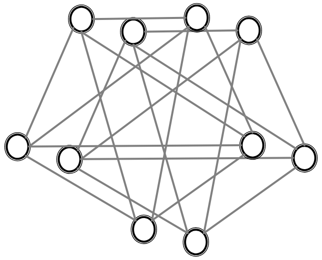
Sparse coding

d



Recognition Surveillance Search and Ranking Bioinformatics

The curse of dimensionality:
…increasingly demand inference with limited samples for very high-dimensional data.

The blessing of dimensionality:
… real data highly concentrate on low-dimensional, sparse, or degenerate structures in the high-dimensional space.

But nothing is free: *Gross errors and irrelevant measurements are now ubiquitous in massive cheap data.*

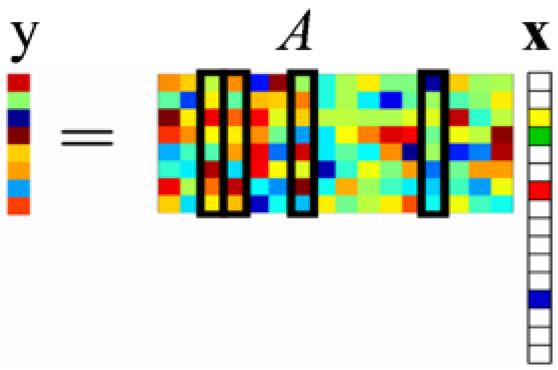
Outline

- ***Sparse representation***
  - ***Sparse coding***
  - ***Optimization for sparse coding***
  - ***Dictionary learning***
  - ***Recent advances in Computer Vision***

- ***Sparse methods on matrices***
  - ***Matrix completion (pLSA, collaborative filtering)***
  - ***Matrix factorization (topic model, LDA, low-rank)***
  - ***Some applications***

欠定线性方程组: $Ax = y$ 有无穷多组解

对解x加约束: 比如 $||x||_0$(x中非零元素的个数)尽可能小



即求解如下优化问题:

$$\min \quad ||x||_0 \, ,$$
$$\text{subject to} \quad Ax = y.$$

2003~2004年，Donoho & Elad做了一个很漂亮的证明，如果矩阵A 满足：$\sigma(A) \geq 2\|x\|_0$（$\sigma(A)$ 为最小的线性相关的列向量集所含的向量个数），上文提及的0范数优化问题具有唯一的解。

2006年，Tao和Donoho的弟子Candes合作证明了在RIP条件下，0范数优化问题与以下1范数优化问题具有相同的解(凸的)：

$$\begin{aligned} \min \quad & \|x\|_1, \\ \text{subject to} \quad & Ax = y. \end{aligned}$$

RIP条件是对于矩阵列向量正交性的一种衡量。

$$(1-\mu_N)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1\mu_N)\|x\|_2^2 \quad \forall x \quad \|x\|_0 \leq N$$

Sparse coding

进一步可以考虑含噪声情况:可以得到相似的结果

$$\min \|x\|_0 \quad s.t. \quad \|Ax-y\|_2^2 \le \varepsilon$$

问题表达为:

$$\underbrace{\begin{pmatrix} \mathbf{y} \end{pmatrix}}_{\mathbf{y}\in\mathbb{R}^m} \approx \underbrace{\left(\begin{array}{c|c|c|c} \mathbf{d}_1 & \mathbf{d}_2 & \cdots & \mathbf{d}_p \end{array}\right)}_{\mathbf{D}\in\mathbb{R}^{m\times p}} \underbrace{\begin{pmatrix} \alpha[1] \\ \alpha[2] \\ \vdots \\ \alpha[p] \end{pmatrix}}_{\alpha\in\mathbb{R}^p,\ \text{sparse}}$$

Two interpretations**:**

- ***Compressed sensing***: *A* as sensing matrix
- ***Sparse representation***: *A* as overcomplete dictionary

The Sparse Decomposition Problem

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^p} \underbrace{\frac{1}{2}||\mathbf{x}-\mathbf{D}\boldsymbol{\alpha}||_2^2}_{\text{data fitting term}} + \underbrace{\lambda\psi(\boldsymbol{\alpha})}_{\text{sparsity-inducing regularization}}$$

ψ induces sparsity in $\boldsymbol{\alpha}$. It can be

- the ℓ_0 "pseudo-norm". $||\boldsymbol{\alpha}||_0 \triangleq \#\{i \text{ s.t. } \boldsymbol{\alpha}[i] \neq 0\}$ (NP-hard)
- the ℓ_1 norm. $||\boldsymbol{\alpha}||_1 \triangleq \sum_{i=1}^p |\boldsymbol{\alpha}[i]|$ (convex)
- ...

This is a selection problem.

求解信号稀疏分解的优化问题：

1. Greedy Algorithms
2. Homotopy and LARS
3. Soft-thresholding based optimization

Matching Pursuit

每次寻找与残差相关性最大的字典原子，以此原子更新编码和残差。

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^p} ||\underbrace{\mathbf{x}-\mathbf{D}\boldsymbol{\alpha}}_{\mathbf{r}}||_2^2 \text{ s.t. } ||\boldsymbol{\alpha}||_0 \leq L$$

1: $\boldsymbol{\alpha} \leftarrow 0$

2: $\mathbf{r} \leftarrow \mathbf{x}$ (residual).

3: **while** $||\boldsymbol{\alpha}||_0 < L$ **do**

4: Select the atom with maximum correlation with the residual

$$\hat{\imath} \leftarrow \arg\max_{i=1,\ldots,p} |\mathbf{d}_i^T \mathbf{r}|$$

5: Update the residual and the coefficients

$$\begin{aligned} \boldsymbol{\alpha}[\hat{\imath}] &\leftarrow \boldsymbol{\alpha}[\hat{\imath}] + \mathbf{d}_{\hat{\imath}}^T \mathbf{r} \\ \mathbf{r} &\leftarrow \mathbf{r} - (\mathbf{d}_{\hat{\imath}}^T \mathbf{r})\mathbf{d}_{\hat{\imath}} \end{aligned}$$

6: **end while**

Orthogonal Matching Pursuit

每次寻找能够最大地减小目标函数的字典原子加入到编码基元中，并用正交投影更新残差。

$$\min_{\boldsymbol{\alpha}\in\mathbb{R}^p} ||\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}||_2^2 \text{ s.t. } ||\boldsymbol{\alpha}||_0 \leq L$$

1: $\Gamma = \emptyset$.
2: **for** *iter* $= 1, \ldots, L$ **do**
3: Select the atom which most reduces the objective

$$\hat{\imath} \leftarrow \arg\min_{i\in\Gamma^C} \left\{ \min_{\boldsymbol{\alpha}'} ||\mathbf{x} - \mathbf{D}_{\Gamma\cup\{i\}}\boldsymbol{\alpha}'||_2^2 \right\}$$

4: Update the active set: $\Gamma \leftarrow \Gamma \cup \{\hat{\imath}\}$.
5: Update the residual (orthogonal projection)

$$\mathbf{r} \leftarrow (\mathbf{I} - \mathbf{D}_\Gamma(\mathbf{D}_\Gamma^T\mathbf{D}_\Gamma)^{-1}\mathbf{D}_\Gamma^T)\mathbf{x}.$$

6: Update the coefficients

$$\boldsymbol{\alpha}_\Gamma \leftarrow (\mathbf{D}_\Gamma^T\mathbf{D}_\Gamma)^{-1}\mathbf{D}_\Gamma^T\mathbf{x}.$$

7: **end for**

Lasso

Nonsmooth optimization

Directional derivatives and subgradients are useful tools for studying ℓ_1-decomposition problems:

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^p} \frac{1}{2} ||\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}||_2^2 + \lambda ||\boldsymbol{\alpha}||_1$$

In this tutorial, we use the **directional derivatives** to derive simple optimality conditions of the Lasso.

变为l-1 norm regularization，直接求导获得最优条件。

拉格朗日乘子法:未知拉格朗日乘子λ

$$\min_{x,\lambda} \frac{1}{2}\|y - Ax\|_2^2 + \lambda\|x\|_1$$

当正则化项有比较nice的性质时，如piecewise linear, piecewise quadratic，即objective function是λ的分段线性函数，这时能够比较好的求解。

拉格朗日乘子法:已知拉格朗日乘子λ，优化无约束凸优化问题

$$\min_{x} \frac{1}{2}||y - Ax||_2^2 + \lambda||x||_1$$

- For this **specific** problem, coordinate descent is **convergent**.
- Supposing $||\mathbf{d}_i||_2 = 1$, updating the coordinate i:

$$\alpha[i] \leftarrow \arg\min_{\beta} \frac{1}{2}||\underbrace{\mathbf{x} - \sum_{j \neq i} \alpha[j]\mathbf{d}_j}_{\mathbf{r}} - \beta\mathbf{d}_i||_2^2 + \lambda|\beta|$$

$$\leftarrow \text{sign}(\mathbf{d}_i^T\mathbf{r})(|\mathbf{d}_i^T\mathbf{r}| - \lambda)^+$$

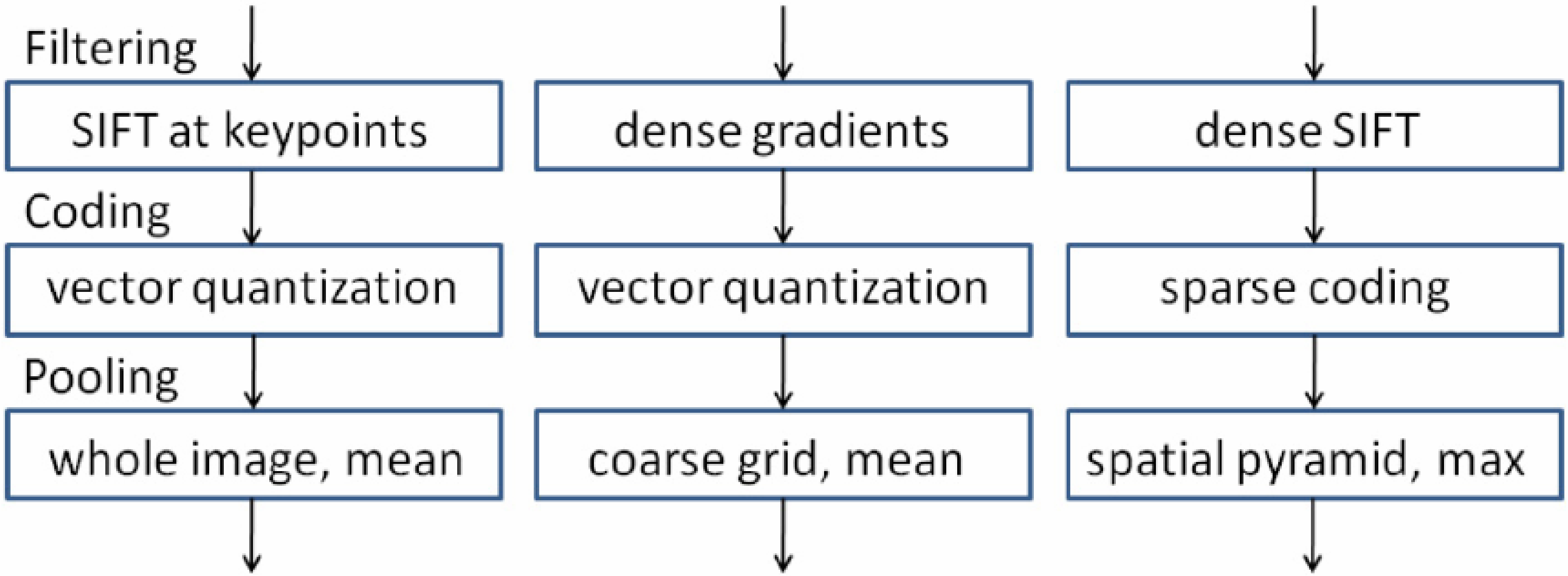
未知矩阵A，如何构造A，使得这一字典（矩阵）下 的表示最稀疏。即为如下优化问题：

$$\min_{x_i,A}\sum_i \frac{1}{2}||y_i - Ax_i||_2^2 + \lambda||x_i||_1$$

这个优化是非凸的（优化变量相乘）。
block descent(batch)：首先固定A ，优化 xi（相当于多个独立的1范数优化问题）；其次将计算出的 xi固定，优化A ，这就是一个（可能带约束）的least square问题。如此反复，直到算法收敛到某个（局部）极小值。

$$\begin{cases} \boldsymbol{\alpha}_t \leftarrow \underset{\boldsymbol{\alpha}\in\mathbb{R}^p}{\arg\min} \frac{1}{2}||\mathbf{x}_t - \mathbf{D}_{t-1}\boldsymbol{\alpha}||_2^2 + \lambda||\boldsymbol{\alpha}||_1, \\ \mathbf{D}_t \leftarrow \underset{\mathbf{D}\in\mathcal{C}}{\arg\min} \frac{1}{t}\sum_{i=1}^{t}\left(\frac{1}{2}||\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}_i||_2^2 + \lambda||\boldsymbol{\alpha}_i||_1\right). \end{cases}$$

Learning Codebooks for Image Classification



Idea

Replacing Vector Quantization by Learned Dictionaries!

Learning Codebooks for Image Classification

Let an image be represented by a set of low-level descriptors $\mathbf{x}_i$ at N locations identified with their indices $i = 1, \ldots, N$.

- hard-quantization:

$$\mathbf{x}_i \approx \mathbf{D}\boldsymbol{\alpha}_i, \quad \boldsymbol{\alpha}_i \in \{0,1\}^p \ \text{ and } \ \sum_{j=1}^{p} \boldsymbol{\alpha}_i[j] = 1$$

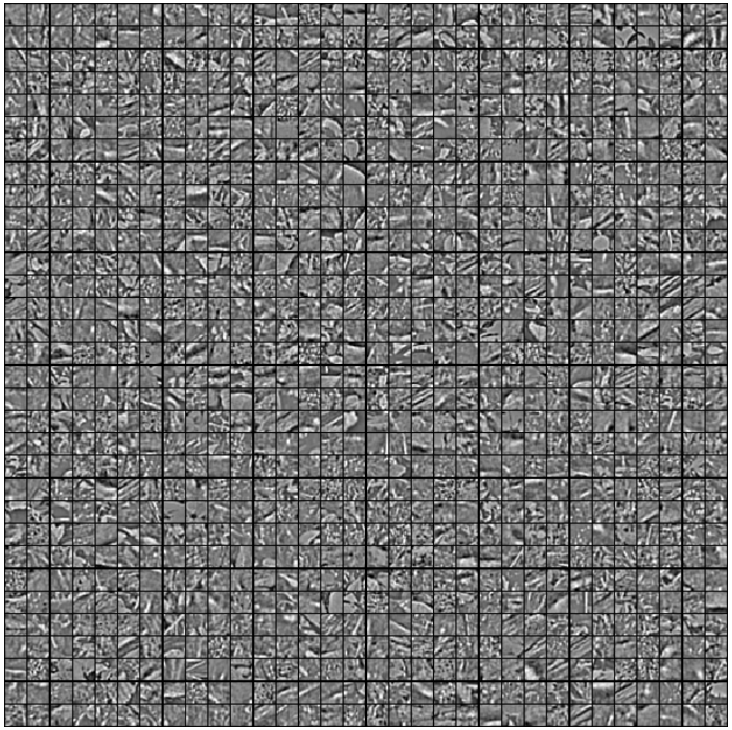
- soft-quantization:

$$\boldsymbol{\alpha}_i[j] = \frac{e^{-\beta\|\mathbf{x}_i - \mathbf{d}_j\|_2^2}}{\sum_{k=1}^{p} e^{-\beta\|\mathbf{x}_i - \mathbf{d}_k\|_2^2}}$$

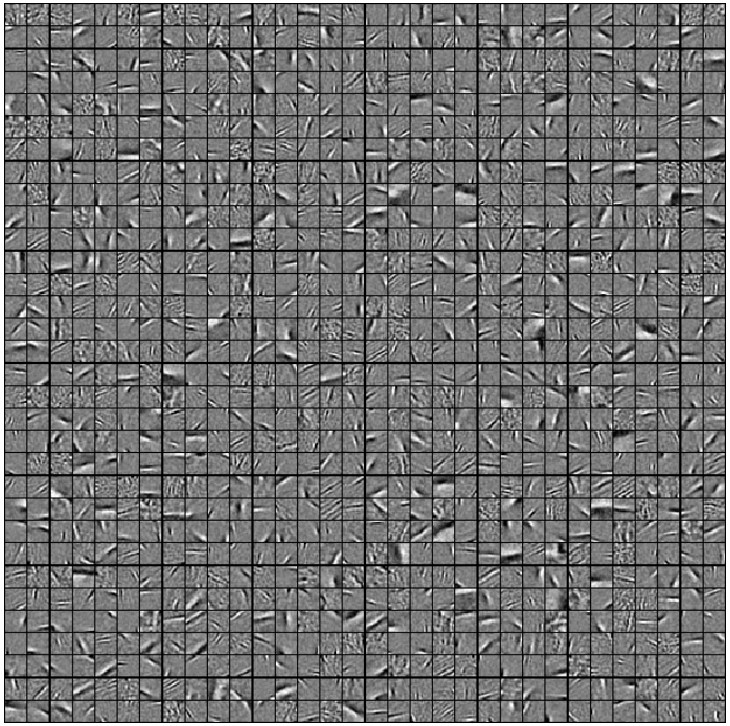
- sparse coding:

$$\mathbf{x}_i \approx \mathbf{D}\boldsymbol{\alpha}_i, \quad \boldsymbol{\alpha}_i = \arg\min_{\boldsymbol{\alpha}} \frac{1}{2}\|\mathbf{x}_i - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1$$

图像分类普通的字典和SC学到的字典对比:普通字典包含大量的纹理细节信息，视词内容仍然为低层图像内容；SC字典包含都是边缘线条信息，基向量内容已经变成抽象高层图像特征，这使得稀疏编码后特征的区分能力更强。



普通字典



SC字典

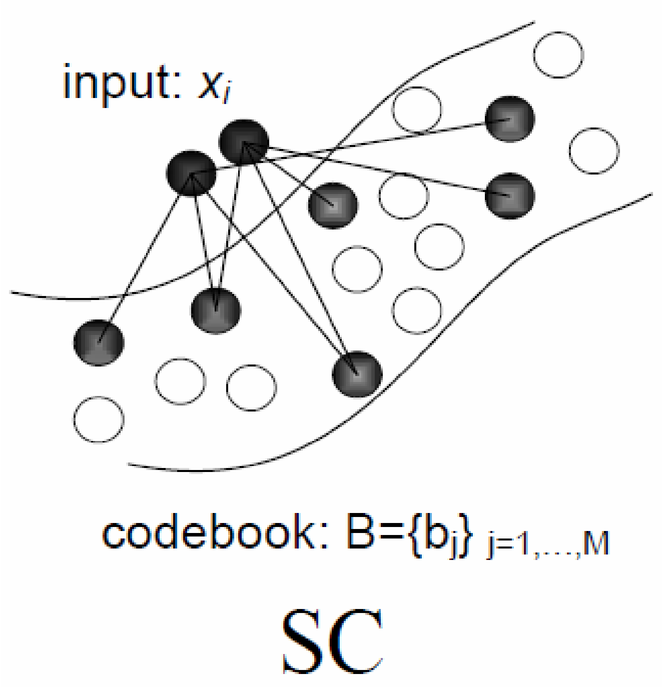
Learning Codebooks for Image Classification

Table from Boureau et al. [2010]

Method	Caltech-101, 30 training examples		15 Scenes, 100 training examples	
	Average Pool	Max Pool	Average Pool	Max Pool
	Results with basic features, SIFT extracted each 8 pixels			
Hard quantization, linear kernel	51.4 ± 0.9 [256]	64.3 ± 0.9 [256]	73.9 ± 0.9 [1024]	80.1 ± 0.6 [1024]
Hard quantization, intersection kernel	64.2 ± 1.0 [256] **(1)**	64.3 ± 0.9 [256]	80.8 ± 0.4 [256] **(1)**	80.1 ± 0.6 [1024]
Soft quantization, linear kernel	57.9 ± 1.5 [1024]	69.0 ± 0.8 [256]	75.6 ± 0.5 [1024]	81.4 ± 0.6 [1024]
Soft quantization, intersection kernel	66.1 ± 1.2 [512] **(2)**	70.6 ± 1.0 [1024]	81.2 ± 0.4 [1024] **(2)**	83.0 ± 0.7 [1024]
Sparse codes, linear kernel	61.3 ± 1.3 [1024]	$\mathbf{71.5 \pm 1.1}$ [1024] **(3)**	76.9 ± 0.6 [1024]	83.1 ± 0.6 [1024] **(3)**
Sparse codes, intersection kernel	70.3 ± 1.3 [1024]	$\mathbf{71.8 \pm 1.0}$ [1024] **(4)**	83.2 ± 0.4 [1024]	$\mathbf{84.1} \pm \mathbf{0.5}$ [1024] **(4)**
	Results with macrofeatures and denser SIFT sampling			
Hard quantization, linear kernel	55.6 ± 1.6 [256]	70.9 ± 1.0 [1024]	74.0 ± 0.5 [1024]	80.1 ± 0.5 [1024]
Hard quantization, intersection kernel	68.8 ± 1.4 [512]	70.9 ± 1.0 [1024]	81.0 ± 0.5 [1024]	80.1 ± 0.5 [1024]
Soft quantization, linear kernel	61.6 ± 1.6 [1024]	71.5 ± 1.0 [1024]	76.4 ± 0.7 [1024]	81.5 ± 0.4 [1024]
Soft quantization, intersection kernel	70.1 ± 1.3 [1024]	73.2 ± 1.0 [1024]	81.8 ± 0.4 [1024]	83.0 ± 0.4 [1024]
Sparse codes, linear kernel	65.7 ± 1.4 [1024]	$\mathbf{75.1 \pm 0.9}$ [1024]	78.2 ± 0.7 [1024]	83.6 ± 0.4 [1024]
Sparse codes, intersection kernel	73.7 ± 1.3 [1024]	$\mathbf{75.7} \pm 1.1$ [1024]	83.5 ± 0.4 [1024]	$\mathbf{84.3} \pm \mathbf{0.5}$ [1024]

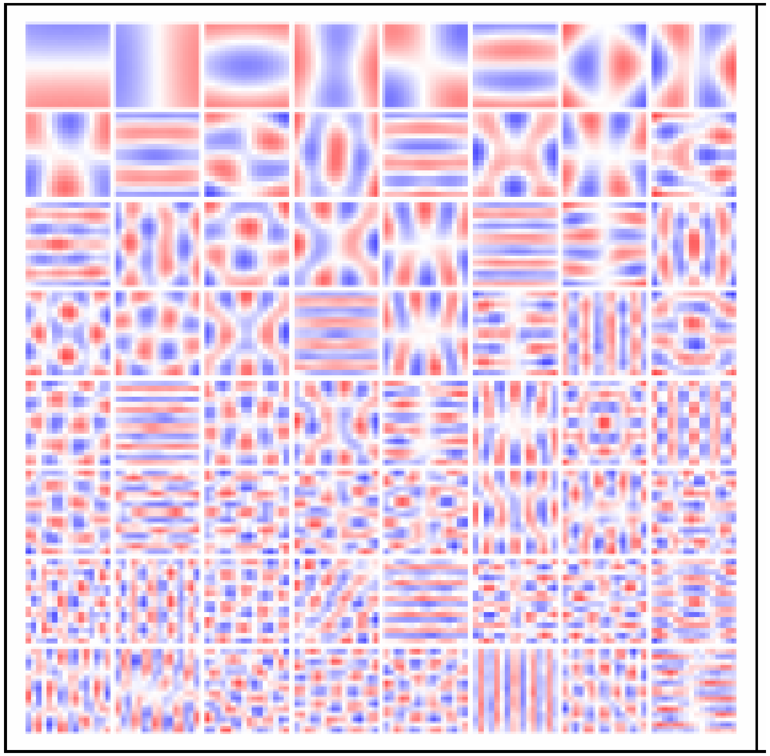
Structured sparsity

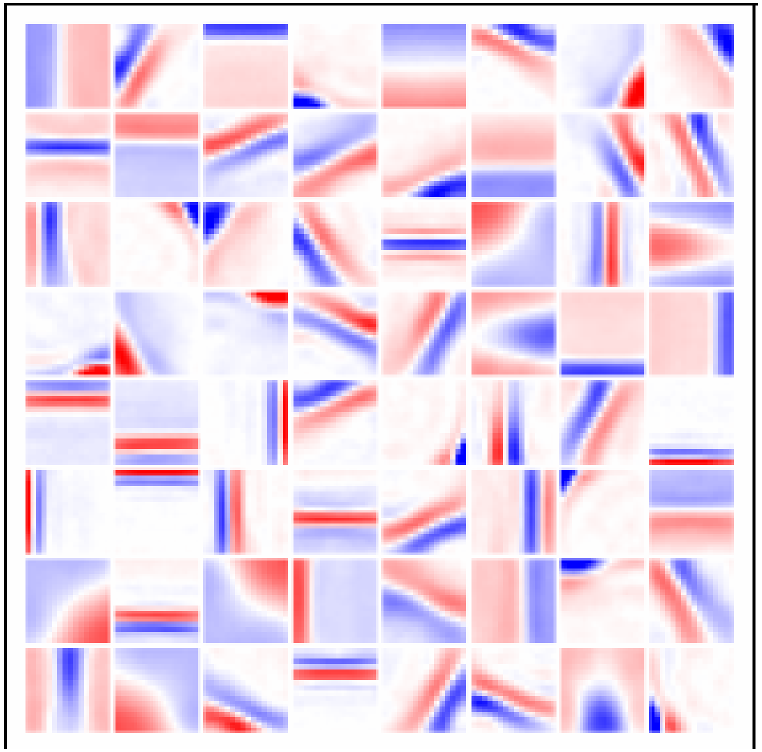
- 由于稀疏字典的过完备性，冗余基的存在使稀疏解不唯一，相似特征可能被量化的不同的子空间中，出现相似特征的输出编码一致性问题。



Structured sparsity

- 字典原子杂乱，无序，不利于编码的分析处理和字典的扩展





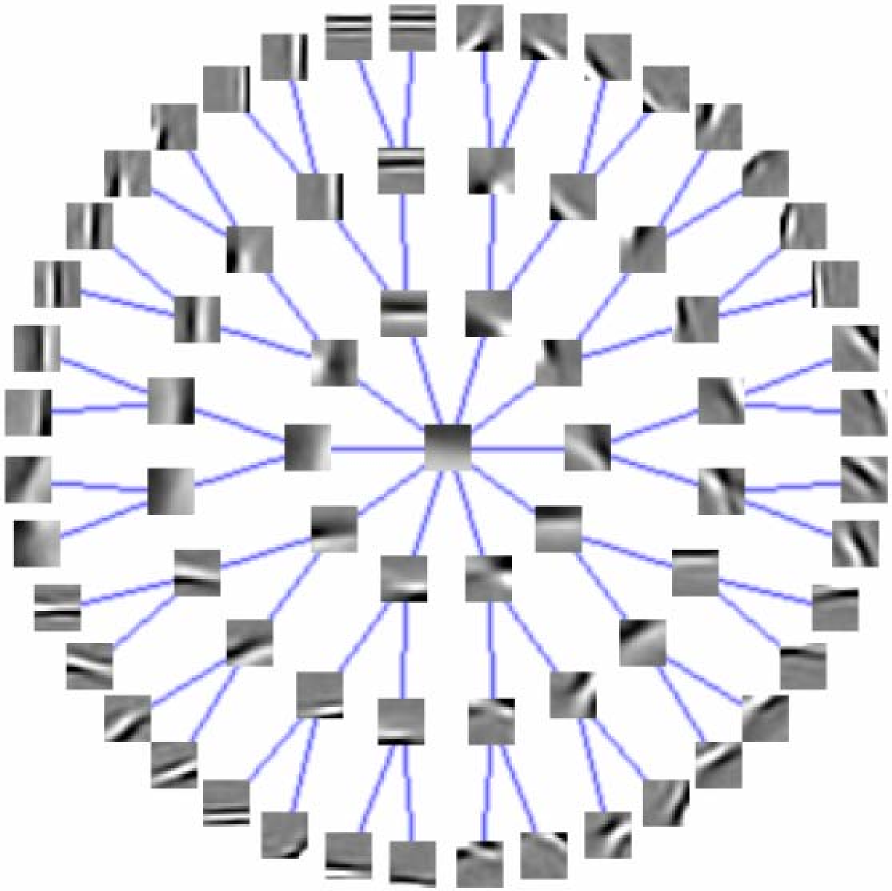
Structured sparsity

- 结构稀疏性对字典作出约束，使其符合一定的规范
- 学习过程中使字典原子嵌入的预先指定的结构中，比如分组稀疏，树结构稀疏等
- 相似图像具有相似的稀疏模式

Structured sparsity

树结构等级字典，对任何输入信号x
的稀疏编码中，非零系数的父节点总是
非零的，而零系数的所有支节点都是
零。

低频特征更靠近根节点
高频特征更接近叶子节点



Outline

- ***Sparse representation***
  - ***Sparse coding***
  - ***Optimization for sparse coding***
  - ***Dictionary learning for signal reconstruction***
  - ***Recent advances in Computer Vision***

- ***Sparse methods on matrices***
  - ***Matrix completion (pLSA, collaborative filtering)***
  - ***Matrix factorization (topic model, LDA, low-rank)***
  - ***Some applications***

- 在很多的具体问题中，信号或者数据往往可以用矩阵来表示，使得对数据的理解、建模、处理和分析更为方便。然而这些数据经常面临缺失、损坏、受噪声污染等等问题。如何在各种情况下得到干净、准确、结构性良好的数据，就是矩阵重建所要解决的问题。

- 大致来讲，矩阵重建分为矩阵填充（Matrix Completion）和矩阵恢复（Matrix Recovery）两大类。前者主要研究如何在数据不完整的情况下将缺失数据进行填充，后者主要研究在某些数据受到严重损坏的情况下恢复出准确的矩阵。

矩阵填充（Matrix Completion）考虑的是这样一个问题，对于某个矩阵，我们只能采样得到矩阵的一部分元素，其它一部分或者大部分元素由于各种原因丢失了或无法得到，如何将这些空缺的元素合理准确地填充。即求解如下优化问题：

$$\begin{aligned} \min \quad & \operatorname{rank}(X), \\ \text{subject to} \quad & X_{ij} = M_{ij}, (i,j) \in \Omega, \end{aligned}$$

这个模型的意义是说，将空缺的元素填充之后使得矩阵的结构尽可能好，即秩尽可能低。然而不幸的是，这是一个NP-Hard的问题，在理论和实践中，均只存在指数复杂度（相对于矩阵维数n）的算法

我们知道，一个矩阵的秩r与它的非零奇异值的个数相同。于是有一个选择是用矩阵的奇异值的和，即核范数，来近似地替代矩阵的秩：

$$
\begin{aligned}
&\min \quad \|X\|_*, \\
&\text{subject to} \quad X_{ij} = M_{ij}, (i,j) \in \Omega,
\end{aligned}
$$

由于核范数是凸的而零范数并非数学意义上的范数，于是原来的问题就转化为一个凸优化问题，但它仍然是一个比较难计算的问题，可以考虑通过各种迭代法，尤其是梯度法进行求解。关于这种近似理论上的可靠性，有一些相关证明。

Algorithm	表达	备注
Singular Value Thresholding	$\min \quad \tau\|X\|_* + \frac{1}{2}\|X\|_F^2,$ $\text{subject to} \quad P_\Omega(X) = P_\Omega(M),$	求解原问题的近似问题
Accelerated Proximal Gradient	$\min \quad F(X) = \frac{1}{2}\|P_\Omega(X-M)\|_F^2 + \mu\|X\|_*.$	无约束优化问题
ALM	$\min \quad \|A\|_*,$ $\text{subject to} \quad A+E=D,\ P_\Omega(E)=0.$	部分增广拉格朗日函数
……	……	……

矩阵恢复是指当矩阵的某些元素被严重破坏后，自动识别出被破坏的元素，恢复出原矩阵。同样，假定原矩阵有非常良好的结构，即是低秩的；另外，假定只有很少一部分元素被严重破坏，即噪声是稀疏的但大小可以任意。于是矩阵恢复可用如下优化问题来描述：

$$\begin{aligned} \min \quad & \operatorname{rank}(A) + \lambda \|E\|_0, \\ \text{subject to} \quad & A + E = D, \end{aligned}$$

用矩阵的核范数近似秩，矩阵的1范数来近似零范数，转化为凸优化问题：

$$\begin{aligned} \min \quad & \|A\|_* + \lambda \|E\|_1, \\ \text{subject to} \quad & A + E = D. \end{aligned}$$

Matrix factorization

Algorithm	表达	备注
Iterative Thresholding	$\min_{A,E} \quad \|A\|_* + \lambda \|E\|_1 + \frac{1}{2\tau} \|A\|_F^2 + \frac{1}{2\tau} \|E\|_F^2,$ subject to $\quad A + E = D,$	此问题的解近似于原问题
Accelerated Proximal Gradient	$\min \quad F(A,E) = \frac{1}{2}\|D - A - E\|_F^2 + \mu(\|A\|_* + \lambda\|E\|_1).$	无约束优化问题
ALM	$L(A,E,Y,\mu) = \|A\|_* + \lambda\|E\|_1 + \langle Y, D - A - E\rangle + \frac{\mu}{2}\|D - A - E\|_F^2.$	部分增广拉格朗日函数
……	……	……

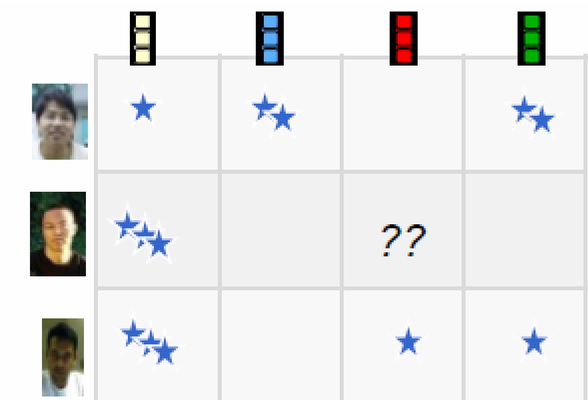
- 矩阵填充的一个著名应用是Netflix推荐系统。Netflix是世界上最大的在线影片租赁服务商，从2006年10月份开始举办Netflix大奖赛。它公开了大约一亿个1～5级的匿名电影评级，来自大约48万个客户对1.8万部电影的评价，比赛要求参赛者预测Netflix客户分别喜欢什么影片。

- 这是一个典型的矩阵填充问题，即矩阵的每一行对应某个用户对电影的评级，每一列表示某电影在所有用户中的评级，但是每个用户只可能对一部分电影进行评价，所以我们可以通过矩阵填充得出用户对每部电影的喜好程度。

- 假设现在有m个用户和n部电影，如果把所有评分列成一张大表，可以得到矩阵 $D \in R^{m\times n}$ 。

Some applications

➢评级矩阵。



$$D \in \mathbb{R}^{m \times n}$$

movies

users

	2		1			4				5	
	5		4				?		1		3
		3		5			2				
4			?			5		3		?	
		4		1	3				5		
			2				1	?			4
	1					5		5		4	
		2		?	5		?		4		
	3		3		1		5		2		1
	3				1			2		3	
	4			5	1			3			
		3				3	?			5	
2	?		1		1						
		5			2	?		4		4	
	1		3		1	5		4		5	
1		2			4				5	?	

- 由AT&T实验室成员组成的BellKors Pragmatic Chaos提出的collaborative filtering 算法比较有效的解决了上述问题，并赢得了100万美元。表示为数学形式即为矩阵填充问题：

$$\begin{aligned} \min \quad & \|X\|_*, \\ \text{subject to} \quad & X_{ij} = M_{ij}, (i,j) \in \Omega, \end{aligned}$$

- 解释起来就是，存在一个latent factor空间，不同的用户群体位于这个latent factor空间的不同位置，体现了不同用户的喜好。如果可以把用户喜好连同潜在的latent factor一同计算出来，预测也自然水到渠成了。从某种角度来看，奇异值分解过程也就是上述的剥离latent factor和用户喜好的过程 。

➢由

Singular Value Decomposition

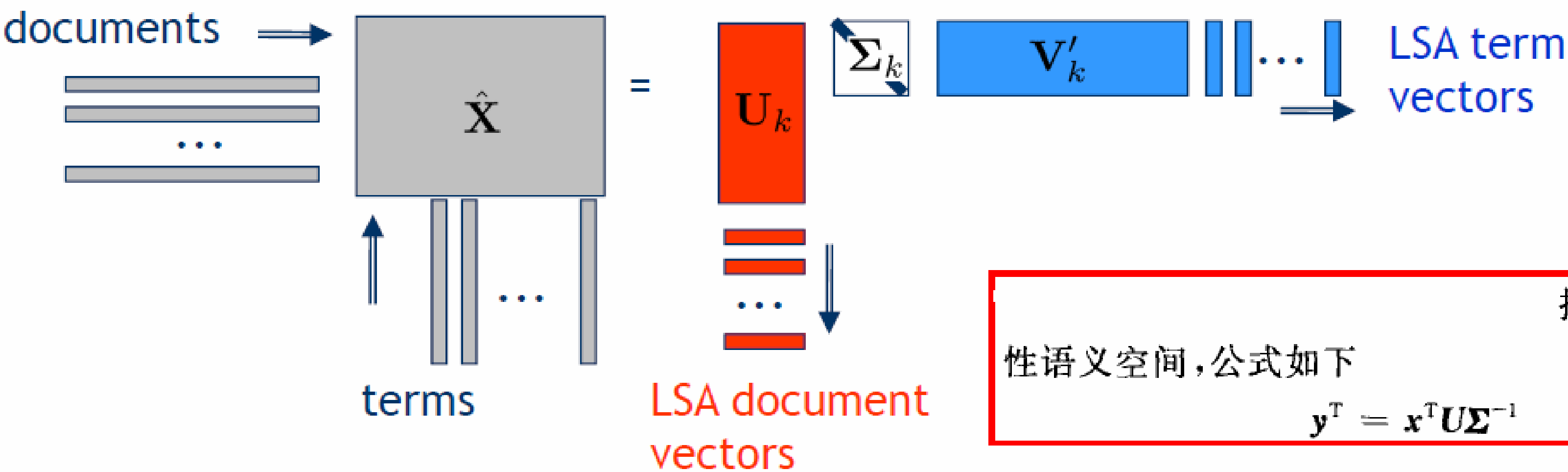
- For an arbitrary matrix X there exists a factorization (Singular Value Decomposition = **SVD**) as follows:

$$X = U \, \Sigma \, V^T \in \mathcal{R}^{n \times m}$$

- Where
  - (i) $U \in \mathcal{R}^{n \times k}$ $\quad \Sigma \in \mathcal{R}^{k \times k}$ $\quad V \in \mathcal{R}^{m \times k}$
  - (ii) $\mathbf{U}'\mathbf{U} = \mathbf{I}$ $\quad \mathbf{V}'\mathbf{V} = \mathbf{I}$ Orthonormal columns
  - (iii) $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_k), \; \sigma_i \geq \sigma_{i+1}$ Singular values (ordered)
  - (iv) $k = \mathrm{rank}(X)$

Latent Semantic Analysis (Indexing)

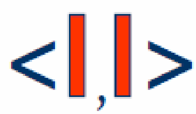
- The LSA via SVD can be summarized as follows:



把文档变换到隐性语义空间，公式如下

$$\boldsymbol{y}^{\mathrm{T}} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{\Sigma}^{-1}$$

- Document **similarity**



- Folding-in **queries**

$$\hat{\mathbf{q}} = \Sigma_k^{-1}\mathbf{V}_k\mathbf{q}$$

M. Berry, S. Dumais, and G. O'Brien. Using linear algebra for intelligent information retrieval. SIAM Review, 37(4):573-595, 1995.

Some applications

Index Words	Titles								
	T1	T2	T3	T4	T5	T6	T7	T8	T9
book			1	1					
dads						1			1
dummies		1						1	
estate							1		1
guide	1					1			
investing	1	1	1	1	1	1	1	1	1
market	1		1						
real							1		1
rich						2			1
stock	1		1					1	
value				1	1				

SVD

book	0.15	-0.27	0.04
dads	0.24	0.38	-0.09
dummies	0.13	-0.17	0.07
estate	0.18	0.19	0.45
guide	0.22	0.09	-0.46
investing	0.74	-0.21	0.21
market	0.18	-0.30	-0.28
real	0.18	0.19	0.45
rich	0.36	0.59	-0.34
stock	0.25	-0.42	-0.28
value	0.12	-0.14	0.23

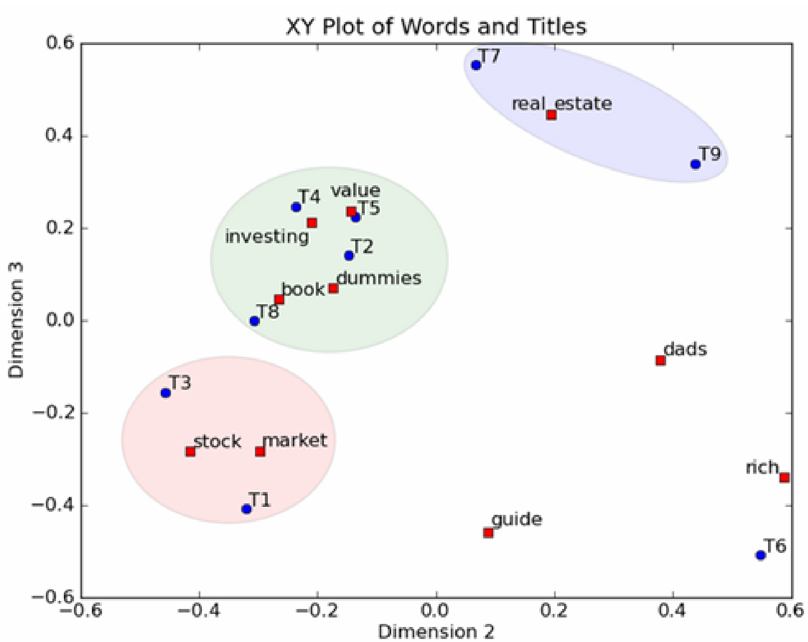
*

3.91	0	0
0	2.61	0
0	0	2.00

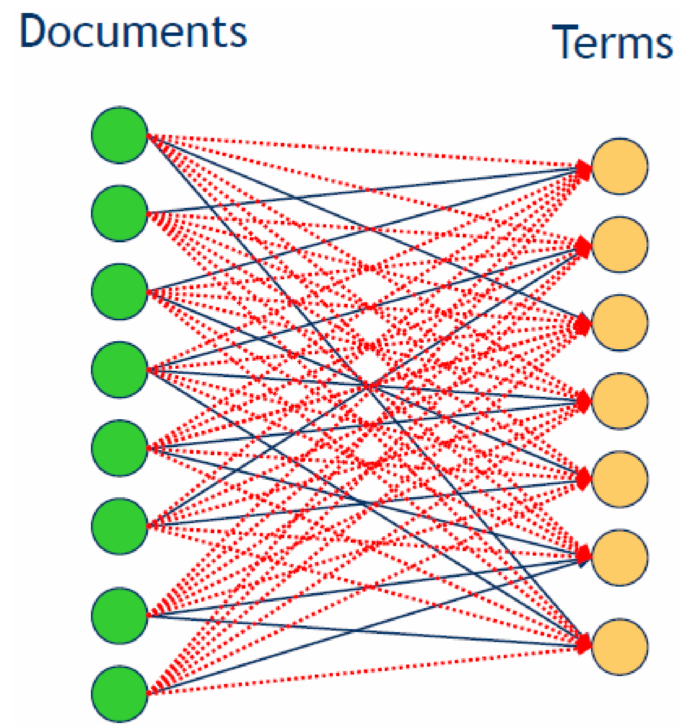
*

T1	T2	T3	T4	T5	T6	T7	T8	T9
0.35	0.22	0.34	0.26	0.22	0.49	0.28	0.29	0.44
-0.32	-0.15	-0.46	-0.24	-0.14	0.55	0.07	-0.31	0.44
-0.41	0.14	-0.16	0.25	0.22	-0.51	0.55	0.00	0.34

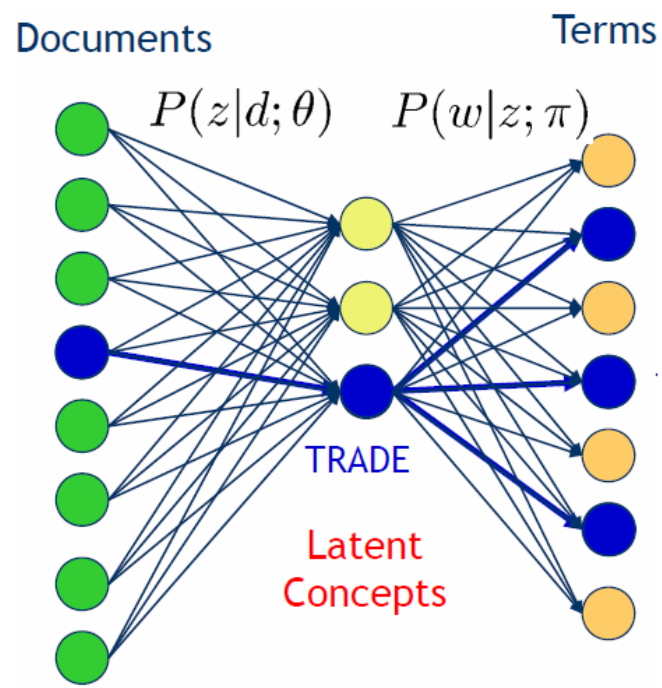
- 左奇异矩阵表示给定主题下的词项概率分布，右奇异矩阵表示给定主题下文档的概率分布。
- 这样将文档和词都转化到了主题空间，即实现了降维，有利于进一步的聚类、相似查询等操作。



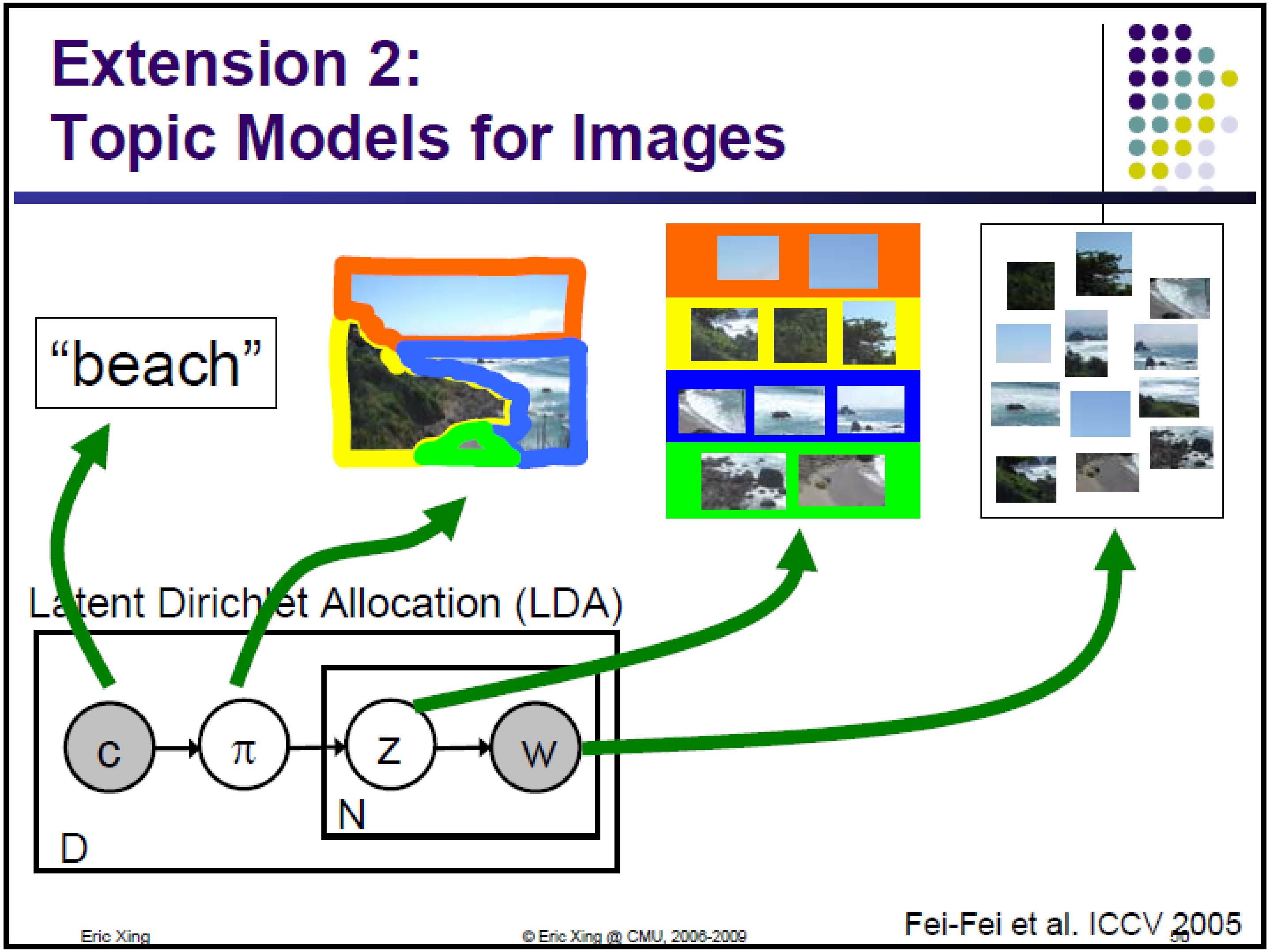
LSA不是概率生成模型，LSA以最优低秩逼近为优化的目标，而pLSA以观测值的似然值为优化目标。pLSA是LSA的概率扩展。



$$\hat{P}_{\mathrm{ML}}(w|d) = \frac{n(d,w)}{\sum_{w'} n(d,w')}$$

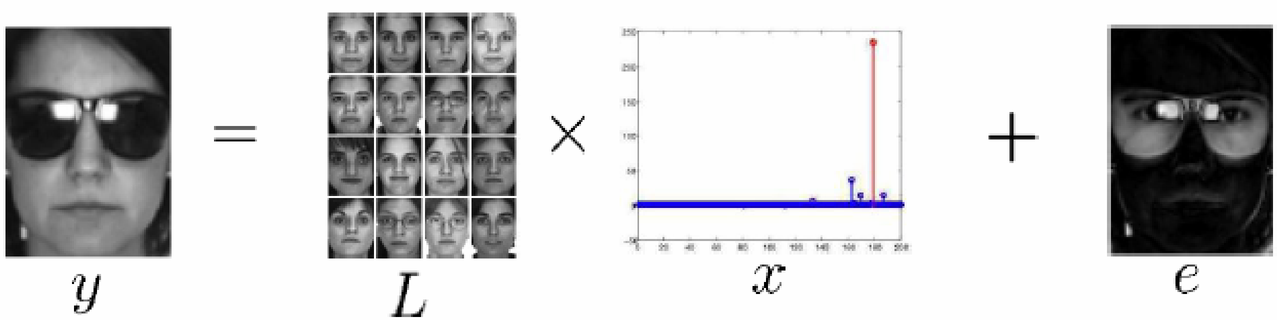


$$\hat{P}_{pLSA}(w|d) = \sum_z P(w|z;\theta)P(z|d;\pi)$$


Extension 2:
Topic Models for Images
"beach"
Latent Dirichlet Allocation (LDA)
c
π
z
w
N
D
Fei-Fei et al. ICCV 2005
Eric Xing
© Eric Xing @ CMU, 2006-2009

由矩阵填充还可以用于视频去噪。根据某一像素值是否背离同一位置处所有像素的“均值”判定该点是否可靠，进而用矩阵填充来得到那些被噪声污染的像素。

Robust recovery: Given $y = Lx_0 + e_0$, $L \in \mathbb{R}^{m \times n}$, $m \ll n$, recover x_0.



Some applications — Background modeling from video

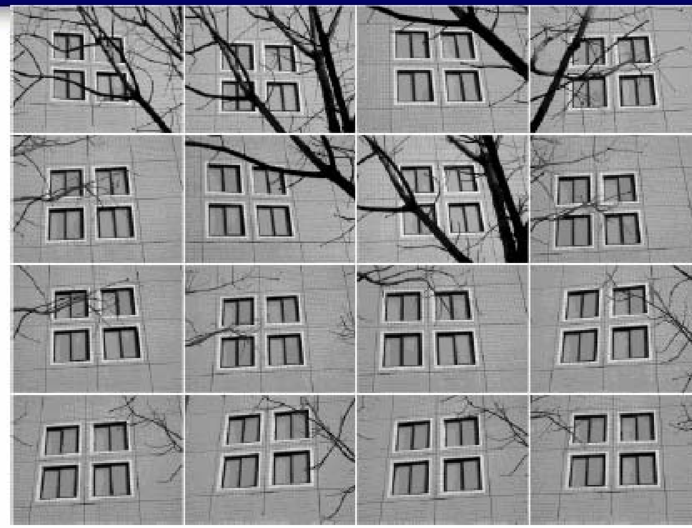
Static camera surveillance video

200 frames,
72 x 88 pixels,

Significant foreground motion

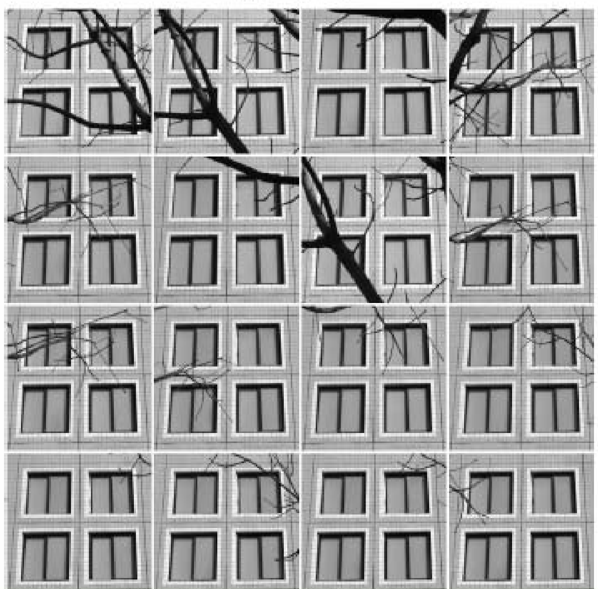
Video D | Low-rank appx. A | Sparse error E

D

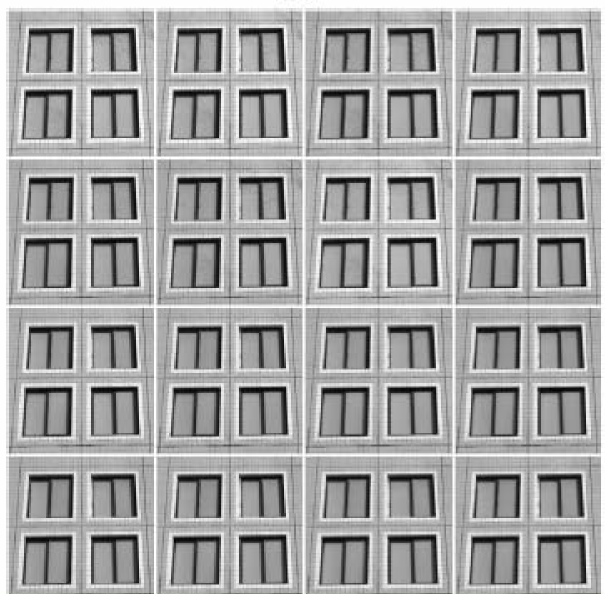


$\tau \in$ 2D homographies

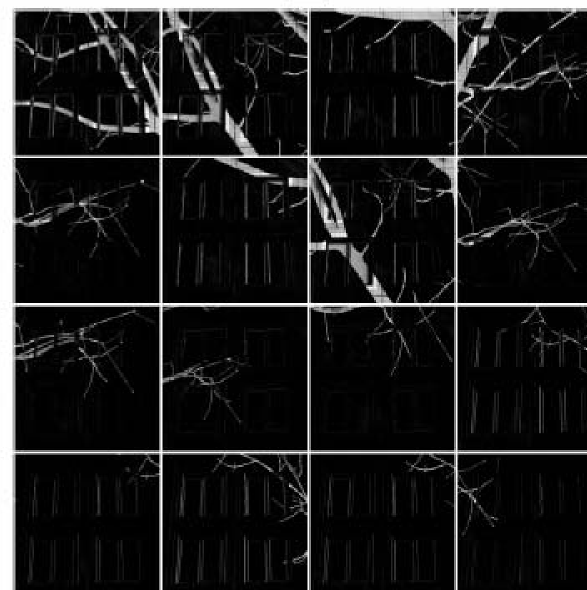
$D \circ \tau$



A



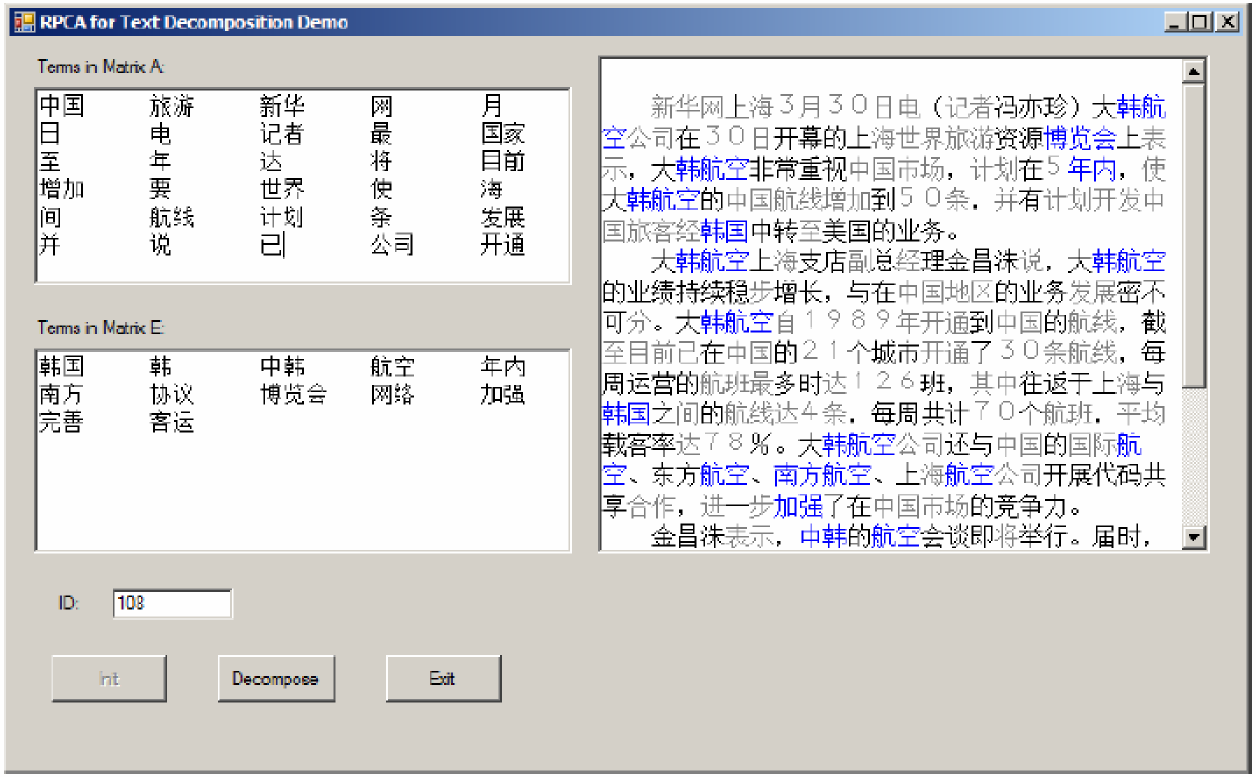
E



- Web document corpus analysis [Min et al.]

$$\min \|A\|_* + \lambda \|E\|_1 \text{ subj } D = A + E, E \geq 0.$$

  - D: tf-idf matrix, 42186 x 18320, 90.3h on an HPC cluster



THE BIG PICTURE – *Connection between rank and sparsity*

	Sparse Vector	***Low-Rank Matrix***
Degeneracy of	individual signal	a batch of signals
Measure	L0 norm	Rank
Convex Surrogate	L1 norm	Nuclear norm
Compressed Sensing	$y = Ax$	$Y = A(X)$
Error Correction	$y = Ax + e$	$Y = A(X) + E$
Sparsity-Rank Tradeoff	$Y = A(X) + B(E) + Z$	

References

[1] Emmanuel Candes, Justin Romberg, and Terence Tao. Stable Signal Recovery from Incomplete and Inaccurate Measurements. Communications on Pure and Applied Mathematics. 2006.

[2] http://www.di.ens.fr/~mairal/tutorial_iccv09/

[3] http://perception.csl.uiuc.edu/matrix-rank/home.html

[4] http://www.cvpr2009.org/call-for-short-courses

[5] http://www.cnblogs.com/LeftNotEasy/archive/2011/01/19/svd-and-applications.html

[6] Deng Cai. Learning with Local Consistency. Chinese Workshop on Machine Learning and Applications. 2010.

[7] http://www.cvchina.info/2010/06/01/sparse-representation-vector-matrix-tensor-1/

[8] http://www.cvchina.info/2010/07/13/sparse_representation_vector_matrix_tensor2/