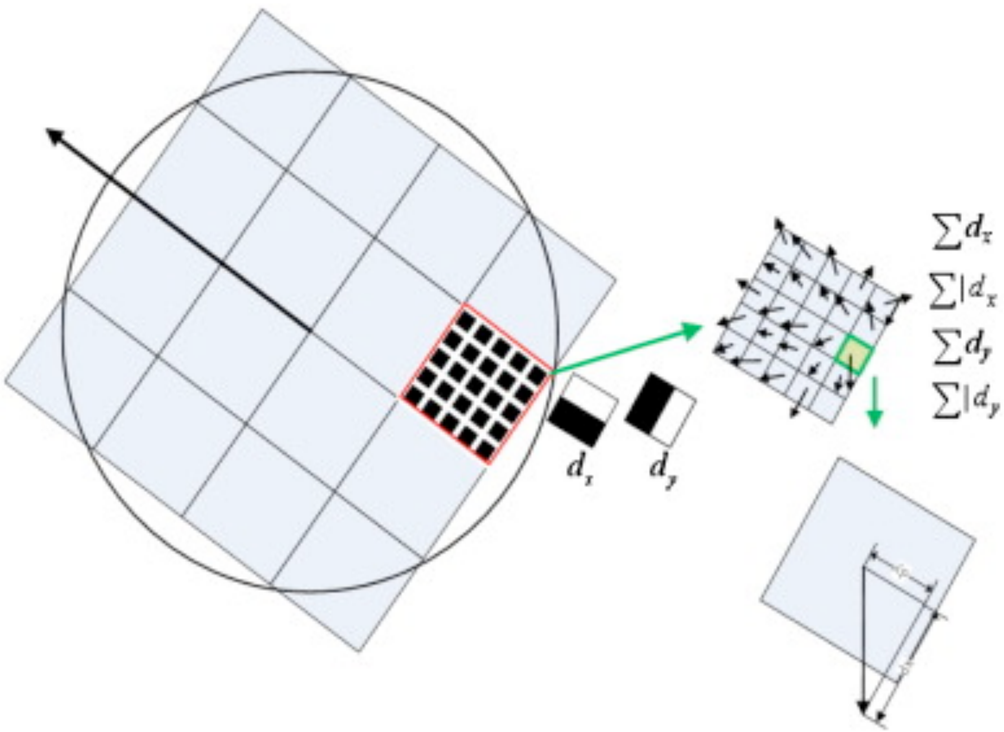
Engin Tola, Vincent Lepetit, and Pascal Fua, DAISY: An Efficient Dense Descriptor Applied to Wide-Baseline Stereo, TPAMI 32(5), 2010.

SURF:

Speeded Up Robust Features

- Approximate derivatives with Haar wavelets
- Exploit integral images

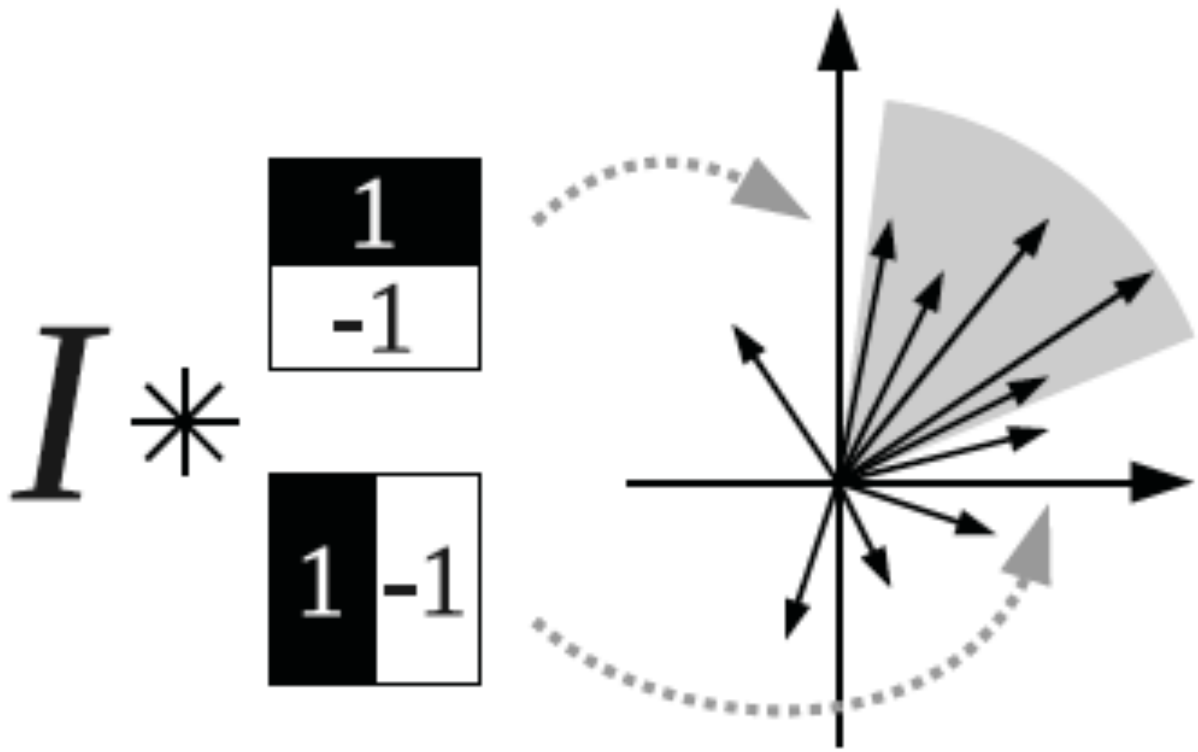
Citations: 4500 (2012)



Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346--359, 2008

SURF:
Speeded Up Robust Features

- Orientation assignment



U-SURF:
Upright SURF

- Rotation invariance is often not needed.

Don’t use more invariance
than needed for a given application

- Orientation estimation takes time.
- Orientation estimation is often a source of errors.

Beyond the classics

- Efficient descriptors
- Compact binary descriptors
- More robust descriptors
- Learned descriptors

Fast and compact descriptors

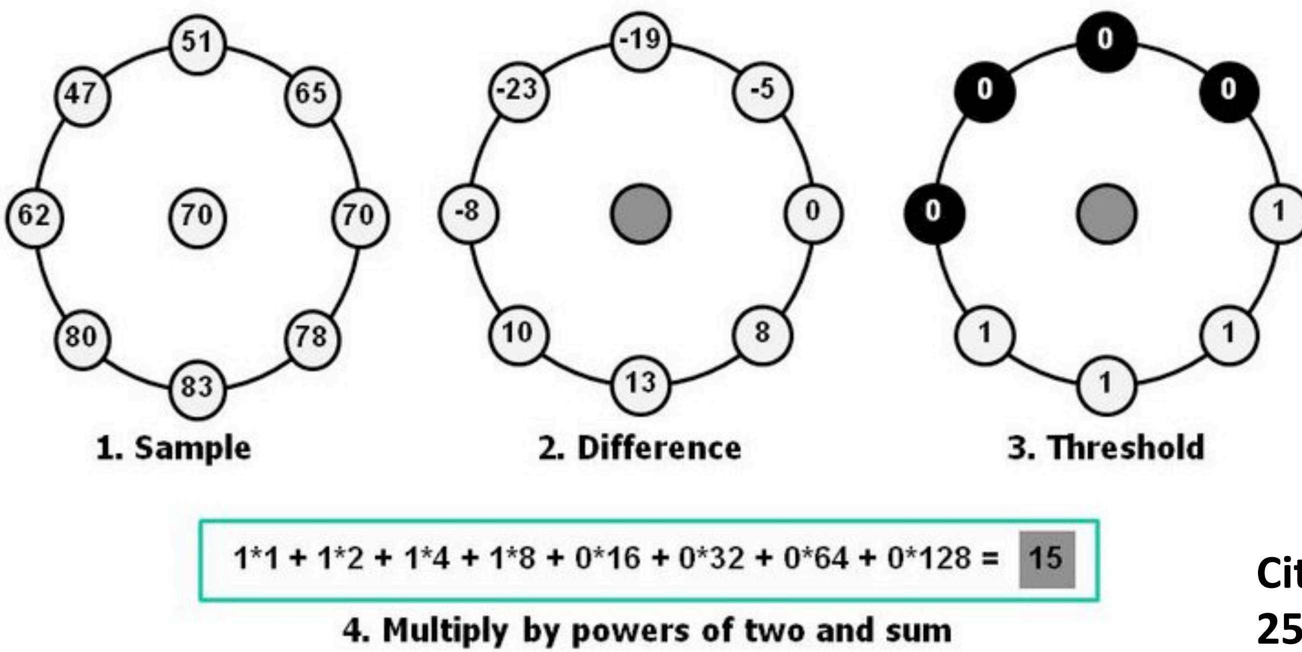
- (Very) large scale applications
  - > memory issues, computation time issues
- Mobile phone applications

Fast and compact descriptors

- Binary descriptors
- Comparison of pairs of intensity values
  - LBP
  - BRIEF
  - ORB
  - BRISK

LBP:
Local Binary Patterns

- First proposed for texture recognition in 1994.

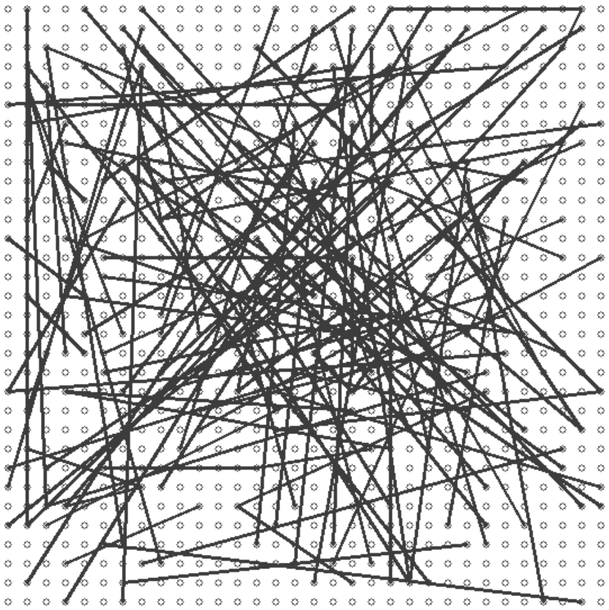


Citations: 2500 (2012)

T. Ojala, M. Pietikäinen, and D. Harwood (1994), "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions", ICPR 1994, pp.582-585.
M Heikkilä, M Pietikäinen, C Schmid, Description of interest regions with LBP, Pattern recognition 42 (3), 425-436

BRIEF:
Binary Robust Independent Elementary Features

- Random selection of pairs of intensity values.
- Fixed sampling pattern of 128, 256 or 512 pairs.
- Hamming distance to compare descriptors (XOR).



Citations: 149 (2012)

M. Calonder, V. Lepetit, C. Strecha, P. Fua, BRIEF: Binary Robust Independent Elementary Features, 11th European Conference on Computer Vision, 2010.

D-BRIEF:
Discriminative BRIEF

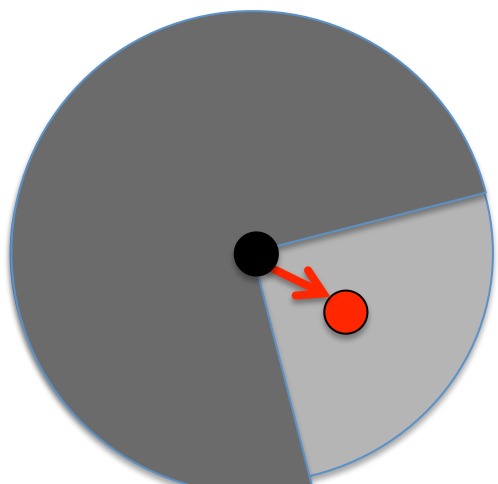
- Learn linear projections that map image patches to a more discriminative subspace
- Exploit integral images

T. Trzcinski and V. Lepetit, Efficient Discriminative Projections for Compact Binary Descriptors
European Conference on Computer Vision (ECCV) 2012

ORB:
Oriented FAST and Rotated BRIEF

- Add rotation invariance to BRIEF
- Orientation assignment based on the intensity centroid

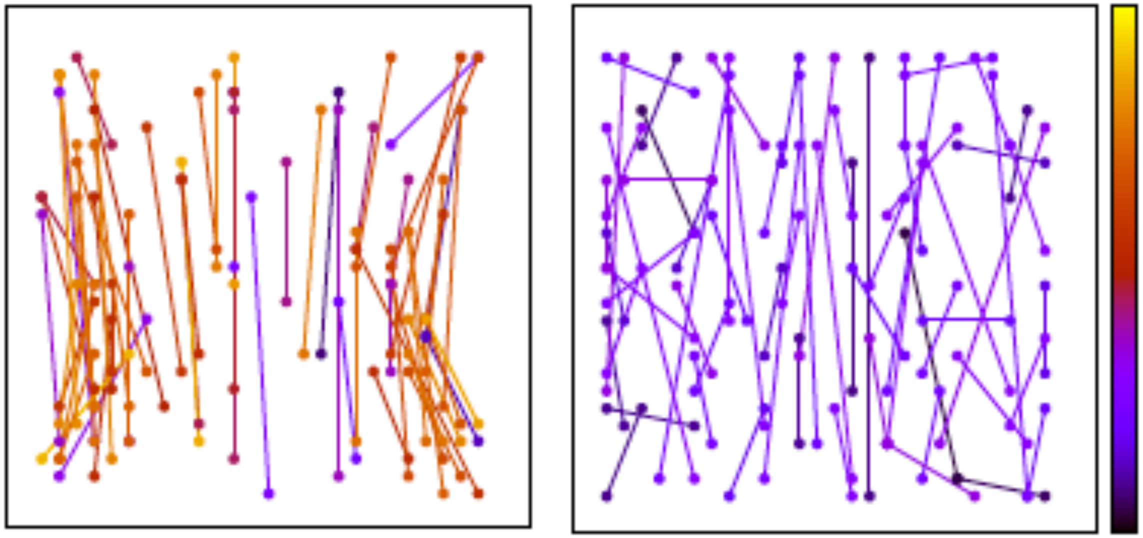
Citations: 43 (2012)



Ethan Rublee, Vincent Rabaud, Kurt Konolige, Gary Bradski, ORB: an efficient alternative to SIFT or SURF, ICCV 2011

ORB:
Oriented FAST and Rotated BRIEF

- Select a good set of pairwise comparisons: minimize correlation under various orientation changes



High variance　　　　High variance + uncorrelated

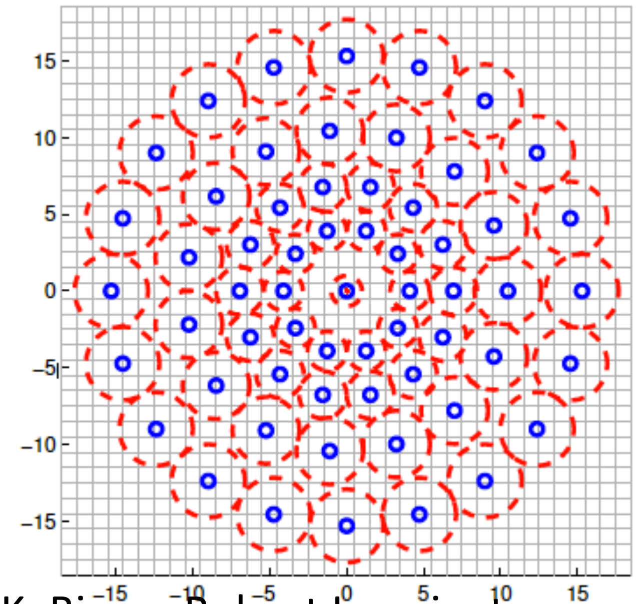
BRISK:
Binary Robust Invariant Scalable Keypoints

Citations: 20 (2012)

- Regular grid
- Orientation assignment based on dominant gradient direction (using long-distance pairs)

$$\mathbf{g} = \begin{pmatrix} g_x \\ g_y \end{pmatrix} = \frac{1}{L} \cdot \sum_{(\mathbf{p}_i, \mathbf{p}_j) \in \mathcal{L}} \mathbf{g}(\mathbf{p}_i, \mathbf{p}_j).$$

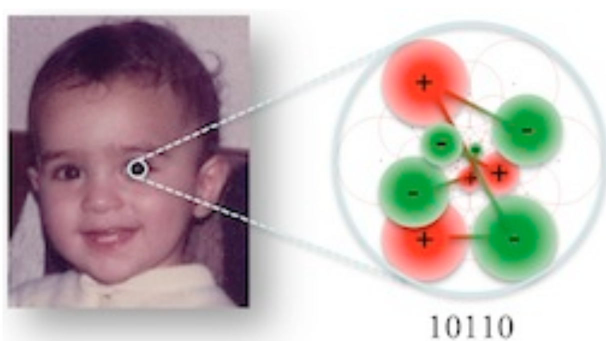
- 512 bit descriptor based on short-distance pairs

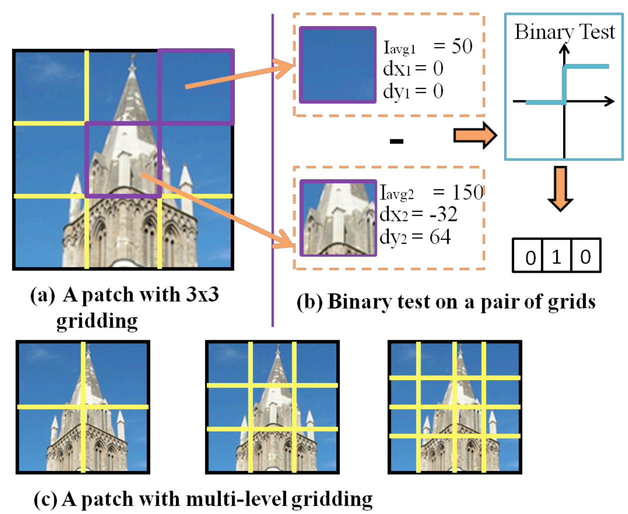


Stefan Leutenegger, Margarita Chli and Roland Y. Siegwart, BRISK: Binary Robust Invariant Scalable Keypoints, ICCV 2011

Various others

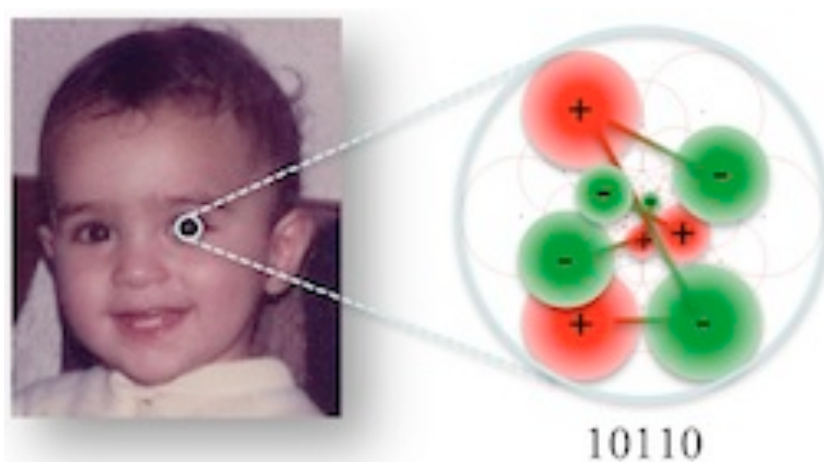
- FREAK: Fast Retina Keypoint
- CARD: Compact and Realtime Descriptor
- LDB: Local Difference Binary

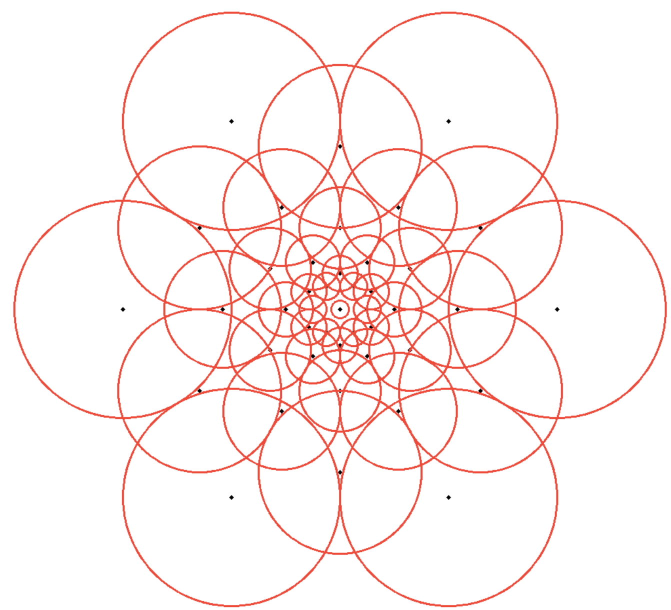




FREAK:
Fast Retina Keypoint

- Inspired by the human visual system





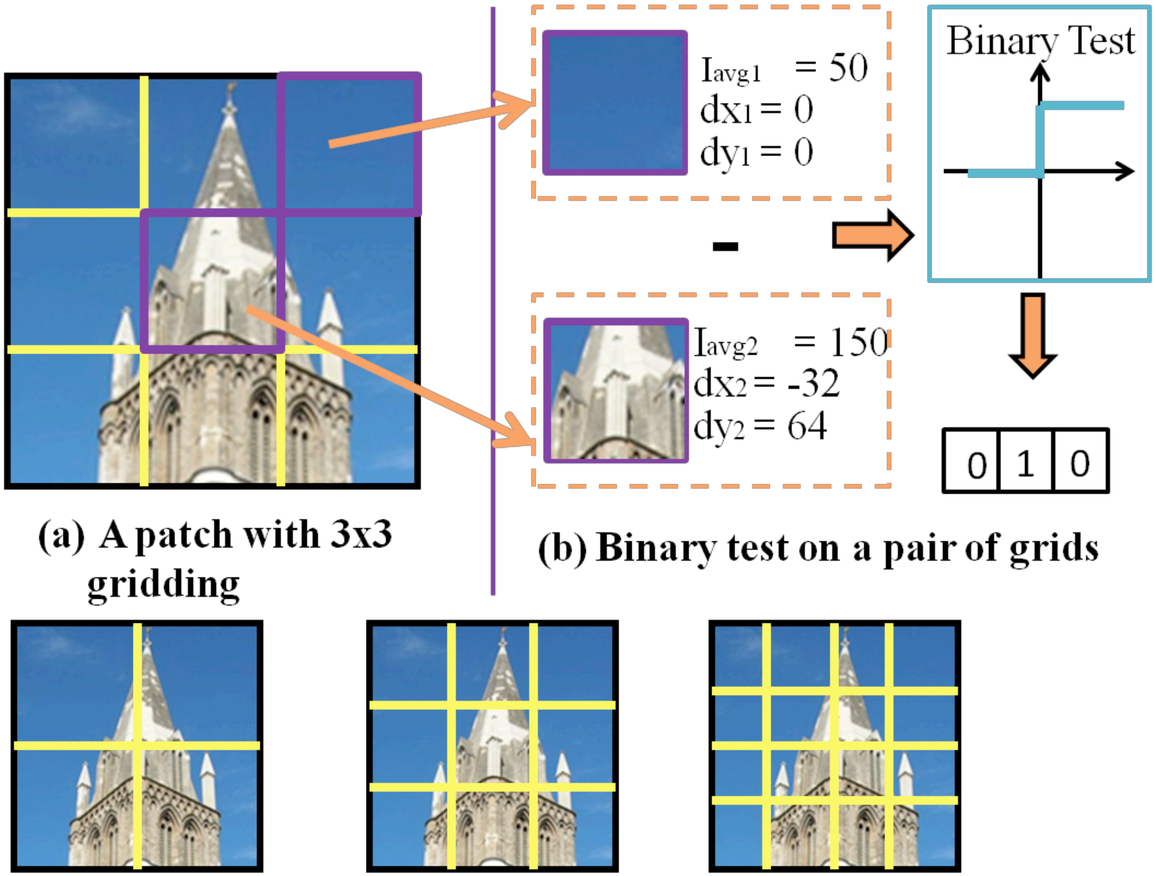
A. Alahi, R. Ortiz, and P. Vandergheynst. FREAK: Fast Retina Keypoint. In IEEE Conference on Computer Vision and Pattern Recognition 2012.

CARD:
Compact and Realtime Descriptor

- Look up tables
- learning-based sparse hashing

LDB:

Local Difference Binary



(a) A patch with 3x3 gridding

(b) Binary test on a pair of grids

(c) A patch with multi-level gridding

Xin Yang and Kwang-Ting Cheng, LDB: An Ultra-Fast Feature for Scalable Augmened Reality on Mobile Devices, International Symposium on Mixed and Augmented Reality 2012 (ISMAR 2012)

Beyond the classics

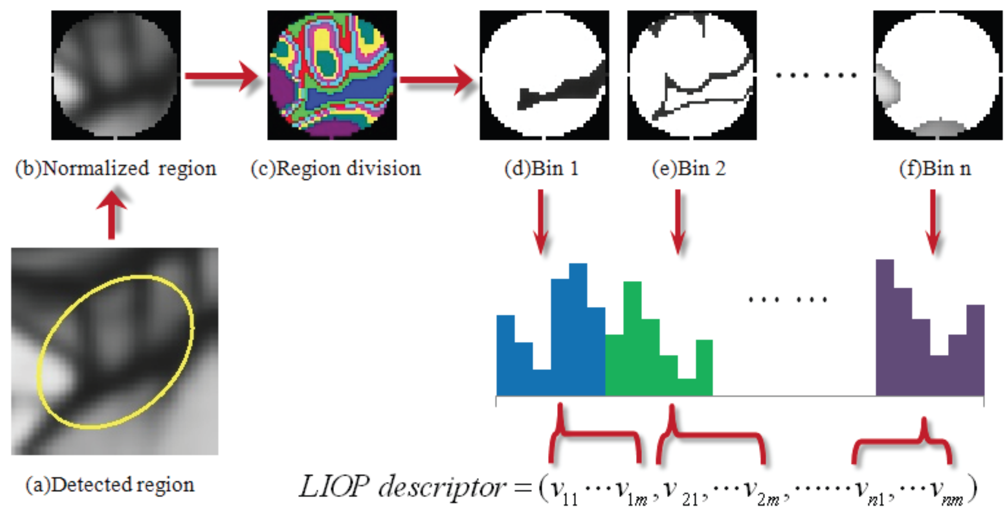
- Efficient descriptors
- Compact binary descriptors
- More robust descriptors
- Learned descriptors

LIOP:

Local Intensity Order Pattern for Feature Description

(and predecessors MROGH and MRRID)

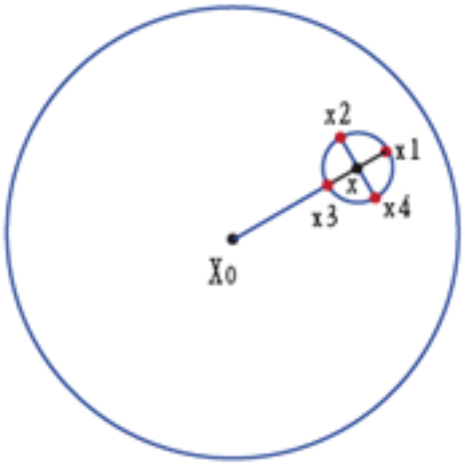
- Robustness to monotonic intensity changes
- Data-driven division into cells



Zhenhua Wang Bin Fan Fuchao Wu, Local Intensity Order Pattern for Feature Description, ICCV 2011.

LIOP:

Local Intensity Order Pattern for Feature Description



(a) Original patch

For a point x:

$$P(x) = (I(x_1), I(x_2), I(x_3), I(x_4))$$
$$= (86, 217, 152, 101)$$

$$\gamma(P(x)) = (1,4,3,2)$$

π	$Ind(\pi)$
1,2,3,4	1
1,2,4,3	2
1,3,2,4	3
1,3,4,2	4
1,4,2,3	5
1,4,3,2	6
2,1,3,4	7
2,1,4,3	8
. . .	. . .
4,3,1,2	23
4,3,2,1	24

(d)Index table of Π^4

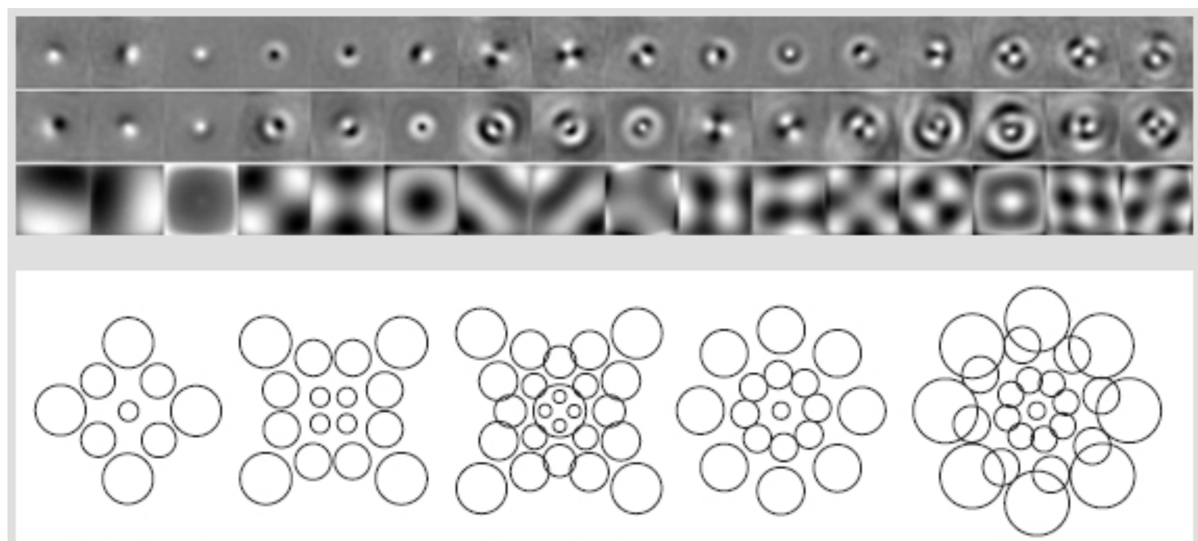
Beyond the classics

- Efficient descriptors
- Compact binary descriptors
- More robust descriptors
- Learned descriptors

Winder & Brown

- Learn configuration and other parameters from training data obtained from 3D reconstructions

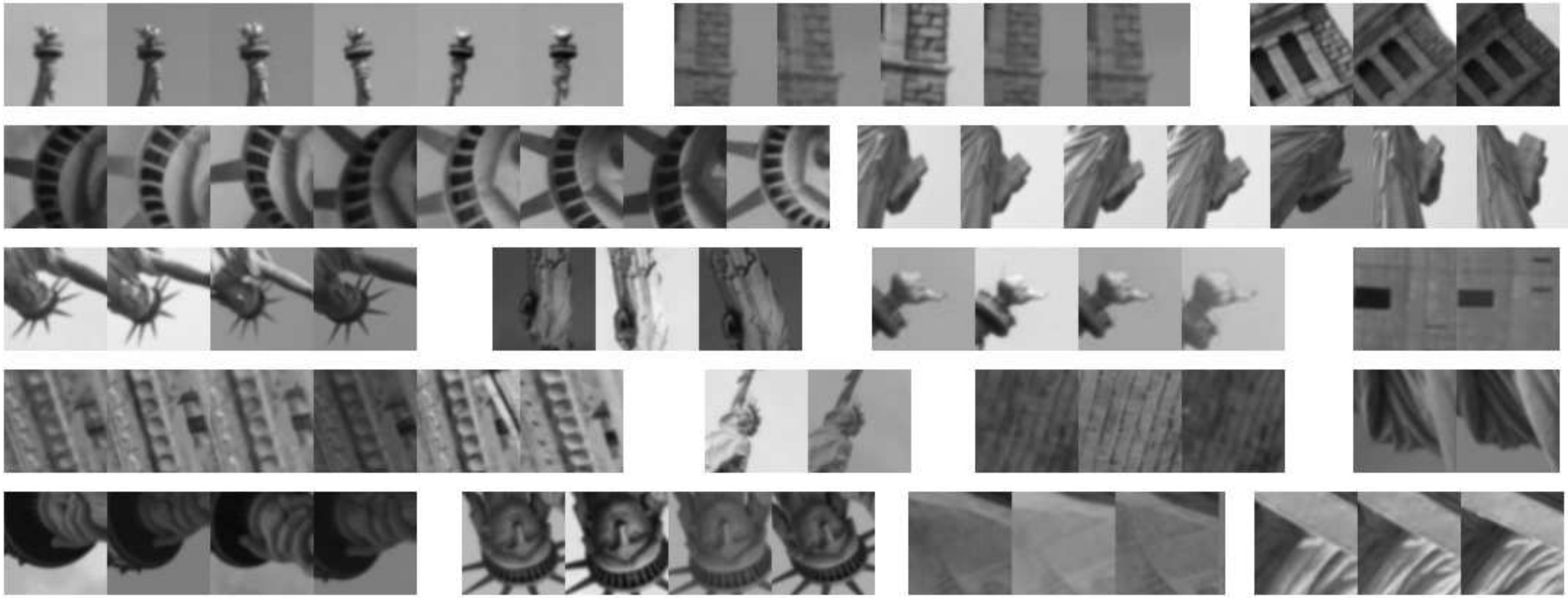
Citations: 194 (2012)



M. Brown, G. Hua and S. Winder, Discriminant Learning of Local Image Descriptors.. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2010.

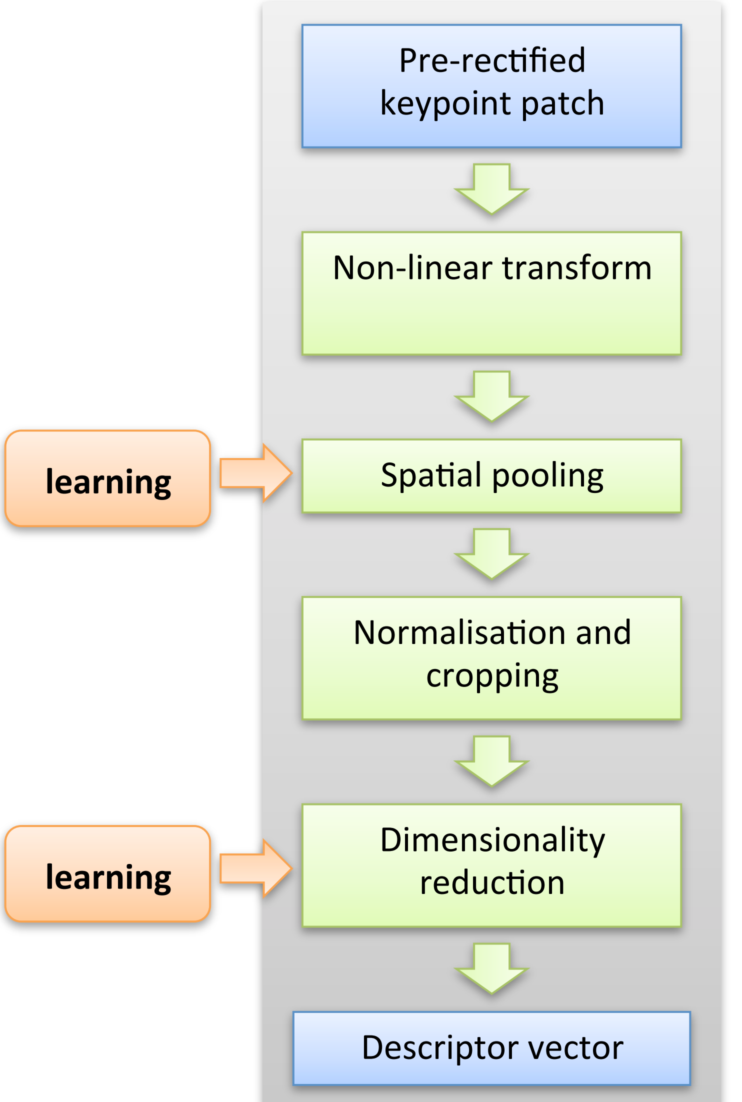
Winder & Brown

- Training data = set of corresponding image patches



Descriptor Learning Using Convex Optimisation

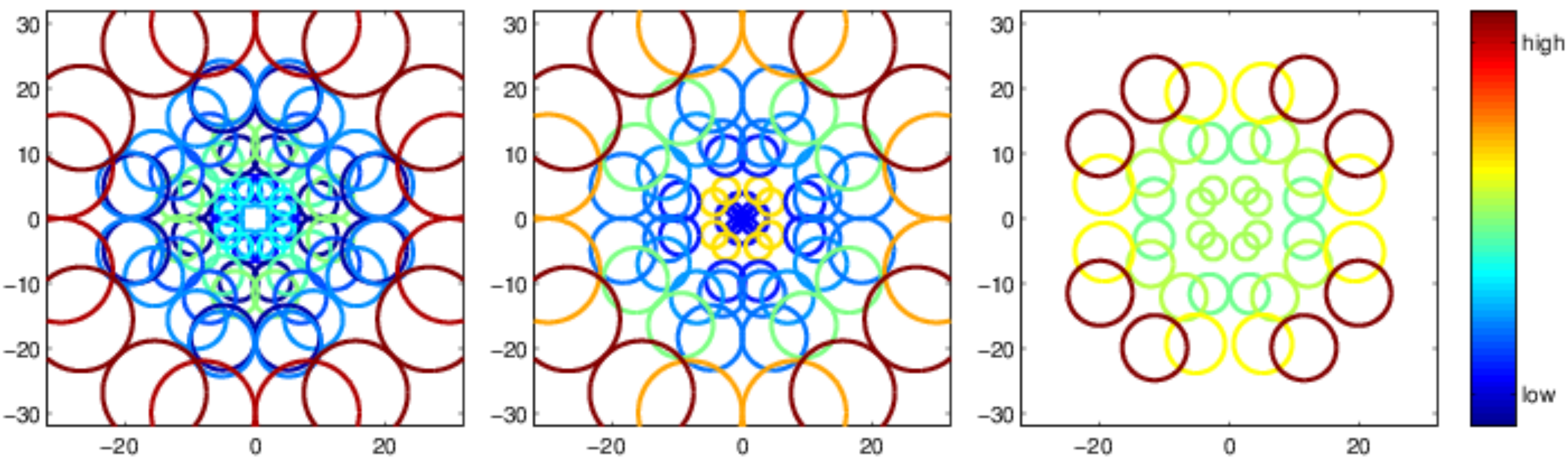
- Convex learning of
  - spatial pooling regions
  - dimensionality reduction
- Learning from very weak supervision



K. Simonyan *et al.*, Descriptor Learning Using Convex Optimisation, ECCV 2012

Learning Spatial Pooling Regions

- **Selection from a large pool using L^1 regularisation**
- Soft-margin constraints:
  - squared L^2 distance between descriptors of matching feature pairs should be smaller than that of non-matching pairs
- Convex objective (optimised with a proximal method)



768-D **576-D** **320-D**

pooling region configurations learnt with different levels of sparsity

Overview

- Introduction
- Modern descriptors
- **Comparison and evaluation**

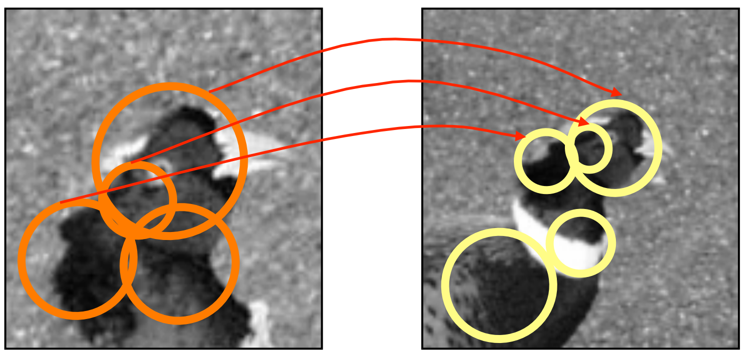
Setting up an evaluation

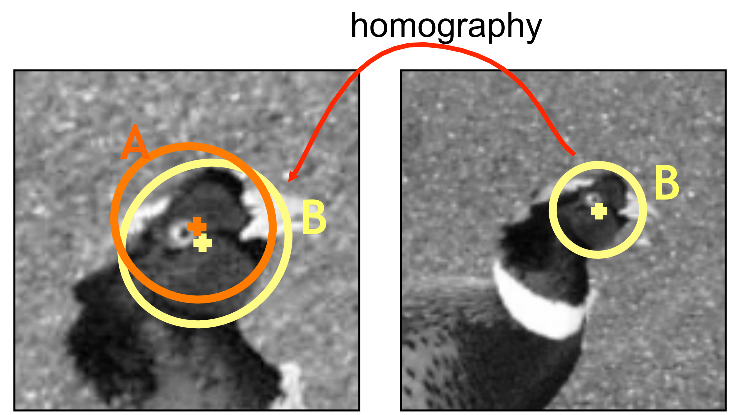
- Which problem? Performance in different application/niches may vary significantly.
  - Category recognition,
  - Matching,
  - Retrieval

- What dataset?
  - Pascal VOC 2007
  - Oxford image pairs
  - Oxford - Paris buildings

- Protocol and criteria?
  - Public dataset,
  - Avoiding risk to over-fitting/optimizing to the data

Detector evaluations





$$precision = \frac{\#correct\ matches}{\#all\ matches}$$

$$recall = \frac{\#correct\ matches}{\#ground\ truth\ correspondences}$$

Two points are correctly matched if
T=40%

$$\frac{A \cap B}{A \cup B} > T$$

Previous Evaluations

- 2D Scene – Homography
  - C. Schmid, R. Mohr, and C. Bauckhage, “Evaluation of interest point detectors,” IJCV, 2000.
  - K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” CVPR, 2003.
  - T. Kadir, M. Brady, and A. Zisserman, “An affine invariant method for selecting salient regions in images,” in ECCV, 2004.
  - K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky,T. Kadir, and L. Van Gool, “A comparison of affine region detectors,” IJCV, 2005.
  - A. Haja, S. Abraham, and B. Jahne, Localization accuracy of region detectors, CVPR 2008
  - T. Dickscheid, FSchindler, Falko, W. Förstner, Coding Images with Local Features, IJCV 2011

Zoom+rotation

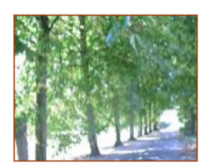
Zoom+rotation

Light

JPEG compression

Blur

Blur

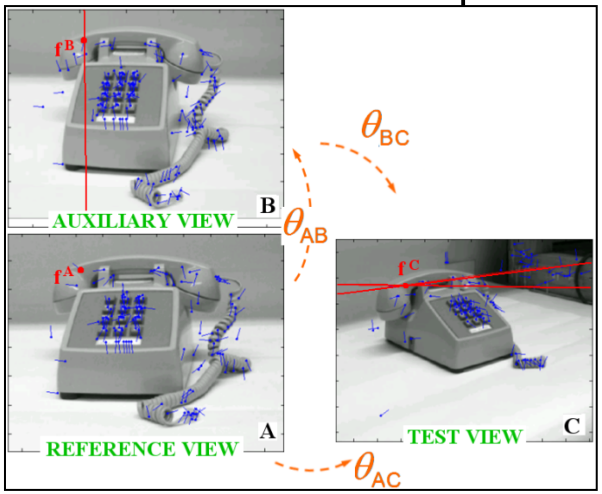


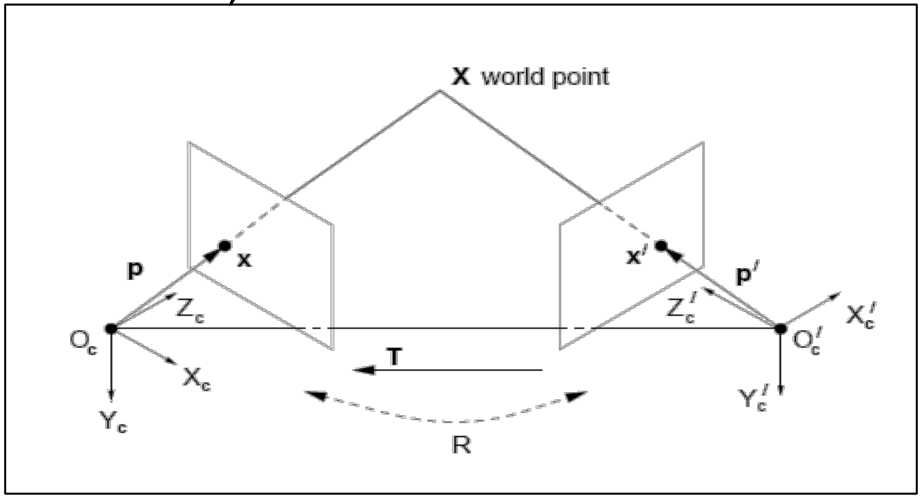
Viewpoint

Viewpoint

Previous Evaluations

- 3D Scene - epipolar constraints
  - F. Fraundorfer and H. Bischof, "Evaluation of local detectors on non-planar, scenes," in AAPR, 2004.
  - P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3D objects," *IJCV*, 2007.
  - S. Winder and M. Brown, "Learning local image descriptors," CVPR, 2007,2009.
  - Dahl, A.L., Aanæs, H. and Pedersen, K.S. (2011): Finding the Best Feature Detector-Descriptor Combination. 3DIMPVT, 2011.

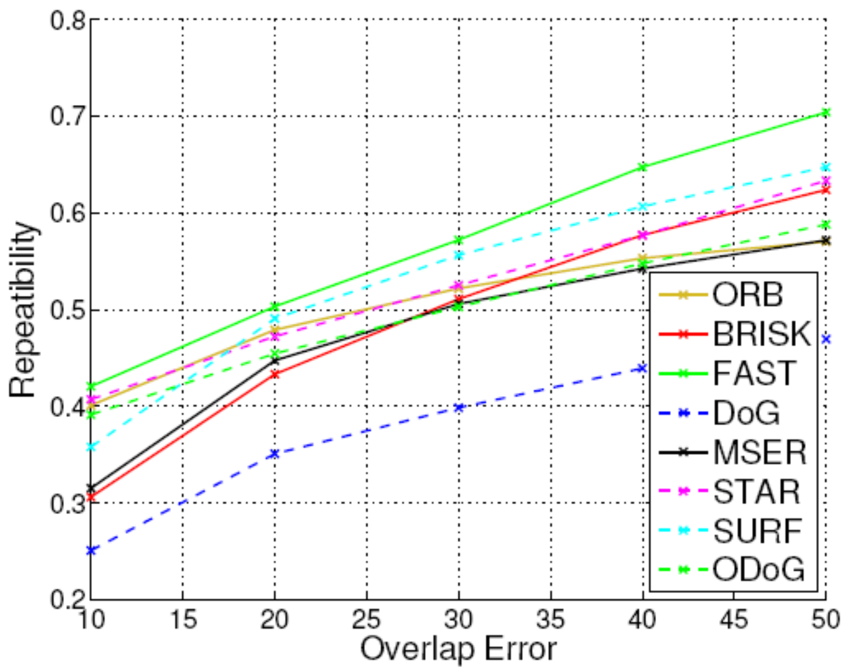




Recent Evaluations

- Recent detectors

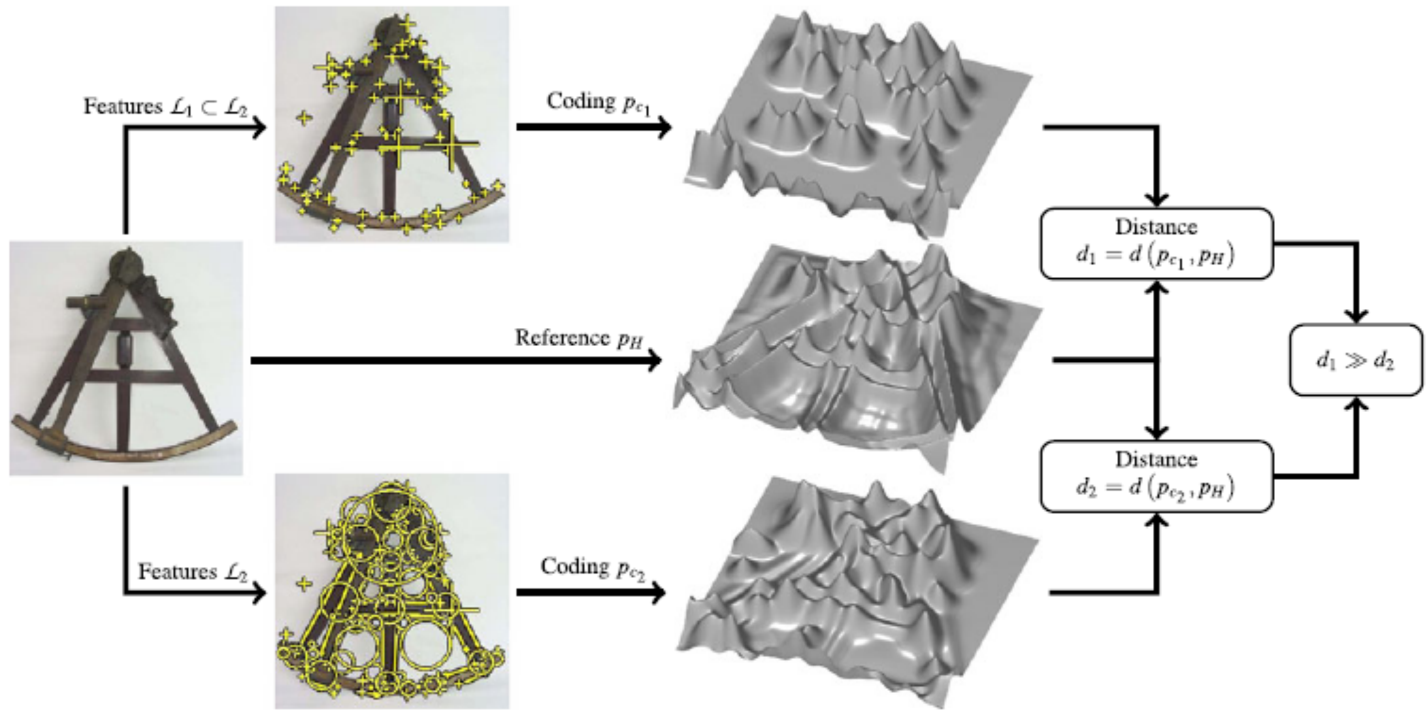
O. Miksik and K. Mikolajczyk, Evaluation of Local Detectors and Descriptors for Fast Feature Matching, ICPR 2012



Detector	Run time [ms.]	Speed-up [-]	# keypoints
SURF	176	1.9	2 911
DoG	338	1.0	1 552
FAST	**2**	169.0	**5 158**
STAR	17	19.9	849
MSER	60	5.6	483
BRISK	**10**	33.8	1 874
ORB	**7**	48.3	594

Recent detector evaluations

- Completeness (coverage), complementarity between detectors

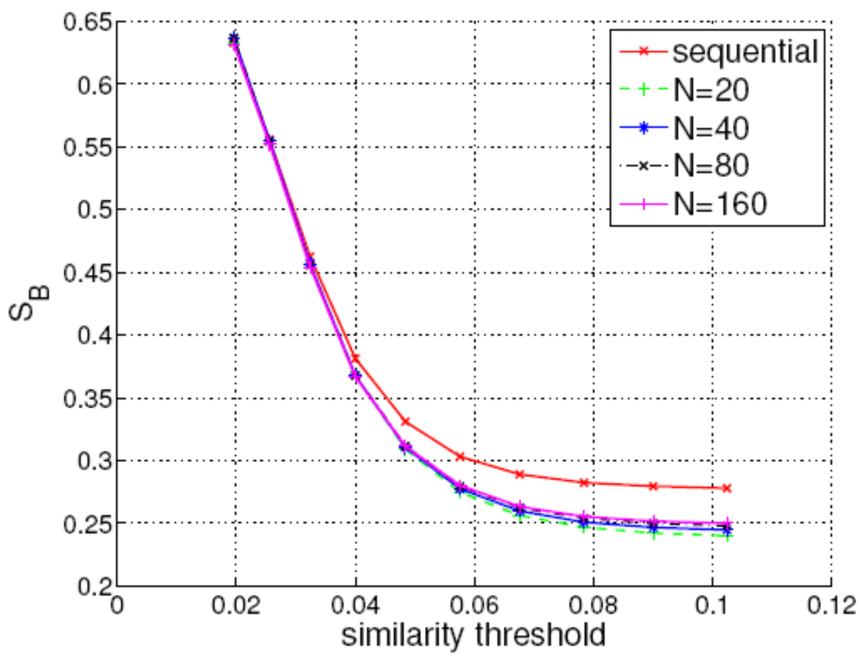


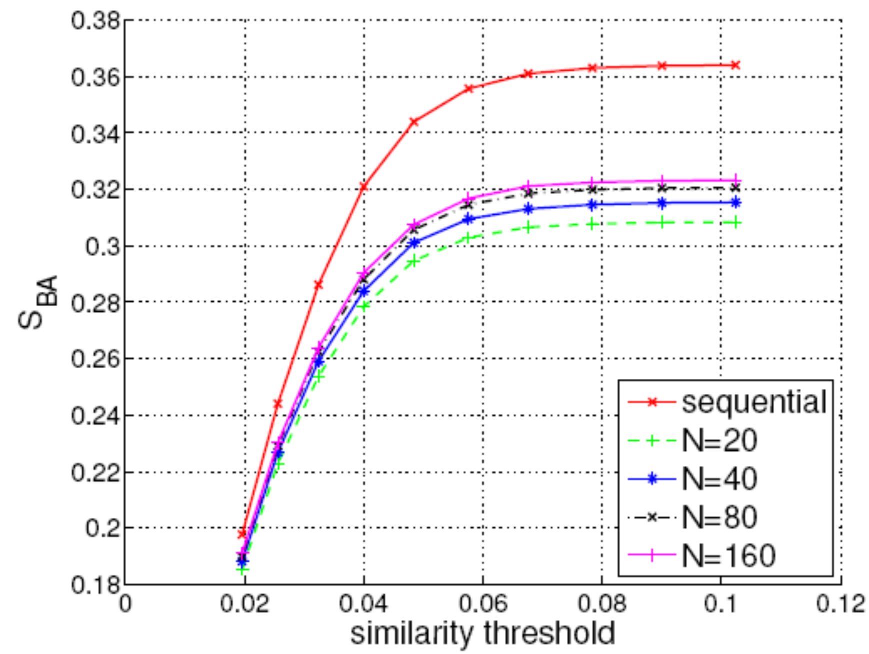
T. Dickscheid, FSchindler, Falko, W. Förstner, Coding Images with Local Features, IJCV 2011

- Completeness, Edgelap (Mikolajczyk), Salient (Kadir), MSER (Matas)
- Complementrity, MSER + SFOP

Descriptor Evaluations

- Matching precision and recall





O. Miksik and K. Mikolajczyk, Evaluation of Local Detectors and Descriptors for Fast Feature Matching, ICPR 2012

Recent Descriptor Evaluations

- **Computation times for the different descriptors for 1000 SURF keypoints**

O. Miksik and K. Mikolajczyk, Evaluation of Local Detectors and Descriptors for Fast Feature Matching, ICPR 2012

J. Heinly E. Dunn, J-M. Frahm, Comparative Evaluation of Binary Features, ECCV2012

Descriptor	Run time [ms.]	Speed-up [-]	Size [bytes]
SURF	117.1	3.83	64
SIFT	448.6	1.00	128
BRIEF	**3.8**	118.05	32
BRISK	**10.6**	42.32	64
ORB	**4.2**	106.80	32
LIOP	1 801.1	0.25	144
MROGH	2 976.8	0.15	192
MRRID	5 625.1	0.08	256

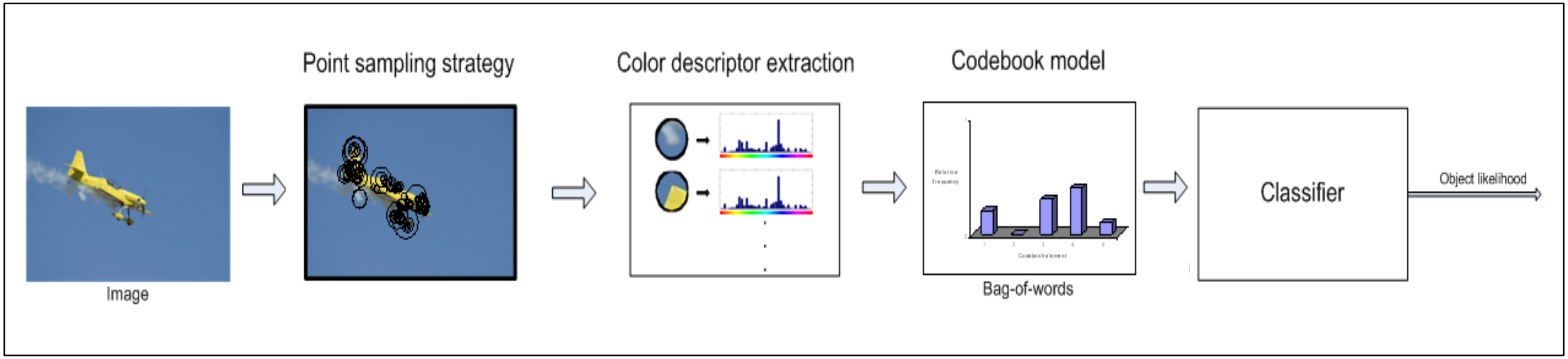
Detector	Descriptor	Precision	Recall	MAP
SURF	SURF	0.485	0.513	0.334
SURF	SIFT	0.525	0.533	0.491
SURF	BRIEF	0.517	0.546	0.514
SURF	ORB	0.448	0.470	0.437
SURF	LIOP	0.581	0.597	0.568
SURF	MROGH	0.540	0.567	0.527
SURF	MRRID	0.550	0.569	0.510
SURF	BRISK	0.536	0.553	0.530
BRISK	BRISK	0.504	0.527	0.492
ORB	ORB	0.493	0.495	0.463
FAST	SIFT	0.366	0.376	0.336

Previous Evaluations

- Image/object categories
  - K. Mikolajczyk, B. Leibe, and B. Schiele, “Local features for object class recognition,” in ICCV, 2005
  - E. Seemann, B. Leibe, K. Mikolajczyk, and B. Schiele, “An evaluation of local shape-based features for pedestrian detection,” in BMVC, 2005.
  - M. Stark and B. Schiele, “How good are local features for classes of geometric objects,” in ICCV, 2007.
  - K. E. A. van de Sande, T. Gevers and C. G. M. Snoek, Evaluation of Color Descriptors for Object and Scene Recognition. CVPR, 2008.

Approach

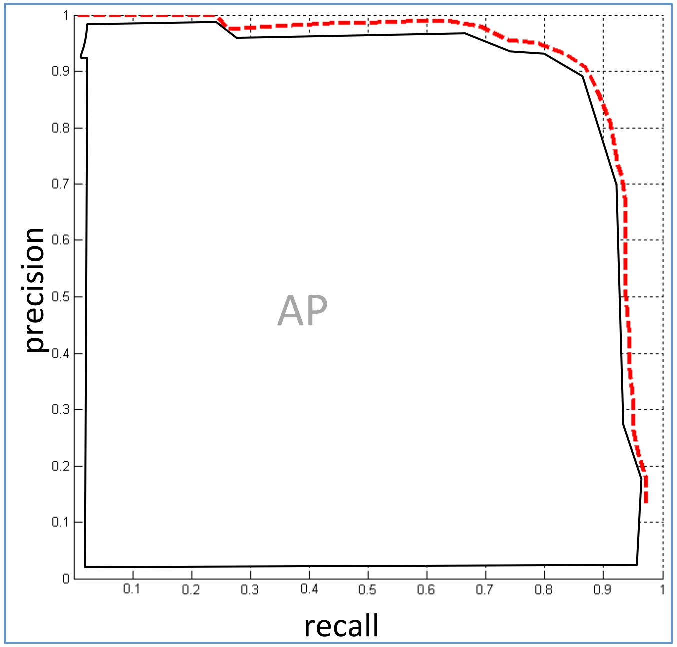
- Bags-of-features
  1. Interest point / region detector
  2. Descriptors
  3. K-means clustering (4000 clusters)
  4. Histogram of cluster occurrences (NN assignment)
  5. Chi-square distance and RBF kernel for KDA or SVM classifier



• J. Zhang and M. Marszalek and S. Lazebnik and C. Schmid,
Local Features and Kernels for Classification of Texture and Object Categories:
A Comprehensive Study, IJCV, 2007
• K. E. A. van de Sande, T. Gevers and C. G. M. Snoek,
Evaluation of Color Descriptors for Object and Scene Recognition. CVPR, 2008

Evaluation

- PASCAL VOC measures
  - Average precision for every object category
  - Mean average precision



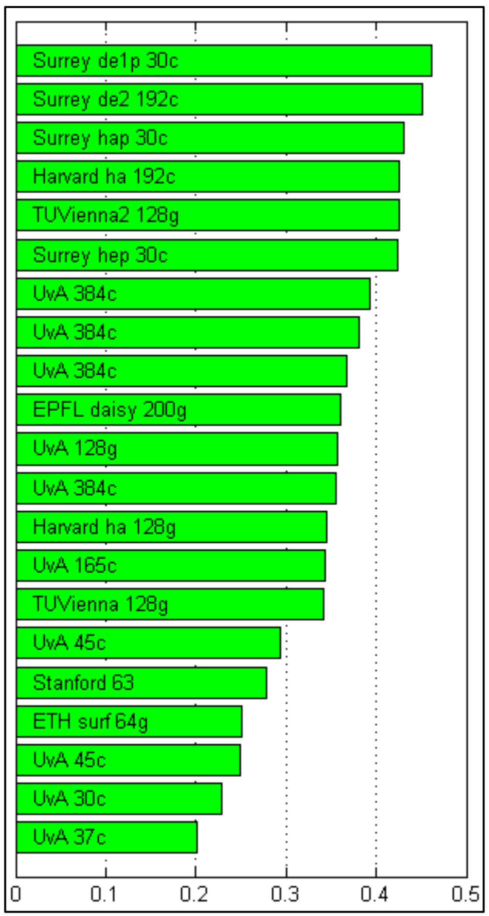
output=>

Category	AP
Aeroplane	0.7920
Bicycle	0.3984
Bird	0.4892
Boat	0.6264
Bottle	0.1884
Bus	0.5255
Car	0.5548
Cat	0.5641
Chair	0.4254
Cow	0.2324
Dining Table	0.2655
Dog	0.3273
Horse	0.4688
MotorBike	0.3994
Person	0.8819
Potted Plant	0.2240
Sheep	0.3077
Sofa	0.3886
Train	0.6832
TV /Monitor	0.5088
MAP	0.4623

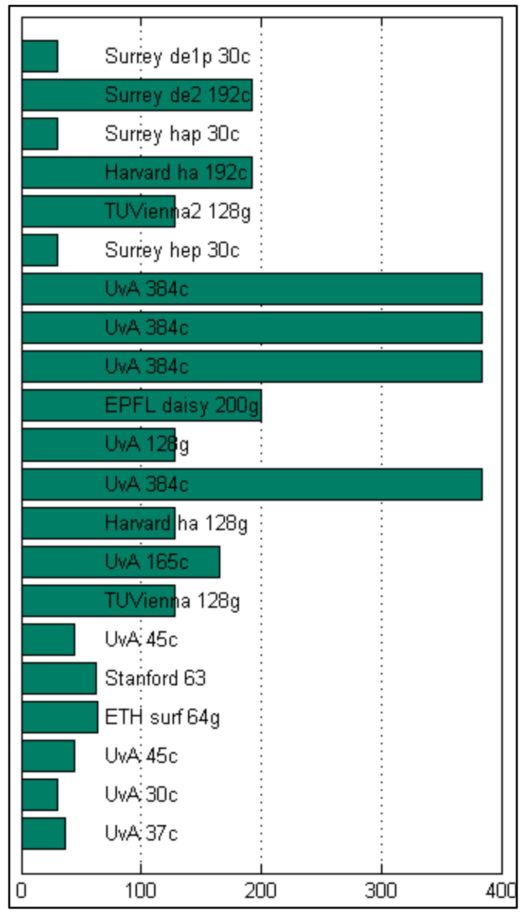
MAP Ranking

color/gray, density, dimensionality ...

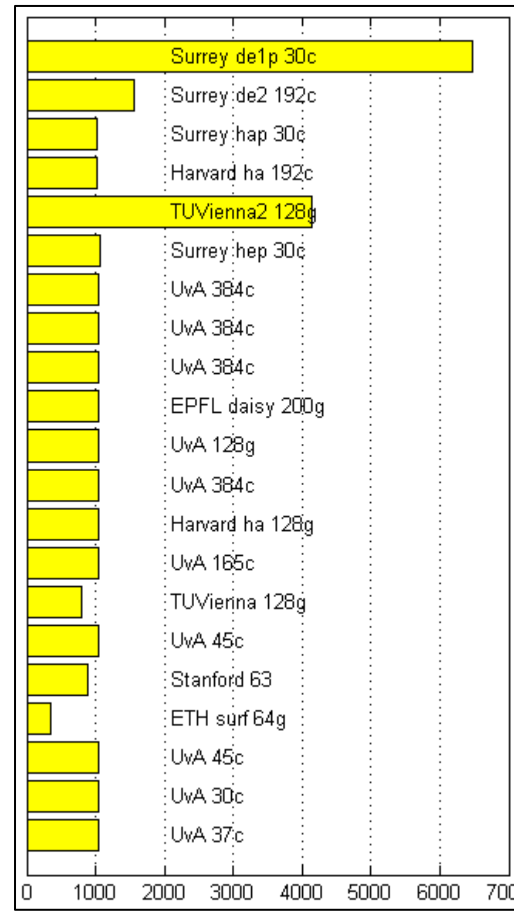
MAP



#dimensions

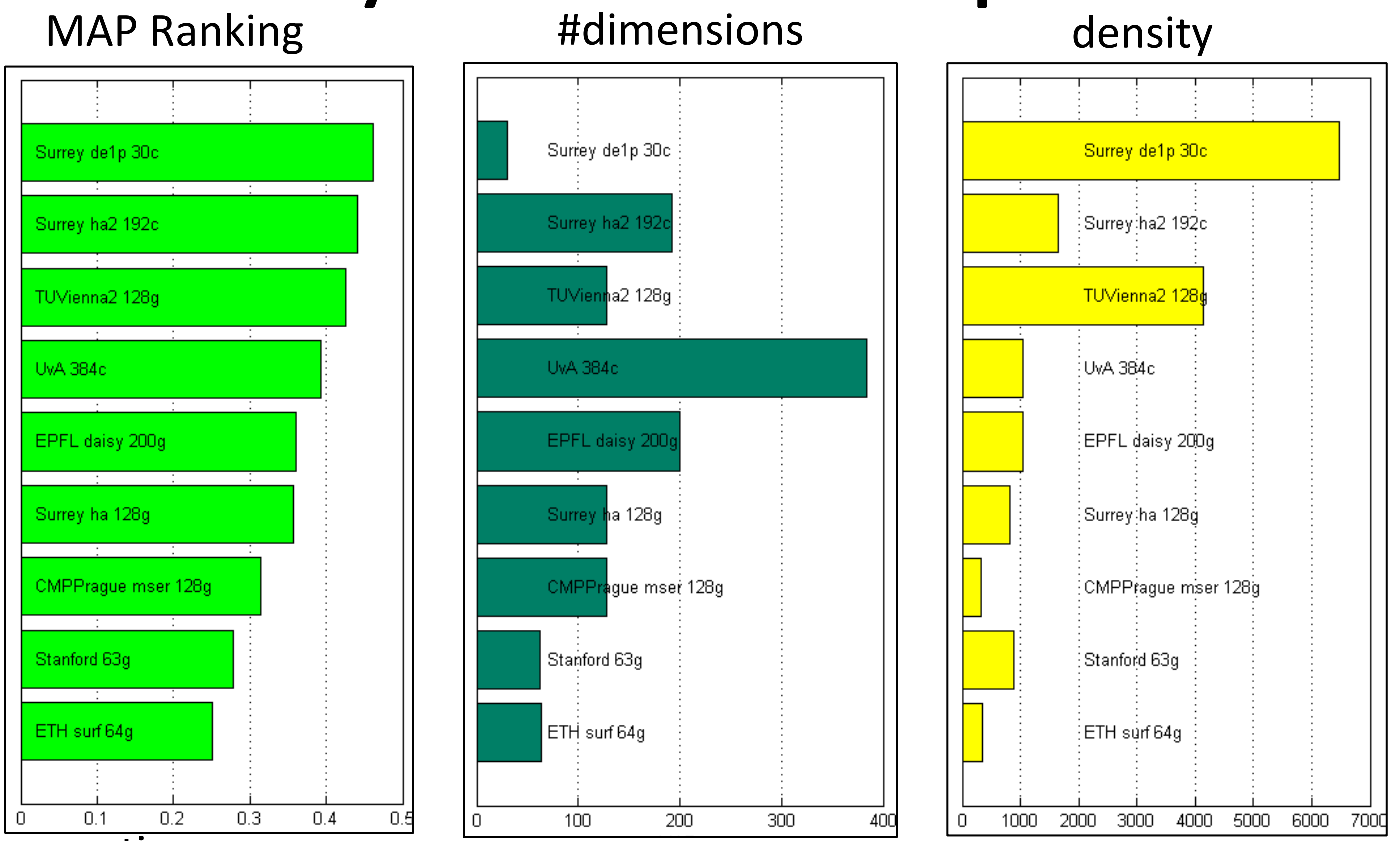


density



- SIFT still dominates (Histograms of gradient locations and orientations)
- Opponent chromatic space (normalized red-green, blue-yellow, and intensity Y

Grayvalue descriptors



- Observations
  - Color improves
  - All based on histograms of gradient locations and orientations
  - Dimensionality not much correlated with the performance
  - Density Strongly correlated (the more the better)
  - Results biased by density
  - <u>Implementation details matter</u>

Break !