

Feature Detectors and Descriptors: The State Of The Art and Beyond

In conjunction with CVPR 2009, Miami, Florida

[home](#)

[about](#)

[program](#)

[submission](#)

[evaluation](#)



In the last decade we have seen enormous progress in the area of invariant local image and video descriptors that led to substantial improvements in many computer vision areas including registration, stereo vision, motion estimation, matching, retrieval, recognition of objects and actions. Techniques for low level image or video description are of wide interest in the computer vision and pattern recognition communities as they are the first stage of many algorithms. The purpose of the workshop is to encourage researchers to present their latest developments in local feature extraction for the general use or for specific applications. There are still many challenges left for a variety of problem classes. In particular, there is a need for diverse measurements from image and video, robust to large changes of viewing conditions, extreme photometric variations, occlusion, background clutter as well as intra-class variations and specific applications i.e., category level correspondence.. We hope the workshop will also serve to establish a standard evaluation protocol and test data for local features.

Performance evaluation:

In conjunction with this workshop we will provide a possibility of an independent evaluation of features by the organizers. The test will include some aspects of recognition/categorization related to PASCAL 2008 challenge as well as matching of new image sets with variations in viewing and acquisition conditions.

Organizers

Krystian Mikolajczyk

CVSSP, UK



Cordelia Schmid

LEAR, France



Jiri Matas

CMP, Czech Republic



Tinne Tuytelaars

VISICS, Belgium



Feature Detectors and Descriptors

The state of the art and beyond

Local covariant detectors and descriptors have been successful in many applications

- Registration
- Stereo vision
- Motion estimation
- Matching
- Retrieval
- Classification
- Detection
- Action recognition
- Robot navigation

However, they have still many issues to address

Different classes of detector in term of generality

- Harris, DoG, EBR, MSER All possible task
- Sharable category-specific features
- Category specific
- Object specific
- Detector of a particular patch

Of course, in the restricted domain, performance goes up, but at a cost (learning required, multiple detectors run, may overfit).

Direct comparison in terms of e.g. repeatability not appropriate.

Detector & Descriptor evaluation issues

Mikolajczyk et al. IJCV 05 protocol, but:

- **Bias towards dense responses:**
*Larger number of detection → better performance
(return all windows and you'll be perfect)*
- **Bias towards large regions:**
No parallax (either flat scene or same viewpoint)
- **Bias towards (current) detector-friendly problems:**
*Scenes selected so that standard detectors perform well.
There are many image categories, where standard detectors are useless.
Increasing generality (applicability) cannot be demonstrated on this dataset*
- **Limited number of scenes, risk of over-fitting**

Detector & Descriptor evaluation issues

Mikolajczyk et al. IJCV 05 protocol (like all others we know)

ignore important aspects of detectors:

- Memory requirements
note papers dealing with the “how do I reduce memory needs” problem, very important for large scale problem
- Speed
this might be overrated, most detector are not the bottleneck of a system, most can be parallelized easily, should be well motivated
- “Coverage”
reasonable spatial distribution of features is beneficial ; metric not clear
- Complementarity
even a “underperforming” feature in terms of repeatability is valuable, if complementary
- Geometric precision
- Degrees of freedom fixed

Can the Mikolajczyk et al. protocol be fixed? *future work ...*

Detector & Descriptor as Components

- Performance in different application/niches may vary significantly; different aspects are important
- Evaluation in “context” of some (general) application such as categorization is important → see Evaluation Report later
- In a system, yet another set of aspects of a detector becomes important: e.g. efficient feedback control, adaptability

Previous Evaluations

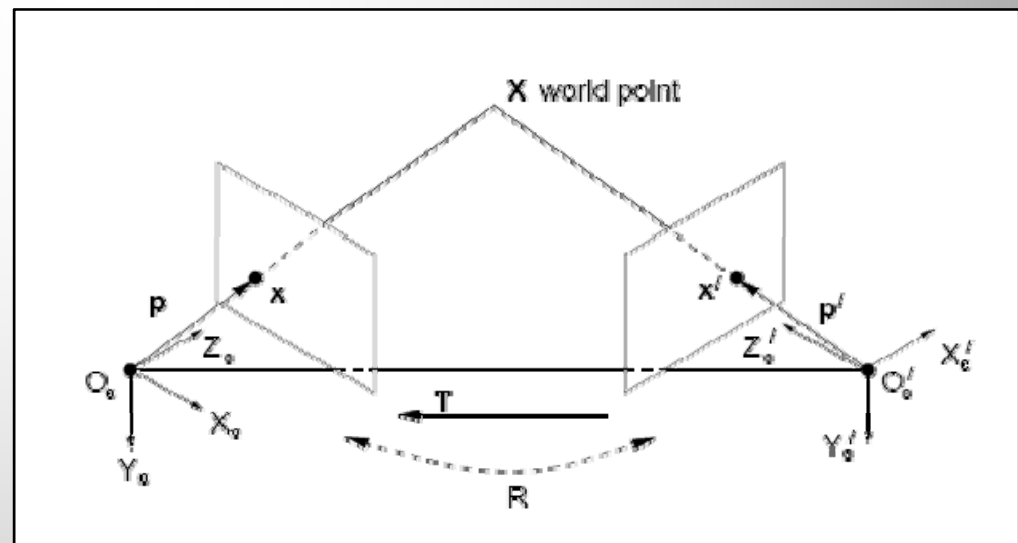
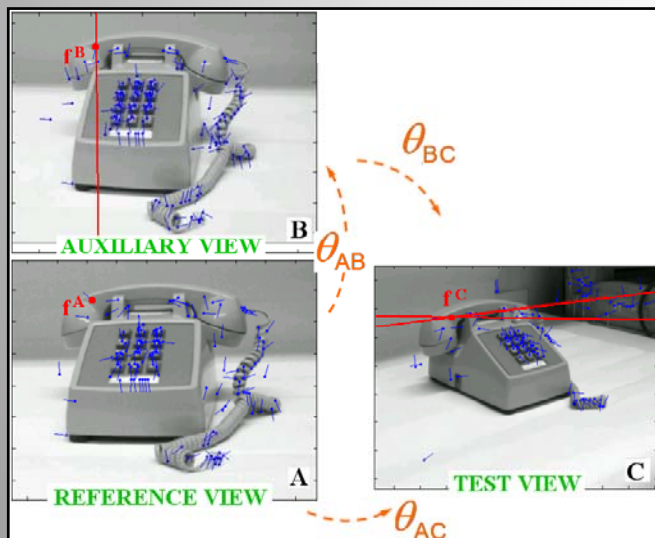
- 2D Scene – Homography

- 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



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.



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.



Recent Developments

- Domains
 - Registration
 - Stereo vision
 - Motion estimation
 - Matching
 - Retrieval
 - Classification
 - Detection
 - Action recognition
 - Robot navigation

Recent Developments

- Learning features from the data
 - Interest point detectors - decision trees, AdaBoost
 - Local descriptors – exhaustive optimizations of descriptor parameters
- Improving discriminability – dimensionality/memory reduction
 - LDA, PCA
- Feature selection
 - General saliency, application driven selection
- Quantization
 - Clustering
- Efficient features extraction algorithms
 - Integral images, sampling strategies, GPU implementations
- Domain specific features
 - 3D, spatio-temporal, IR, biology, medical imaging

Feature Detectors and Descriptors Evaluation for image classification

Krystian Mikolajczyk and Mark Barnard
University of Surrey, UK

CVPR, Miami, 2009

Setting up an evaluation

- Which problem?
 - Category recognition
 - Bags-of features
- What dataset?
 - Pascal VOC 2007
- Protocol and criteria?
 - Public dataset,
 - Avoiding risk to over-fitting/optimizing to the data

Dataset

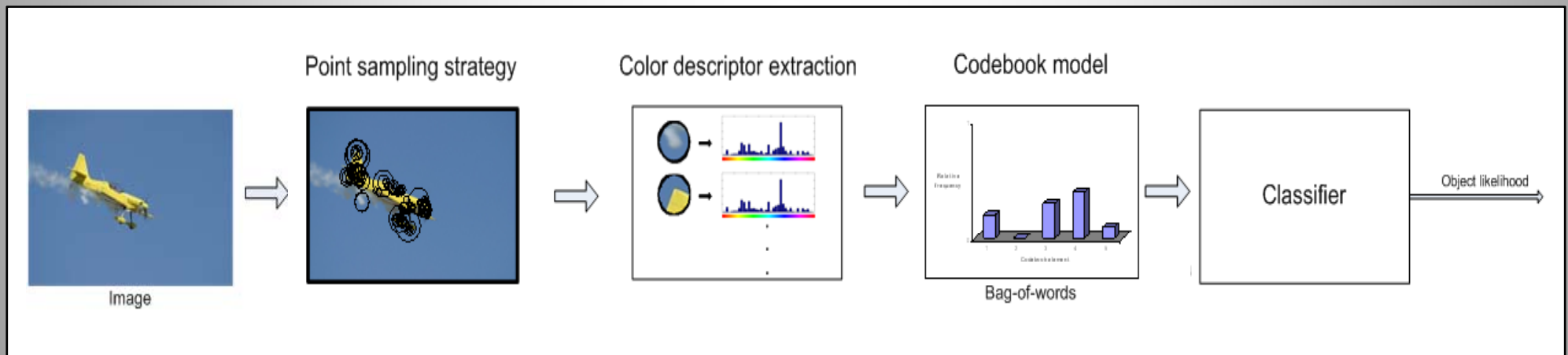
- PASCAL VOC 2007,
 - 20 object categories,
 - 5011 training/validation images
 - 4952 test images
 - 24,640 annotated objects



Approach

- Bags-of-features

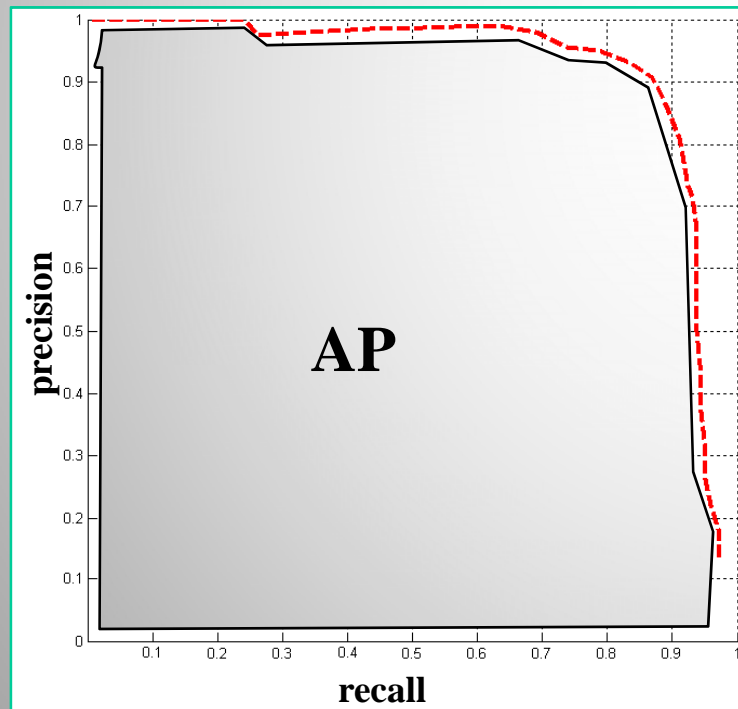
1. Interest point / region detector
2. Descriptors
3. K-means clustering (4000 clusters)
4. Histogram of cluster occurrences (NN assignment)
5. Chi-square distance and RBF kernel for KDA or SVM classifier



- 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


Recognition Evaluation

- www.featurespace.org
 - Protocol, data, binaries

Feature Detectors and Descriptors: The State Of The Art and Beyond

In conjunction with CVPR 2009, Miami, Florida




[home](#) [about](#) [program](#) [submission](#) [evaluation](#)



In the last decade we have seen enormous progress in the area of invariant local image and video descriptors that led to substantial improvements in many computer vision areas including registration, stereo vision, motion estimation, matching, retrieval, recognition of objects and actions. Techniques for low level image or video description are of wide interest in the computer vision and pattern recognition communities as they are the first stage of many algorithms. The purpose of the workshop is to encourage researchers to present their latest developments in local feature extraction for the general use or for specific applications. There are still many challenges left for a variety of problem classes. In particular, there is a need for diverse measurements from image and video, robust to large changes of viewing conditions, extreme photometric variations, occlusion, background clutter as well as intra-class variations and specific applications i.e., category level correspondence.. We hope the workshop will also serve to establish a standard evaluation protocol and test data for local features.

Performance evaluation:
In conjunction with this workshop we will provide a possibility of an independent evaluation of features by the organizers. The test will include some aspects of recognition/categorization related to PASCAL 2008 challenge as well as matching of new image sets with variations in viewing and acquisition conditions.

Organizers

Krystian Mikolajczyk CVSSP, UK 	Cordelia Schmid LEAR, France 	Jiri Matas CMP, Czech Republic 	Tinne Tuytelaars VISICS, Belgium 
---	--	---	---

Evaluated Features

- 33 different features types
 - University of Amsterdam
 - TU Vienna
 - CMP Prague
 - ETH Zurich
 - EPFL Lausanne
 - University of Surrey
 - Stanford University
 - Harvard Medical School

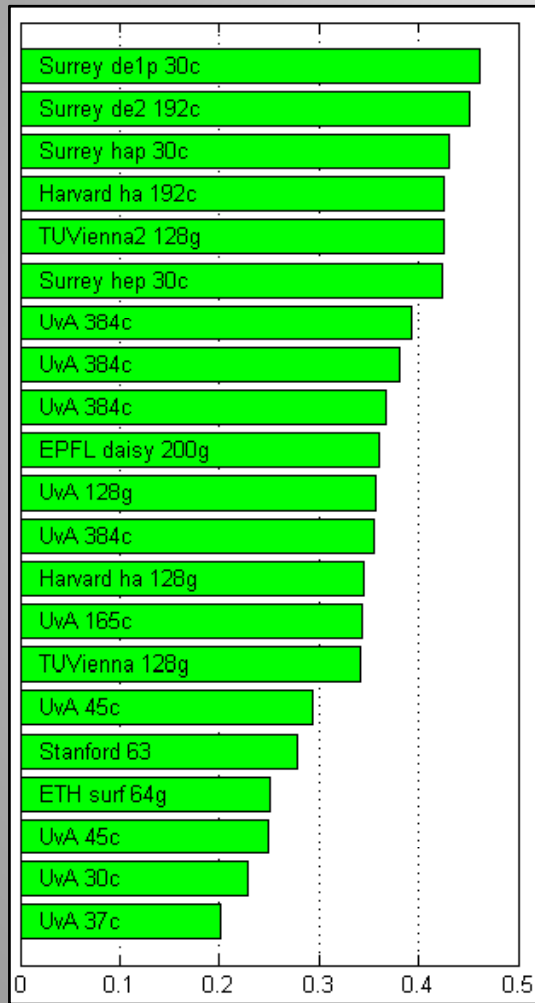
Features

- **Region detectors** (combined with SIFT descriptor)
 - MSER – CMP Prague
 - Computed in opponent chromatic space
 - Harris-Laplace, Hessian-Laplace, Surrey UK
 - Improved localization and scale precision
 - Dog (D. Lowe)
 - Sparse color features, TU Vienna
 - Color Harris detector
- **Region descriptors** (combined with Harris-Laplace detector)
 - Original Sift (D. Lowe)
 - Color descriptors (histograms, moments, SIFT) - UvA Amsterdam
 - DAISY – EPFL Lausanne
 - SURF – ETH Zurich
 - Color descriptors - Surrey UK
 - Ordinal SIFT – Harvard
 - CHoG– Stanford

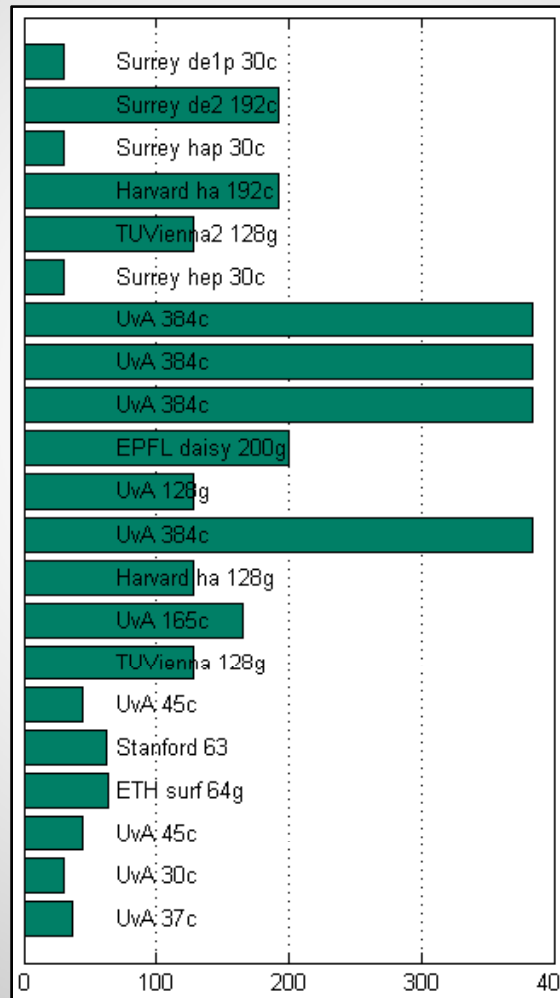
MAP Ranking

color/gray, density, dimensionality ...

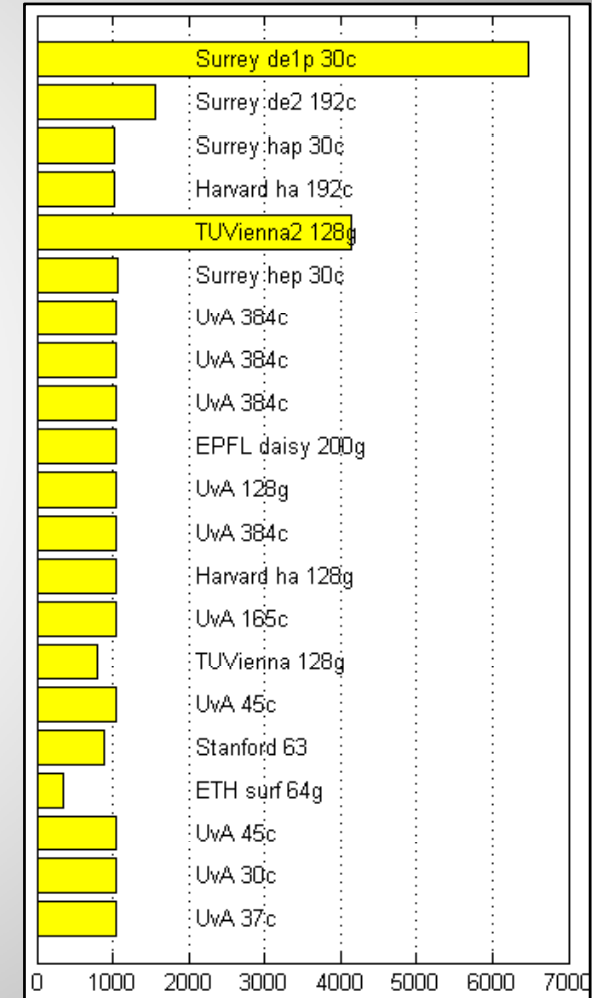
MAP



#dimensions

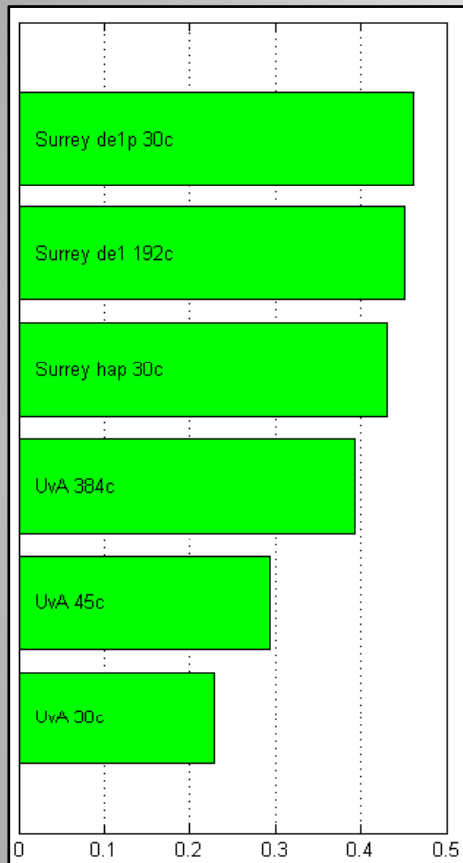


density

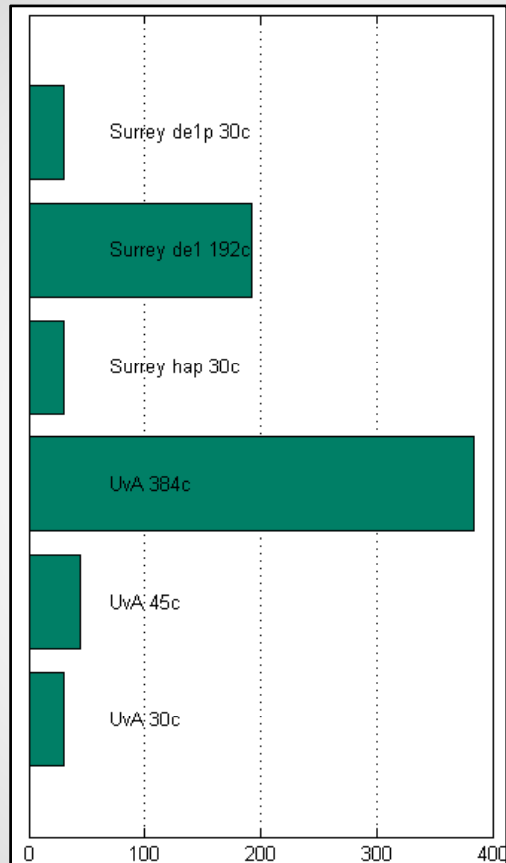


P07/DM Dimensionality

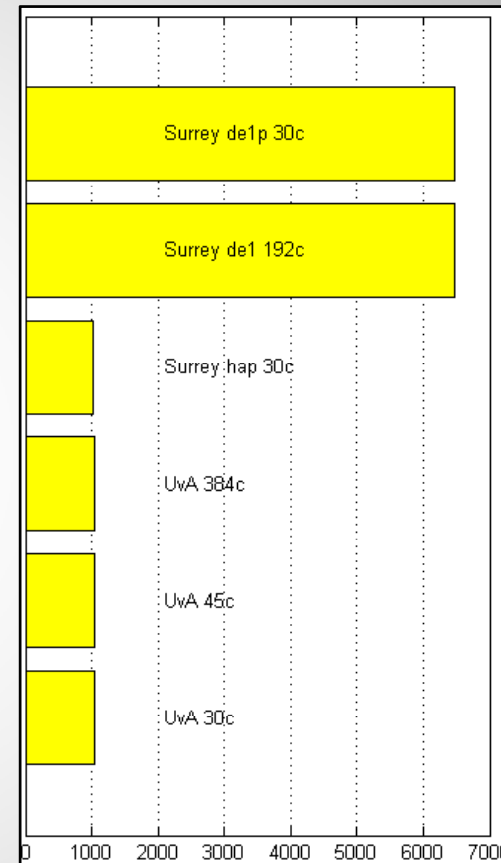
MAP



#dimensions



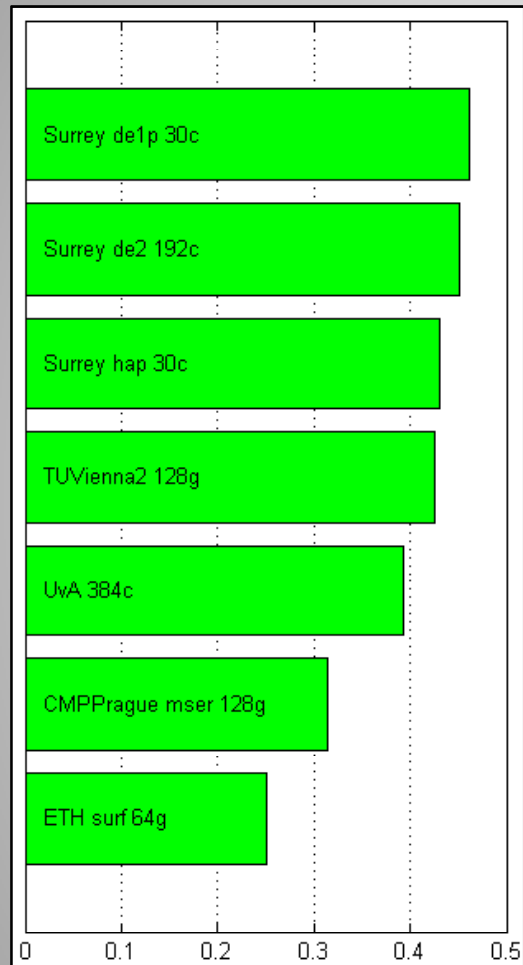
density



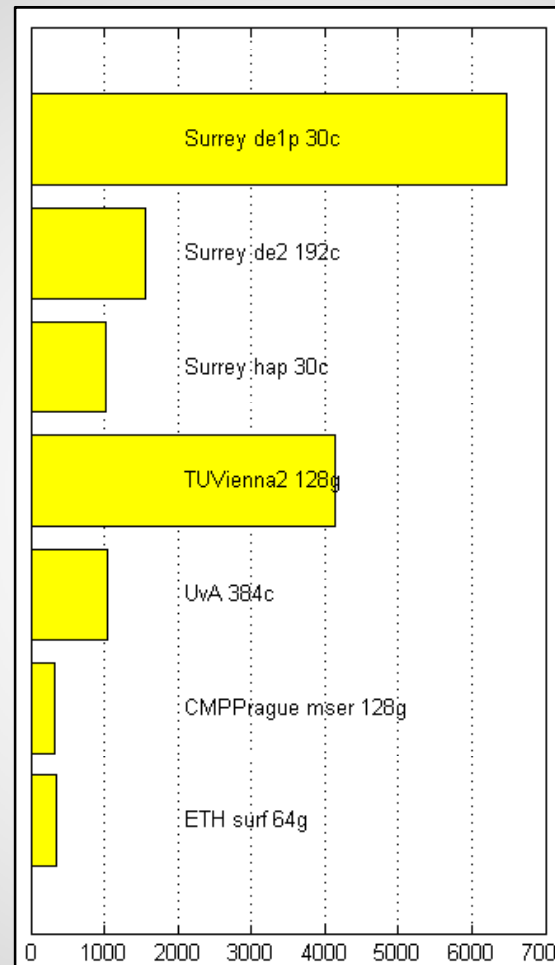
- Observations
 - Dimensionality not much correlated with the performance

P07/DS Density

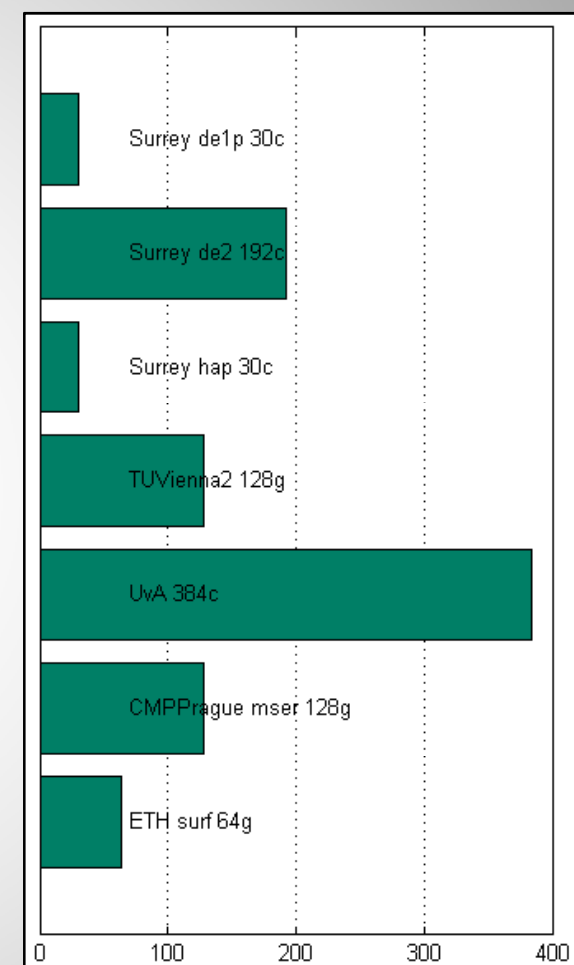
MAP Ranking



density



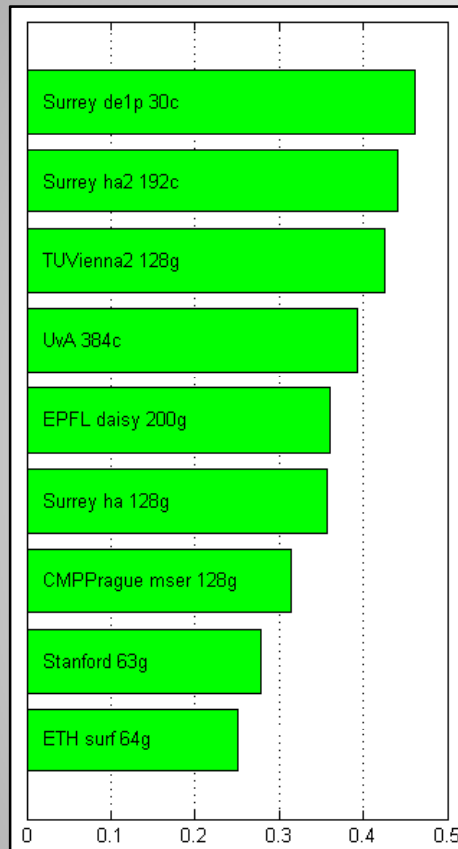
#dimensions



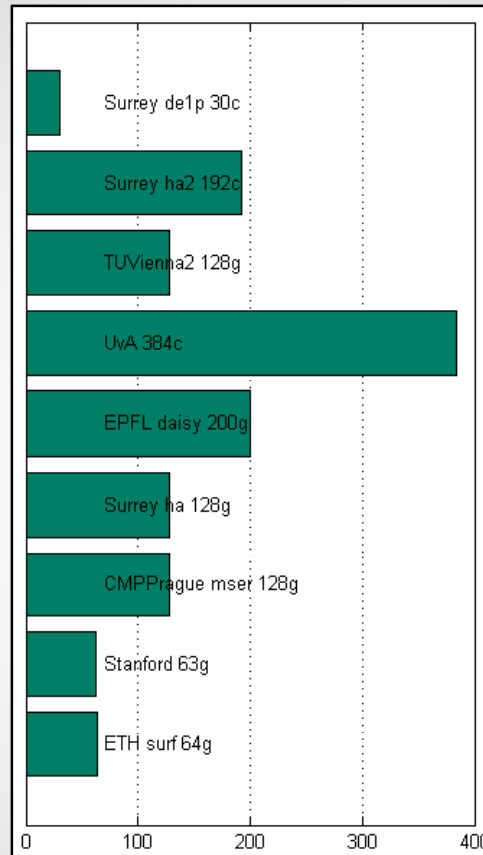
- Observations
 - Strongly correlated (the more the better)

P07/GD Grayvalue descriptors

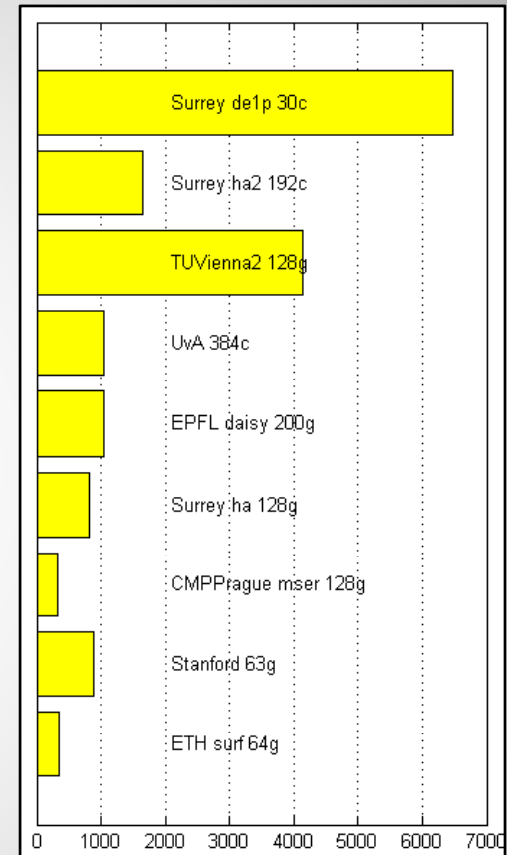
MAP Ranking



#dimensions



density

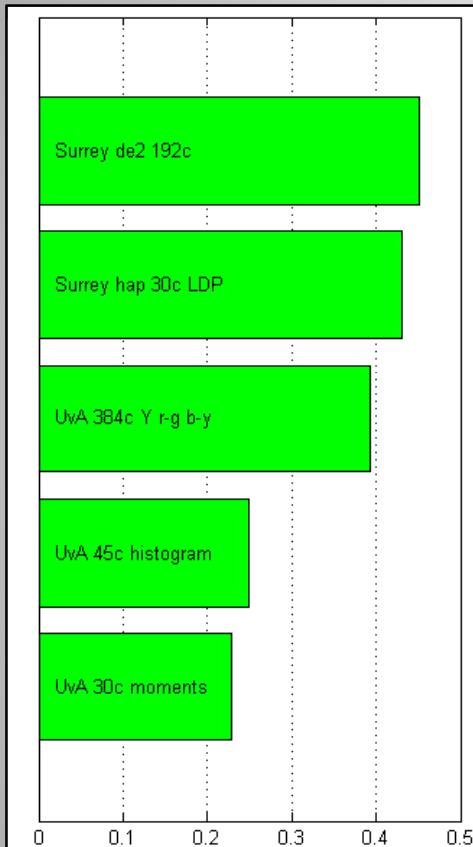


- Observations

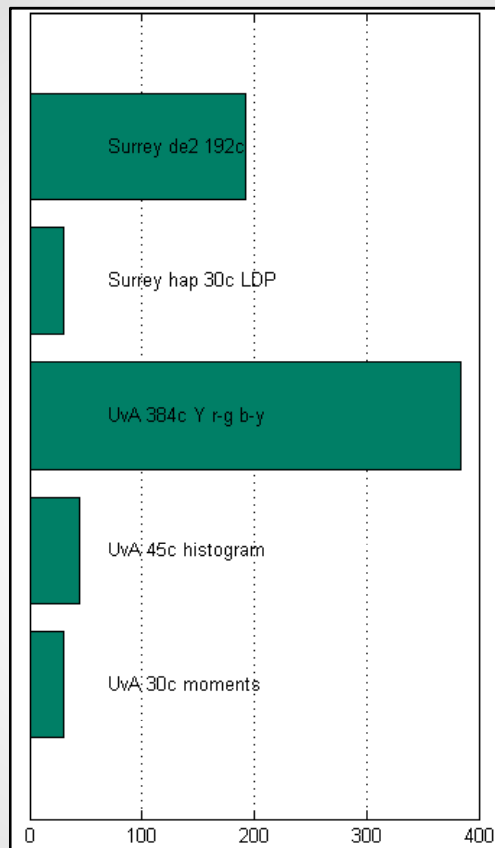
- Color improves
- All based on histograms of gradient locations and orientations
- Results biased by density
- Implementation details matter

P07/CD Color Descriptors

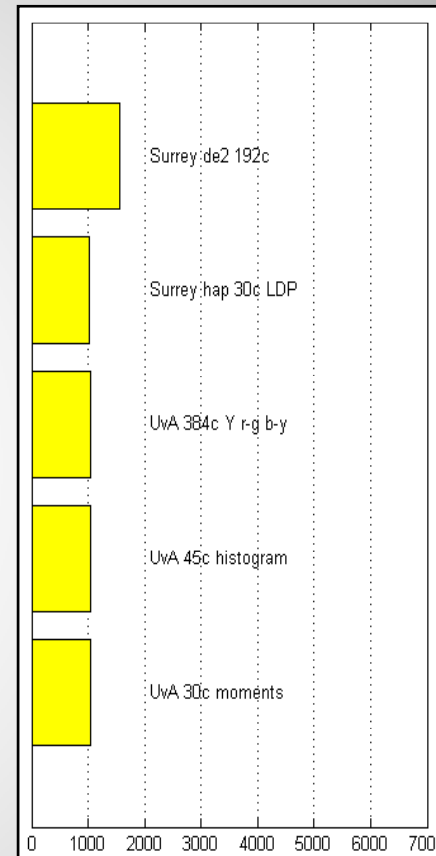
MAP Ranking



#dimensions



density

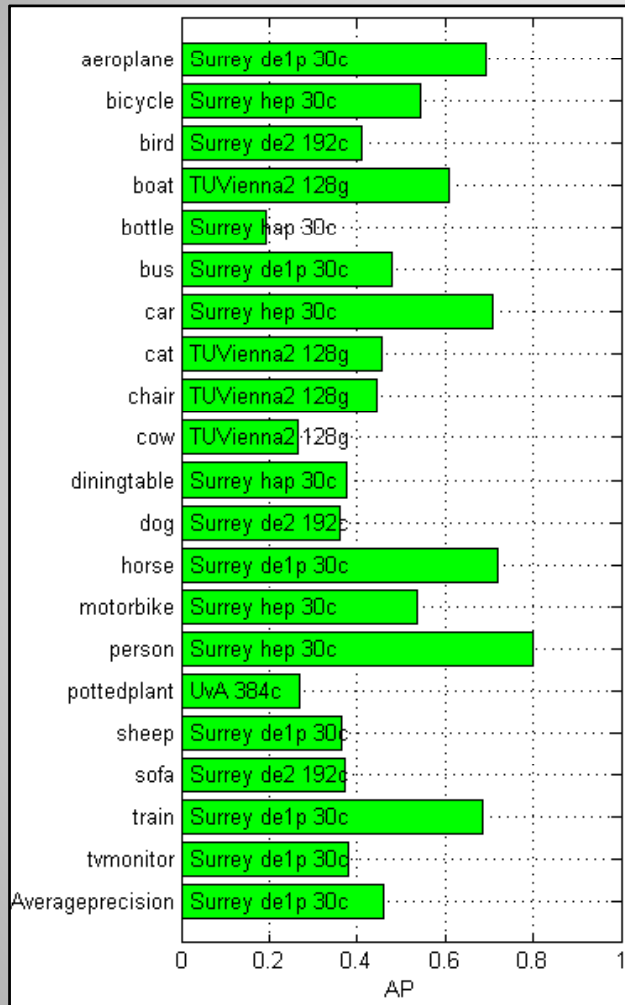


• Observations

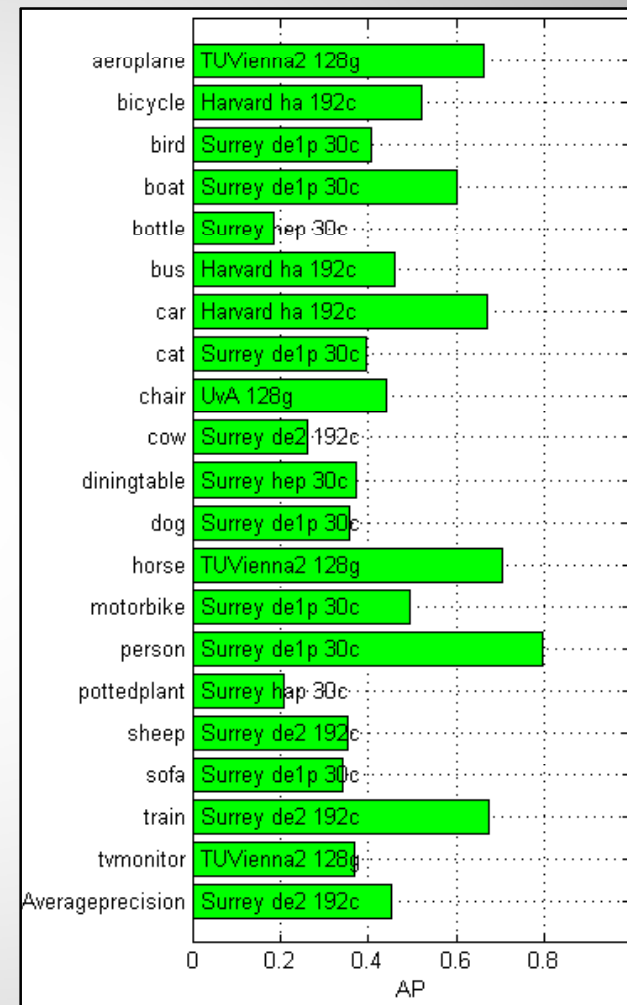
- SIFT still dominates (Histograms of gradient locations and orientations)
- Opponent chromatic space (normalized red-green, blue-yellow, and intensity Y)

P07/OC per object category

Ranked 1st



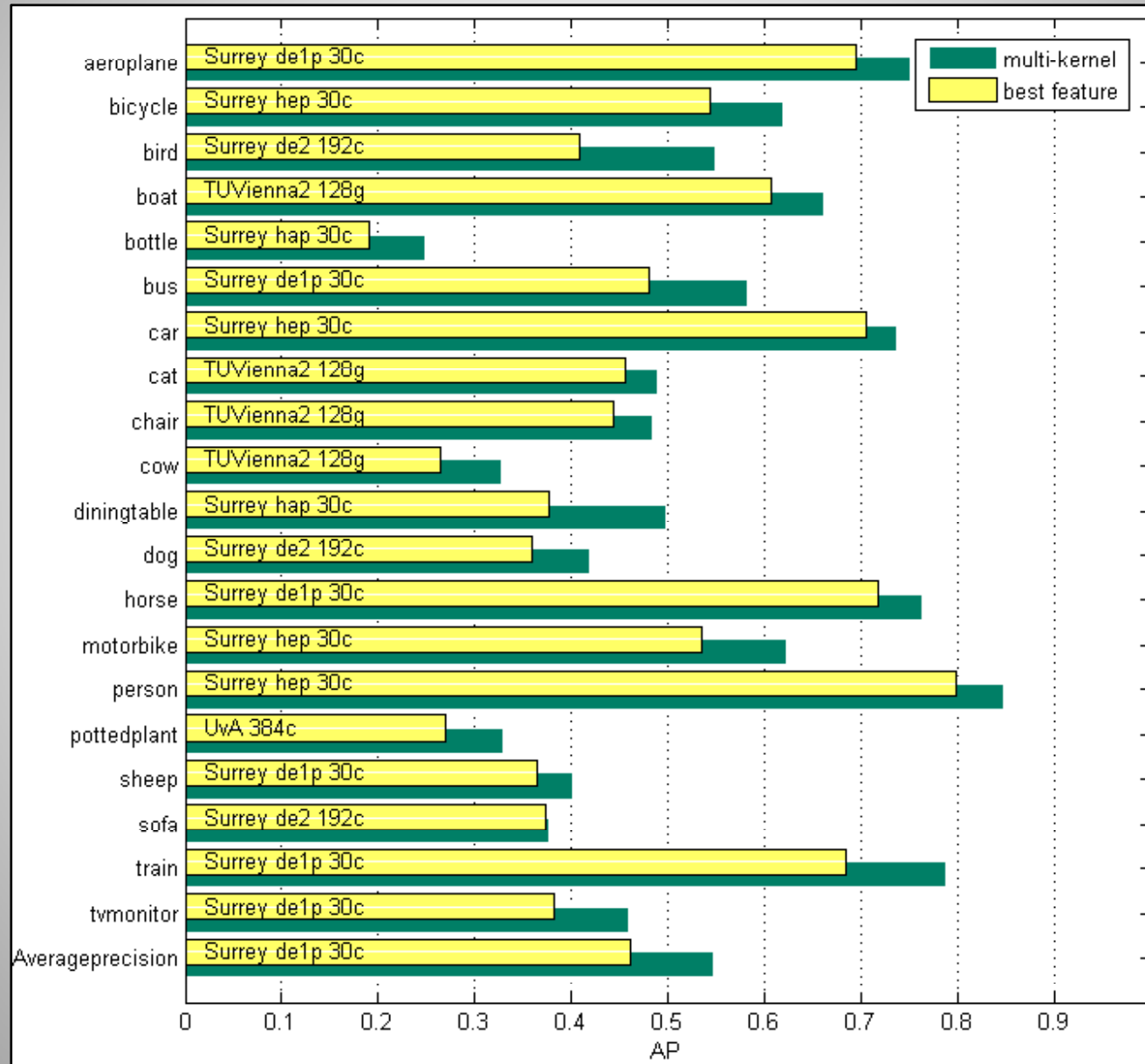
Ranked 2nd



Observations

- No single best feature, different features suitable for different categories
- Bias from the number of training images

P07/MK Multi-Kernel Fusion per object category



Conclusions

- Histograms of gradient location-orientation dominate
- Color brings improvement for most classes
 - opponent chromatic space
- Feature number
 - the more the better
- Similar ranking in image matching
 - performance generalizes across applications