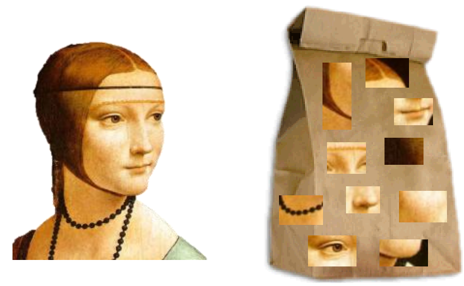
EEE422 (EEE6082) Computational Vision

Bag of Features

Ling Shao

Bag-of-features models



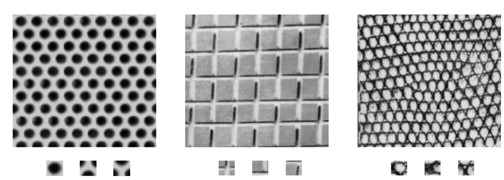
Many slides adapted from Fei-Fei Li, Rob Fergus, and Antonio Torralba

Overview: Bag-of-features models

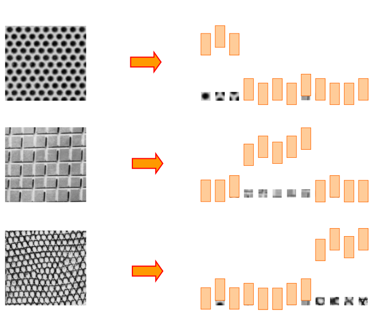
- Origins and motivation
- Image representation
  - Feature extraction
  - Visual vocabularies
- Discriminative methods
  - Nearest-neighbor classification
  - Distance functions
  - Support vector machines
  - Kernels
- Generative methods
  - Naïve Bayes
  - Probabilistic Latent Semantic Analysis
- Extensions: incorporating spatial information

Origin 1: Texture recognition

- Texture is characterized by the repetition of basic elements or *textons*
- For stochastic textures, it is the identity of the textons, not their spatial arrangement, that matters



Origin 1: Texture recognition

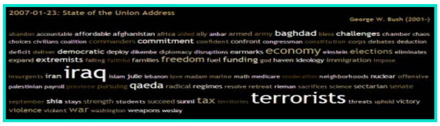


Origin 2: Bag-of-words models

- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

Origin 2: Bag-of-words models

- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)



Origin 2: Bag-of-words models

- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

Origin 2: Bag-of-words models

- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

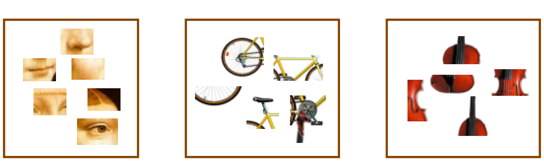
Bags of features for object recognition

face, flowers, building

- Works pretty well for image-level classification

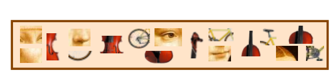
Bag of features: outline

1. Extract features



Bag of features: outline

1. Extract features
2. Learn "visual vocabulary"

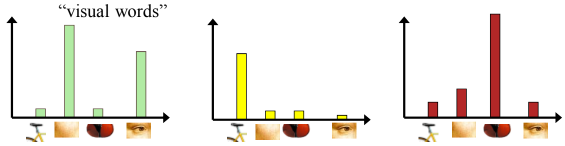


Bag of features: outline

1. Extract features
2. Learn “visual vocabulary”
3. Quantize features using visual vocabulary

Bag of features: outline

1. Extract features
2. Learn “visual vocabulary”
3. Quantize features using visual vocabulary
4. Represent images by frequencies of “visual words”

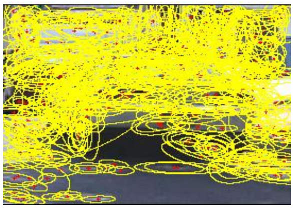


1. Feature extraction

- Regular grid
  - Vogel & Schiele, 2003
  - Fei-Fei & Perona, 2005

1. Feature extraction

- Regular grid
  - Vogel & Schiele, 2003
  - Fei-Fei & Perona, 2005
- Interest point detector
  - Csurka et al. 2004
  - Fei-Fei & Perona, 2005
  - Sivic et al. 2005



1. Feature extraction

- Regular grid
  - Vogel & Schiele, 2003
  - Fei-Fei & Perona, 2005
- Interest point detector
  - Csurka et al. 2004
  - Fei-Fei & Perona, 2005
  - Sivic et al. 2005
- Other methods
  - Random sampling (Vidal-Naquet & Ullman, 2002)
  - Segmentation-based patches (Barnard et al. 2003)

1. Feature extraction

Compute SIFT descriptor

[Lowe’99]

Normalize patch

Detect patches

[Mikojaczyk and Schmid ’02]

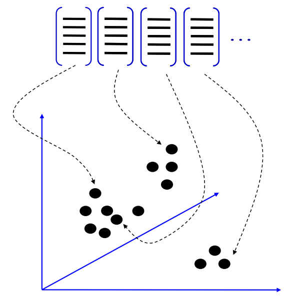
[Mata, Chum, Urban & Pajdla, ’02]

[Sivic & Zisserman, ’03]

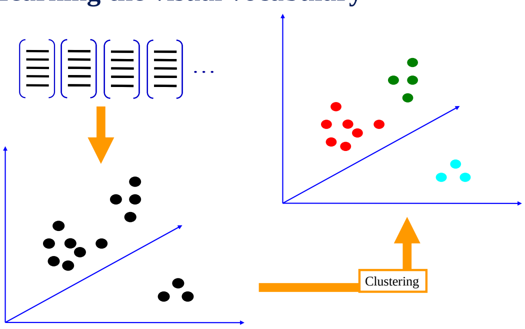
Slide credit: Josef Sivic

1. Feature extraction

2. Learning the visual vocabulary

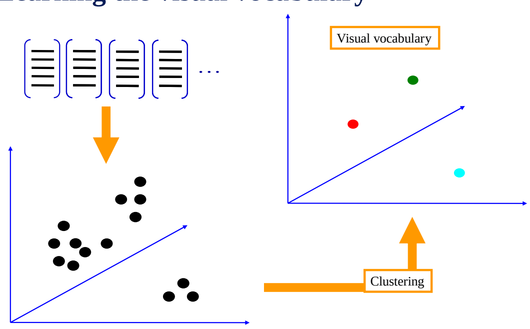


2. Learning the visual vocabulary



Slide credit: Josef Sivic

2. Learning the visual vocabulary



Slide credit: Josef Sivic

K-means clustering

- Want to minimize sum of squared Euclidean distances between points x_i and their nearest cluster centers m_k

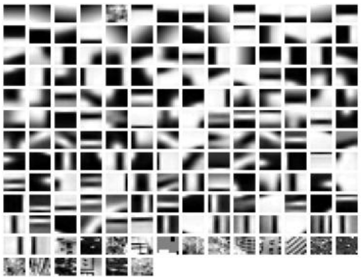
$$D(X,M) = \sum_{\text{cluster } k} \sum_{\substack{\text{point } i \text{ in} \\ \text{cluster } k}} (x_i - m_k)^2$$

- Algorithm:
- Randomly initialize K cluster centers
- Iterate until convergence:
  - Assign each data point to the nearest center
  - Recompute each cluster center as the mean of all points assigned to it

From clustering to vector quantization

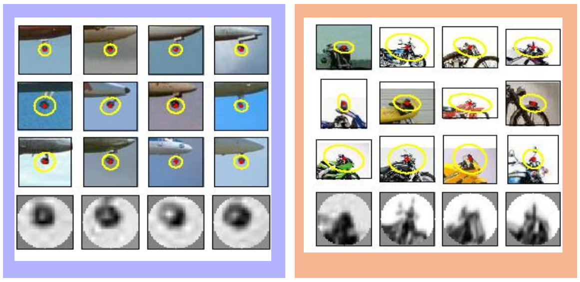
- Clustering is a common method for learning a visual vocabulary or codebook
  - Unsupervised learning process
  - Each cluster center produced by k-means becomes a codevector
  - Codebook can be learned on separate training set
  - Provided the training set is sufficiently representative, the codebook will be "universal"
- The codebook is used for quantizing features
  - A *vector quantizer* takes a feature vector and maps it to the index of the nearest codevector in a codebook
  - Codebook = visual vocabulary
  - Codevector = visual word

Example visual vocabulary



Fei-Fei et al. 2005

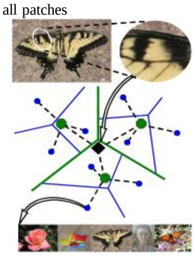
Image patch examples of visual words



Sivic et al. 2005

Visual vocabularies: Issues

- How to choose vocabulary size?
  - Too small: visual words not representative of all patches
  - Too large: quantization artifacts, overfitting
- Generative or discriminative learning?
- Computational efficiency
  - Vocabulary trees (Nister & Stewenius, 2006)



3. Image representation

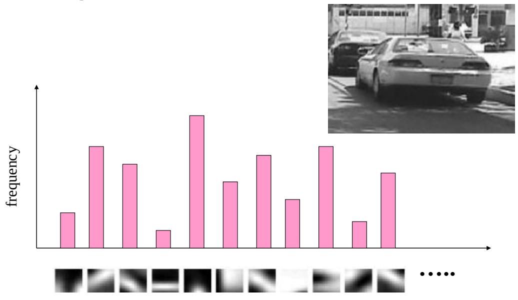
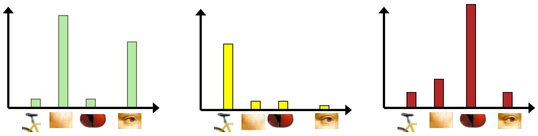
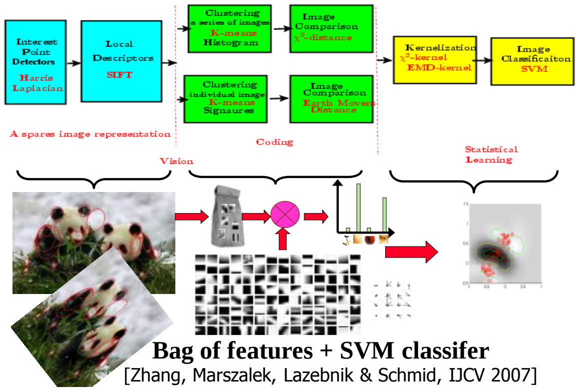


Image classification

- Given the bag-of-features representations of images from different classes, how do we learn a model for distinguishing them?





Bag of features + SVM classifer
[Zhang, Marszalek, Lazebnik & Schmid, IJCV 2007]

Evaluation: Invariance

Datasets	n	Scale Invariance			Scale and Rotation			Affine Invariance		
		HS	LS	HS+LS	HSR	LSR	HSR+LSR	HA	LA	HA+LA
UIUCTex	20	89.7±1.6	91.2±1.5	92.2±1.4	97.1±0.6	97.7±0.6	98.0±0.5	97.5±0.6	97.5±0.7	98.0±0.6
Xerox7	10 fold	92.0±2.0	93.9±1.5	94.7±1.2	88.1±2.1	92.4±1.7	92.2±2.3	88.2±2.2	91.3±2.1	91.4±1.8

- Best invariance level depends on datasets
- Scale invariance is often sufficient for object categories
- Affine invariance is rarely an advantage

Comparison with state-of-the-art

- Existing Datasets

Methods	Xerox7	CalTech6	Graz	CalTech101
Our method	94.3	97.9	90.0	53.9
Others	82.0 Csurka.et al. (ICPR 2004)	96.6 Csurka.et al (ICPR 2004)	83.7 Opelt et al eccv 04	43 Grauman and Darrel iccv 05

- PASCAL VOC Challenges
  - Task: Predict the existence of an object given an image
  - 2005: Four Categories
  - 2006: Ten Categories; More Participants

Pattern Analysis, Statistical Modeling and Computational Learning
→ Europe-wide Network of Excellence with 57 partners

PASCAL VOC 2005

- Testset 2: "Harder" Google images

Submission	motorbikes	bicycles	people	cars
	EER	EER	EER	EER
Aachen1	0.767	0.667	0.663	0.703
Aachen2	0.769	0.665	0.669	0.716
Darmstadt1	0.663	–	–	0.551
Darmstadt2	0.683	–	–	0.658
Edinburgh	0.698	0.575	0.519	0.633
HUT1	0.614	0.527	0.601	0.655
HUT2	0.624	0.604	0.614	0.676
HUT3	0.594	0.524	0.574	0.644
HUT4	0.635	0.616	0.587	0.692
INRIA-Zhang	0.798	0.728	0.719	0.720
MPITuebingen	0.698	0.616	0.591	0.677

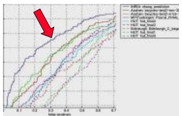


Figure 11: Competition 2.2: test2: bicycles (all entries)

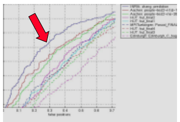


Figure 13: Competition 2.3: test2: people (all entries)

PASCAL VOC 2006

Cat

Horse

Cow

Sheep

Bike

Dog

Bus

Person

Motorbike

Car

- More than 20 participants from different institutions
- More challenges: Highly deformed pose, shape, viewpoint changes

AUC by Method and Class

	bicycle	bus	car	cat	cow	dog	horse	motor bike	person	sheep
AP06_Batra	0.791	0.637	0.833	0.733	0.756	0.644	0.607	0.672	0.550	0.792
AP06_Lee	0.845	0.916	0.897	0.859	0.838	0.766	0.694	0.829	0.622	0.875
Cambridge	0.873	0.864	0.887	0.822	0.850	0.768	0.754	0.844	0.715	0.866
INRIA_Larlus	0.903	0.948	0.943	0.870	0.880	0.743	0.850	0.890	0.736	0.892
INRIA_Marszalek	0.929	0.984	0.971	0.922	0.938	0.856	0.908	0.964	0.845	0.944
INRIA_Moosmann	0.903	0.933	0.957	0.883	0.895	0.825	0.824	-	0.780	0.930
INRIA_Nowak	0.924	0.973	0.971	0.906	0.892	0.797	0.904	0.961	0.814	0.940
INSARouen	-	-	0.895	-	-	0.764	-	-	-	0.869
MUL_1vALL	0.857	0.852	0.914	0.562	0.632	0.584	0.525	0.831	0.616	0.758
MUL_1v1	0.864	0.945	0.928	0.826	0.789	0.764	0.733	0.906	0.718	0.872
QMUL_HSLS	0.944	0.984	0.977	0.936	0.936	0.874	0.922	0.966	0.845	0.946
QMUL_LSPCH	0.948	0.981	0.975	0.937	0.938	0.876	0.926	0.969	0.855	0.956
RWTH_DiscHist	0.874	0.955	0.930	0.879	0.910	0.799	0.854	0.938	0.764	0.906
RWTH_GMM	0.882	0.935	0.942	0.866	0.856	0.825	0.802	0.905	0.718	0.892
RWTH_SparseHists	0.863	0.941	0.935	0.883	0.883	0.704	0.844	0.858	0.776	0.907
Siena	0.671	0.749	0.842	0.696	0.774	0.677	0.644	0.701	0.660	0.768
TKK	0.857	0.928	0.943	0.871	0.892	0.811	0.806	0.908	0.781	0.900
UVA_big5	0.897	0.929	0.945	0.845	0.862	0.785	0.806	0.923	0.774	0.885
UVA_weibull	0.855	0.880	0.910	0.818	0.849	0.762	0.759	0.888	0.723	0.811
XRCE	0.943	0.978	0.967	0.933	0.940	0.866	0.925	0.957	0.863	0.951