UvA & Surrey @ PASCAL VOC 2008

Visual Features

Koen van de Sande
Jasper Uijlings
Xirong Li
Theo Gevers
Arnold Smeulders
University of Amsterdam

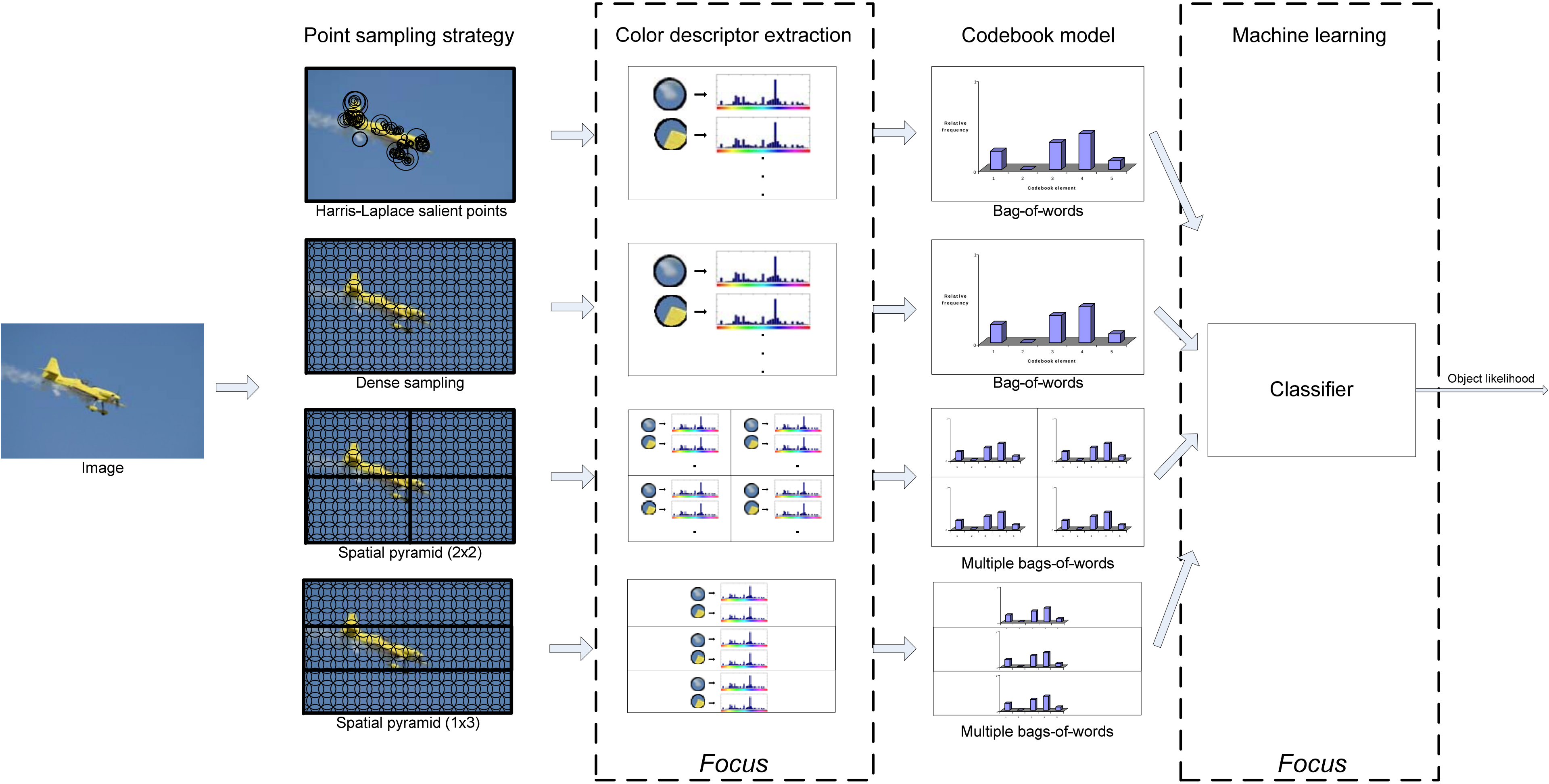
Machine Learning

Muhammad Atif Tahir
Fei Yan
Krystian Mikolajczyk
Josef Kittler

University of Surrey

VIDIVIDEO

Pipeline Overview



Related work

Real-world scenes:

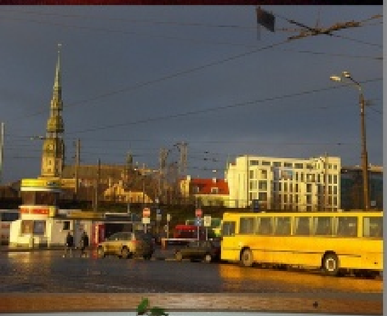
Large variations in viewing and lighting conditions
➔ image description complicated

Viewing conditions:

- Orientation/scale of object changes
- Salient point methods can robustly detect regions which are [LoweIJCV2004], [ZhangIJCV2007] :
  - Translation-invariant
  - Rotation-invariant
  - Scale-invariant
- Dense sampling at multiple scales ‘brute force’ solution

Illumination changes:

➔ **How do changes in lighting conditions affect object detection?**



Color descriptors

Illumination changes:

- Object detection impaired if region description is not robust
- SIFT is most well-known descriptor, state-of-the-art performance [MikolajczykPAMI2005,ZhangIJCV2007]
- Evaluations compare intensity-based descriptors only

Color descriptors have been proposed to:

- Increase illumination invariance
- Increase discriminative power

In "*Evaluation of Color Descriptors for Object and Scene Recognition*" [VanDeSandeCVPR2008]:

- Invariance properties of color descriptors shown analytically using a taxonomy of invariant properties within the diagonal model of illumination change
- Distinctiveness of color descriptors shown on VOC2007

Diagonal model

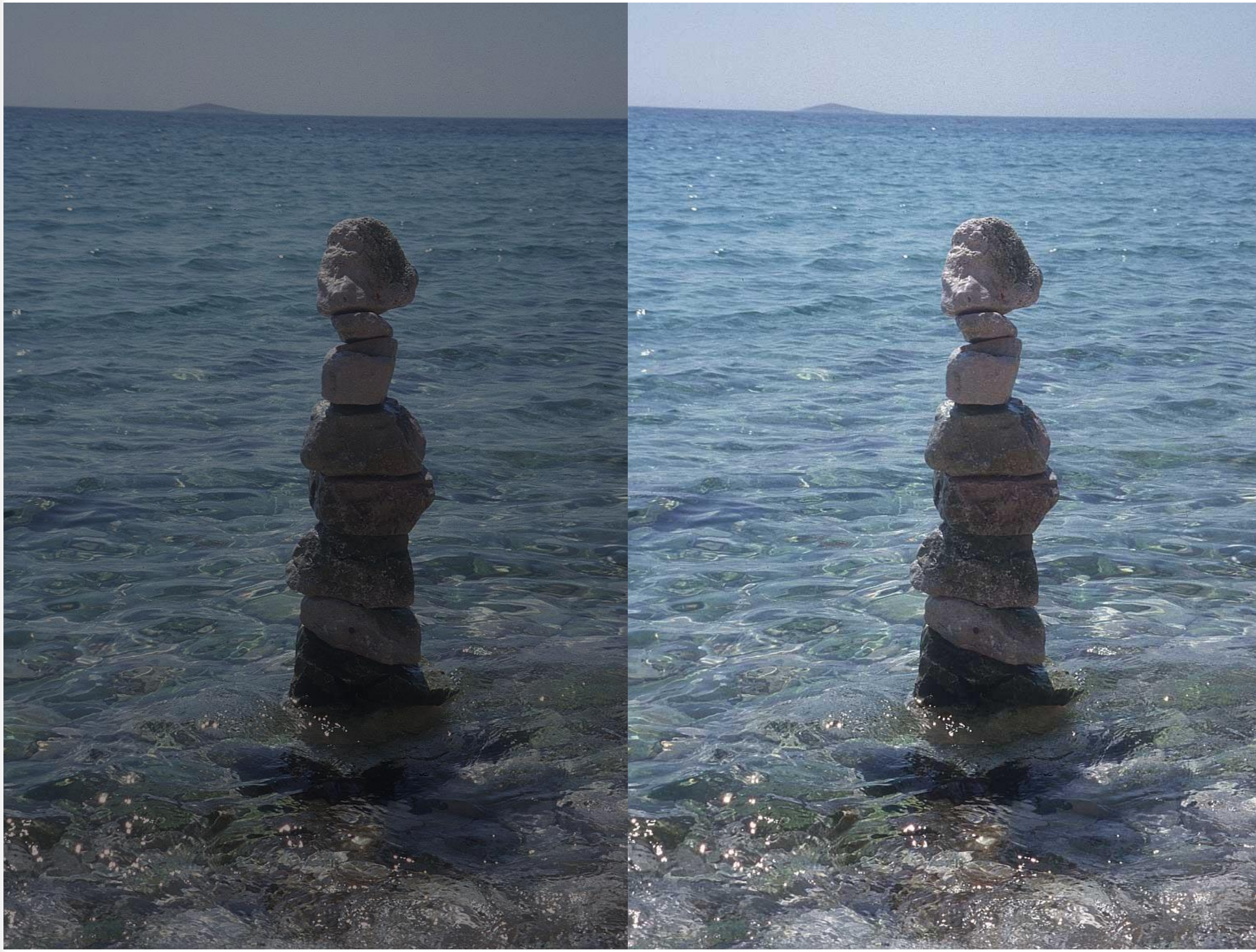
- Diagonal-offset model of illumination change

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} + \begin{pmatrix} o_1 \\ o_2 \\ o_3 \end{pmatrix}$$

- Can model shadows, shading, light color changes, highlights

u = unknown illuminant
c = canonical illuminant

Example: Light intensity change



Photometric Analysis

Light intensity change (a = b = c)

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix}$$

Examples: shadows, shading

$$I^c = a\, I^u$$

Color Descriptor Taxonomy

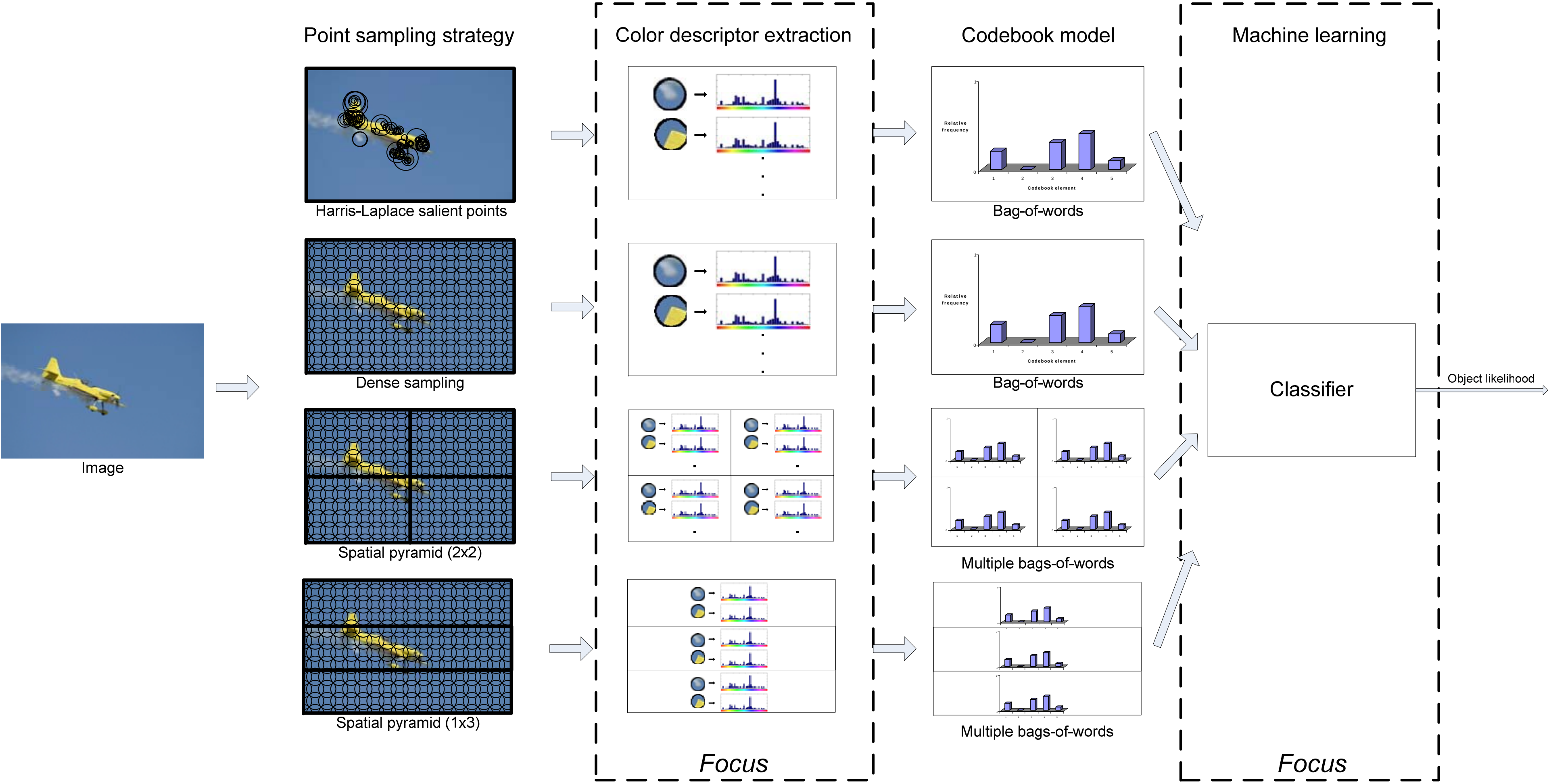
- Invariance properties of the descriptors used
- See [VanDeSandeCVPR2008] for additional color descriptors

	Light intensity change $\begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$	Light intensity shift $\begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} o_1 \\ o_1 \\ o_1 \end{pmatrix}$	Light intensity change and shift $\begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} o_1 \\ o_1 \\ o_1 \end{pmatrix}$	Light color change $\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$	Light color change and shift $\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} o_1 \\ o_2 \\ o_3 \end{pmatrix}$
SIFT	+	+	+	+	+
OpponentSIFT	+/-	+	+/-	+/-	+/-
WSIFT	+	+	+	+/-	+/-
*rg*SIFT	+	+	+	+/-	+/-
Transformed color SIFT	+	+	+	+	+

Descriptors	**MAP on VOC2008val**
Intensity SIFT	42,3
All five (=Soft5ColorSIFT)	45,5

By adding color: +8%

Pipeline Overview



Feature Components

Point sampling strategy:

- Harris-Laplace detector
- Dense sampling every 6 pixels at multiple scales

Spatial pyramid:

- 1x1 (whole image)
- 2x2 (image quarters) [LazebnikCVPR2006]
- 1x3 (horizontal bars) [MarszalekVOC2007]

Descriptors:

- Intensity-based SIFT [LoweIJCV2004]
- OpponentSIFT
- WSIFT
- *rg*SIFT
- Transformed color SIFT

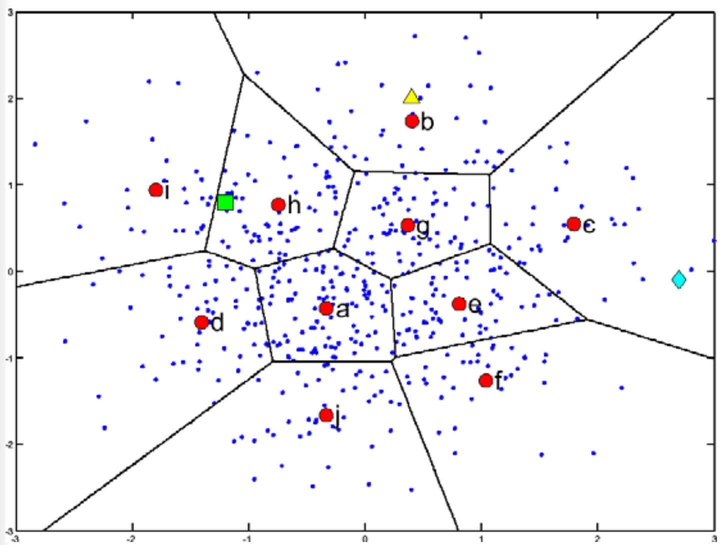
Cf. [VanDeSandeCVPR2008] for evaluation of color descriptors

30 possible combinations of <sampling, pyramid, descriptor>

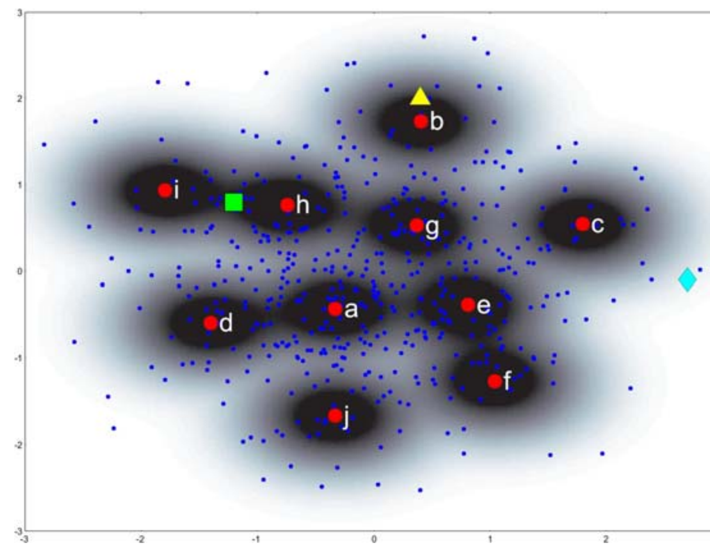
Feature Components (2)

Bag-of-words model:

- Use kernel codebooks [VanGemertECCV2008]
- Soft assignment to codebook elements using Gaussian kernel
- Codebook size = 4000, created using k-means



Codebook



Kernel codebook

Assignment	MAP on VOC2008val
Codebook	43,4
Kernel codebook (=Soft5ColorSIFT)	45,5

+5%

Classification

Classifier baseline

Soft5ColorSIFT run

(SIFT, OpponentSIFT, WSIFT, rgSIFT, Transformed color SIFT)

- Combine 30 feature components using equal weight
- Single γ^2 SVM classifier
- Same as the flat fusion done in [MarszalekVOC2007]

	MAP on VOC2008val
Soft5ColorSIFT	45,5

Linear Discriminant Analysis

- LDA is a traditional statistical method that is proved successful in classification problems
  - The objective is to maximize the between-class covariance
  - and simultaneously minimize the within-class covariance
- The classical LDA is a linear method and fails for non linear problems

Kernel Discriminant Analysis

- Many nonlinear extensions of LDA have been proposed e.g.
  - Kernel Fisher Discriminant Analysis [Mika et al 1999, NNSP]
  - Generalized Discriminant Analysis [Baudat and Anouar 2000, Neural Computation]
  - KDA using QR decomposition [Xiong et al. 2004, Advances in NIPS]
  - KDA using Spectral Regression [Deng et al. 2007 ICDM]

KDA (cont.)

- The idea of non linear extensions is to solve LDA in a kernel feature space
- Need to handle the singularity problem
  - Widely used approaches are Singular Value Decomposition and Regularization techniques
  - That normally requires eigen value decomposition
  - Computationally expensive for very large data sets

KDA using Spectral Regression

- Recently KDA using SR is introduced for spoken letter and face recognition by Deng Cai (ICDM 2007)
- Avoids eigen-decomposition of the kernel-matrix
- The main idea is to use Cholesky Decomposition to solve linear equations

$$(K + \delta \mathrm{I})\alpha = \mathrm{y}$$

KDA using Spectral Regression

- The equation $(K+\delta \mathrm{I})\alpha = \mathrm{y}$ has close connection with regularized regression [Vapnik, Statistical learning theory, 1998]
- Projection functions are optimal for separating training samples with different labels
- To avoid overfit, regularization is necessary

KDA using Spectral Regression

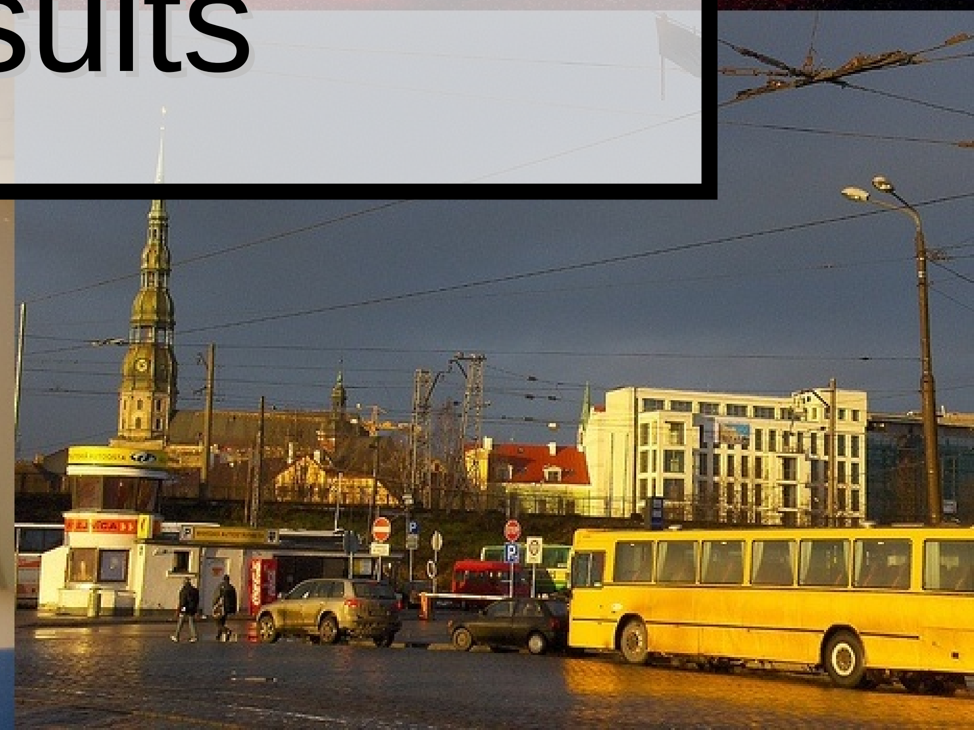
- Theoretical analysis has shown that SRKDA has achieved 27-times speedup over conventional KDA

- Also competitive with Support Vector Machine in terms of classification accuracy

	MAP on VOC2008val
Soft5ColorSIFT	45,5
SRKDA	46,3

Results



Object Category	SurreyUvA_SRKDA	UvA_Soft5ColorSift	UvA_TreeSFS
Aeroplane	79,5	79,7	80,8
Bicycle	54,3	52,1	53,2
Bird	61,4	61,5	61,6
Boat	64,8	65,5	65,6
Bottle	30,0	29,1	29,4
Bus	52,1	46,5	49,9
Car	59,5	58,3	58,5
Cat	59,4	57,4	59,4
Chair	48,9	48,2	48,0
Cow	33,6	27,9	30,1
Dining table	37,8	38,3	39,6
Dog	46,0	46,6	45,0
Horse	66,1	66,0	67,3
Motorbike	64,0	60,6	60,4
Person	86,8	87,0	87,1
Potted plant	29,2	31,8	30,1
Sheep	42,3	42,2	41,5
Sofa	44,0	45,3	45,4
Train	77,8	72,3	74,3
TV/Monitor	61,2	64,7	59,8
MAP	**54,9**	**54,1**	**54,4**

(also uses
randomized forests)

VOC2007 vs. VOC2008 data

- Runs Soft5ColorSIFT and 20072008Soft5ColorSIFT
- 30 components combined using equal weight
- Single γ^2 SVM classifier

Train set	MAP on VOC2007test
2007 train+val	60,5*
2008 train+val	55,8
2007+2008 train+val	63,8

* 2007 Challenge best = 59,4 [MarszalekVOC2007]

Train set	MAP on VOC2008test
2007 train+val	?
2008 train+val	54,1
2007+2008 train+val	58,6

TRECVID2008 benchmark

Using same visual features [MediamillTRECVID2008]:

- Highest overall MAP in TRECVID2008 HLF (“concept detection”) task
- Highest AP for 9 out of 20 concepts, not all with same parameter settings
- Many factors can influence final results, see [TRECVID]

Conclusions

- Adding color information in descriptors on top of intensity information improves ~8%
- In Pascal VOC Challenge, SRKDA gives better mean average precision (MAP) than Support Vector Machines
- Adding kernels based on diverse features increases the MAP




Questions?

Visit http://www.science.uva.nl/~ksande
for color descriptor executables (in a few weeks)

References

- [VanDeSandeCVPR2008] K. E. A. van de Sande, T. Gevers and C. G. M. Snoek, "*Evaluation of Color Descriptors for Object and Scene Recognition*", CVPR 2008
- [VanGemertECCV2008] J.C. van Gemert, J.M. Geusebroek, C.J. Veenman, A.W.M. Smeulders, "*Kernel Codebooks for Scene Categorization*", ECCV 2008
- [CaiICDM2007] "*Efficient Kernel Discriminant Analysis via Spectral Regression*", International Conference on Data Mining 2007
- [MarszalekVOC2007] M. Marszalek, C. Schmid, H. Harzallah and J. van de Weijer, "*Learning Object Representations for Visual Object Class Recognition*", Visual Recognition Workshop in conjunction with ICCV 2007
- [TRECVID] A. F. Smeaton, P. Over and W. Kraaij, "*Evaluation campaigns and TRECVid*", MIR 2006
- [MikolajczykPAMI2005] K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors", PAMI 2005
- [LoweIJCV2004] D. G. Lowe, "*Distinctive Image Features from Scale-Invariant Keypoints*", IJCV 2004
- [ZhangIJCV2007] J. Zhang, M. Marszalek, S. Lazebnik and C. Schmid, "*Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study*", IJCV 2007
- [MediamillTRECVID2008] C. G. M. Snoek et al, "*The MediaMill TRECVID 2008 Semantic Video Search Engine*", TRECVID Workshop 2008