第18回画像センシングシンポジウム　2012年06月08日

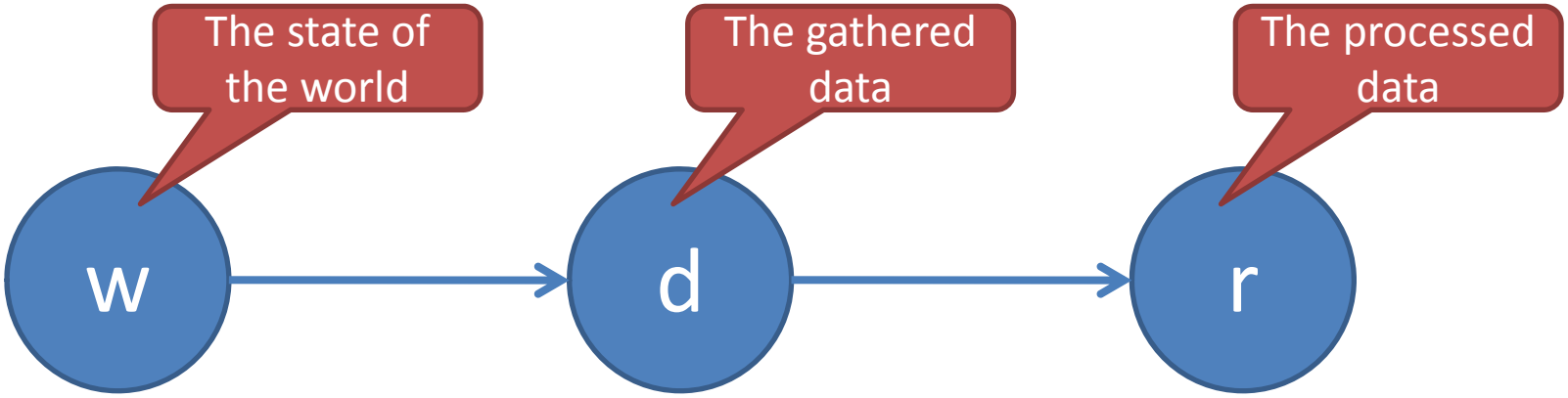
大規模画像データを用いた一般画像認識

東京大学/JSTさきがけ
原田達也

知識はどこから降ってくる？

- 人による教示
  - クラウドソーシング
- インターネットにある膨大な情報を利用
  - 大規模画像処理
- 対象とは別の知識を活用
  - 転移学習

The data processing theorem



Markov chain

$$P(w, d, r) = P(w)P(d|w)P(r|d)$$

The average information

$$I(W; D) \geq I(W; R)$$

The data processing theorem states that data processing can only destroy information.

David J.C. MacKay. Information Theory, Inference, and Learning Algorithms. Cambridge University Press 2003.

画像認識のプロセスと必要機能

訓練時

識別時

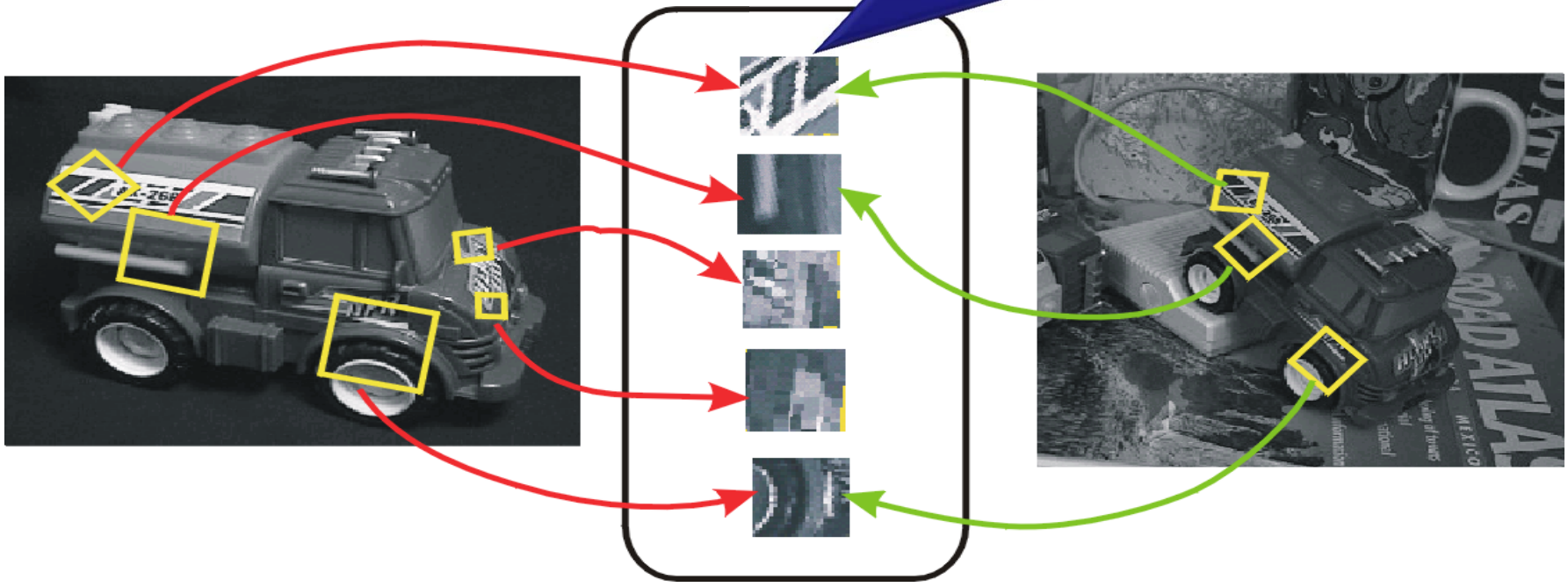
- 処理を重ねる毎にデータの持つ情報は減少
  - データ，特徴抽出，モデルの順に高い質が必要
  - データの大規模化，リッチな画像表現
- 複雑なモデルは大規模データの前では役立たず
  - スケーラビリティの重要性
  - 学習・識別が効率的な線形識別機の利用
- 線形識別機で複雑かつ多クラス問題に対応する必要
  - リッチな画像表現の高次元化

画像表現

局所特徴量による画像認識

- 検出器，記述子として，移動，回転，スケールに不変な特徴量を算出可能なものを利用
  - 例）SIFT
- 訓練画像データの局所記述子とテストデータの局所記述子を比較する
- 局所記述子の空間配置の一貫性を調べる
  - Hough 変換
  - RANSAC

一般物体認識は正確なマッチングが不可能
→局所特徴の分布の近さを考える！

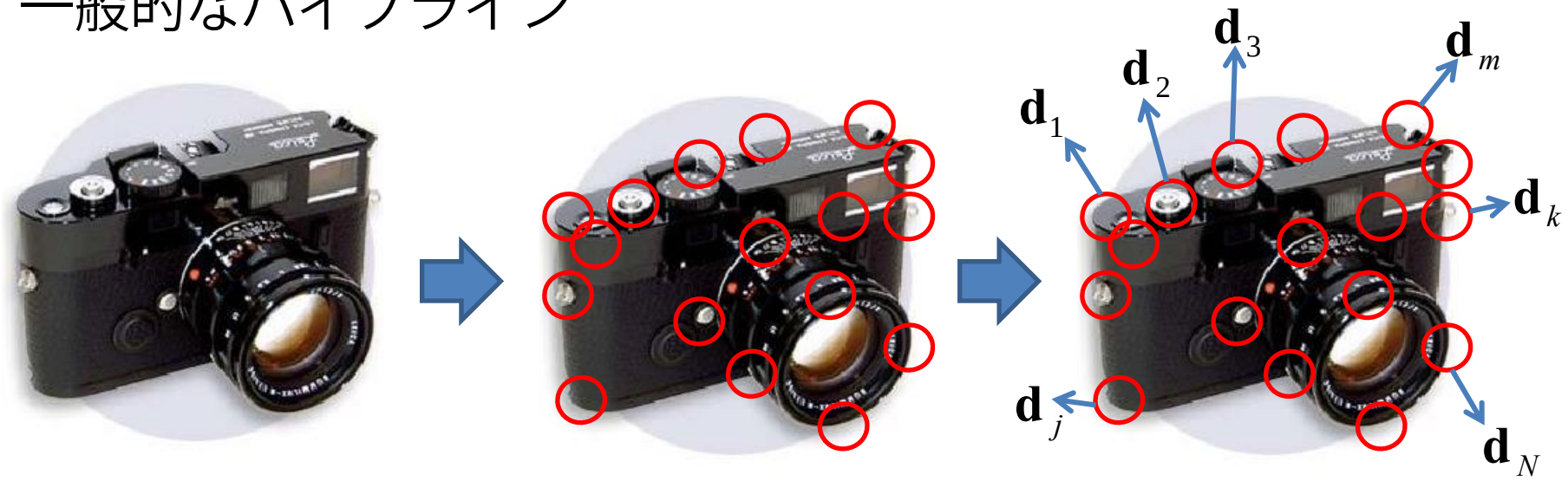


Local Features,
e.g. SIFT

D. Lowe

画像特徴とは？

一般的なパイプライン



1) Input Image

2) Detection

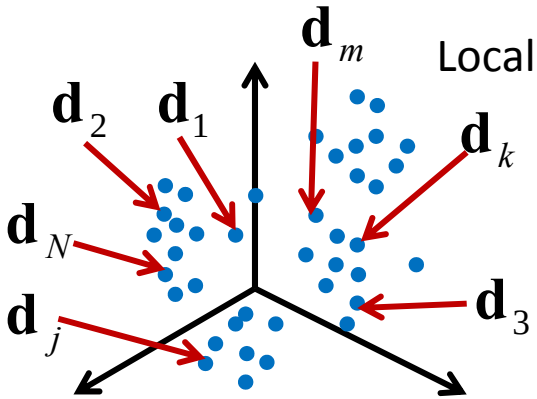
3) Description

4) Local descriptors in feature space

5) PDF estimation

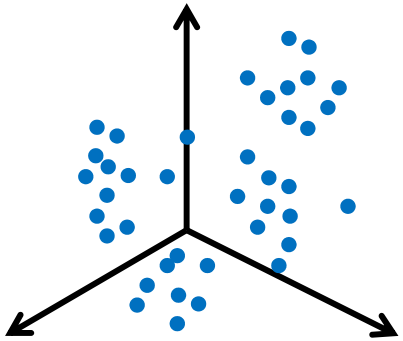
6) Feature vector

画像表現

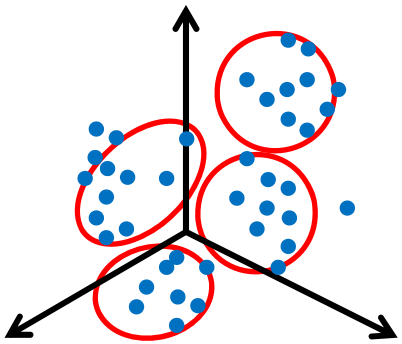


Local descriptors in feature space

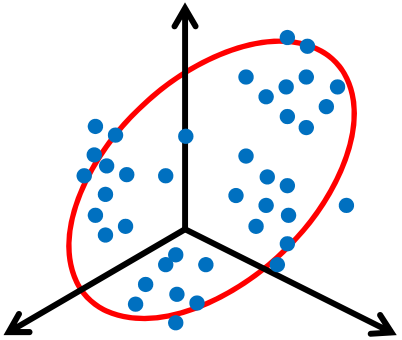
Descriptor matching



Codebook



Global feature



# of anchor points: large
Computational complexity: large

# of anchor points: small
Computational complexity: small

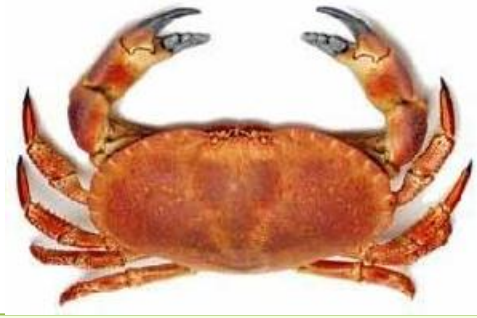
SVM-KNN
Naïve Bayes Nearest Neighbor
Graph Matching Kernel

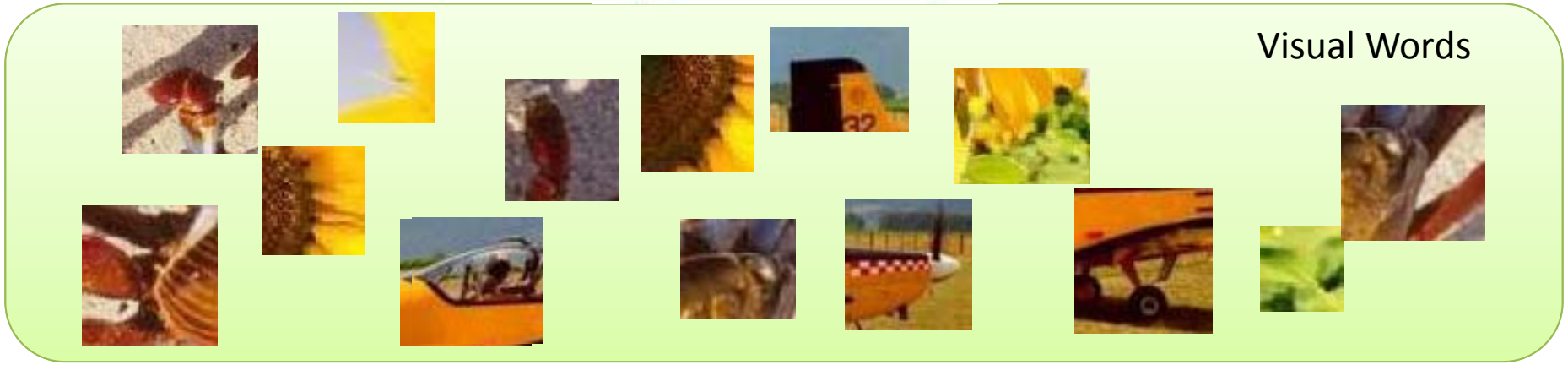
Bag of Visual Words
Gaussian Mixture Model
ScSPM, Super Vector, LLC
Fisher Vector

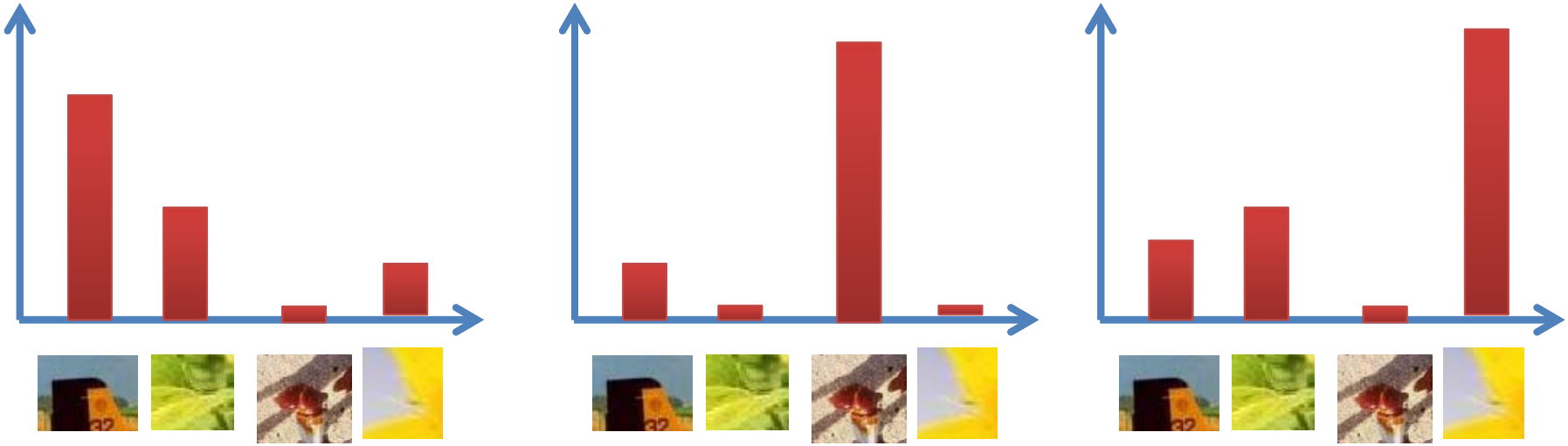
HLAC
GLC
Global Gaussian

画像表現
コードブック

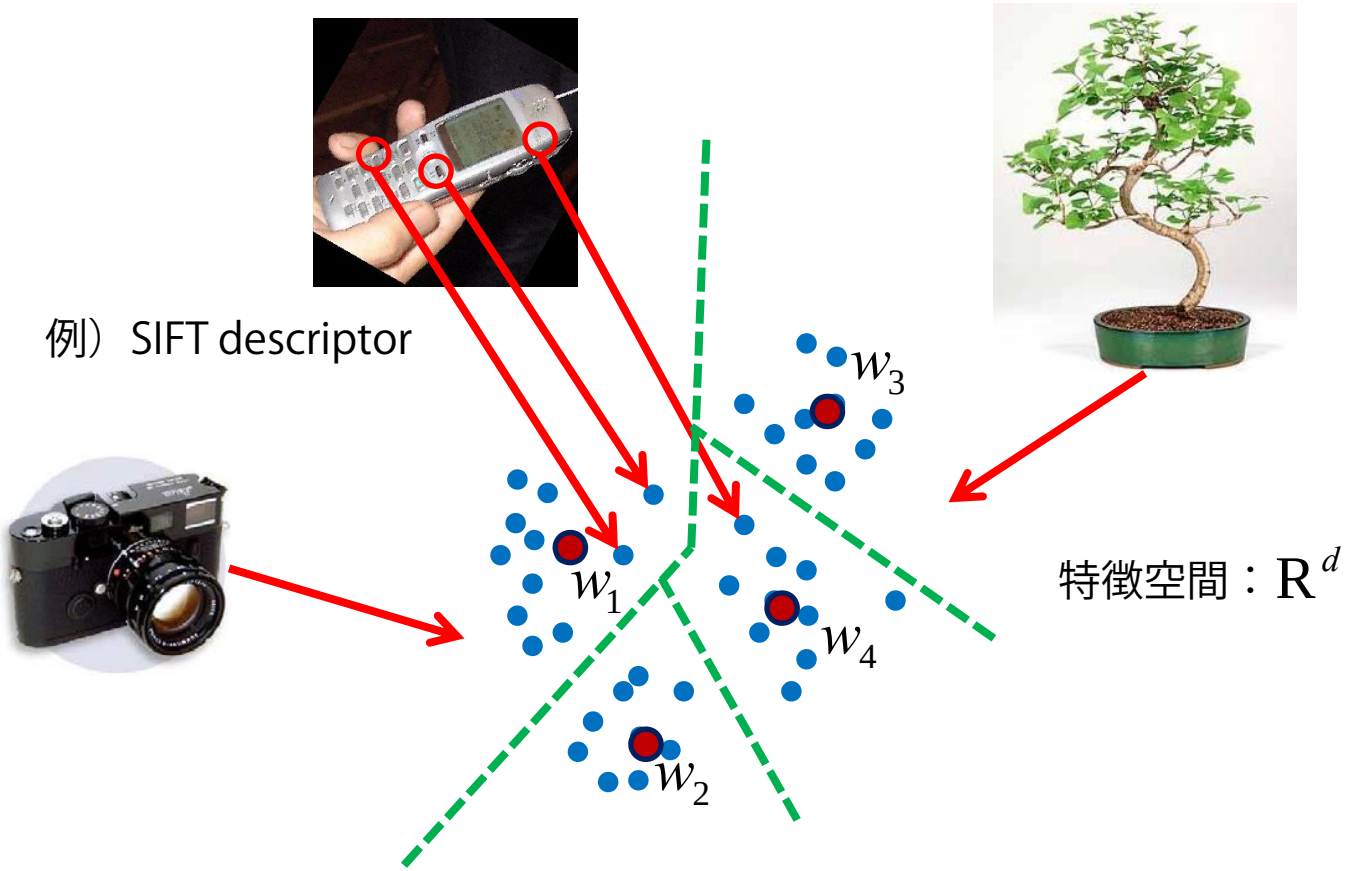
Bag of Visual Words?







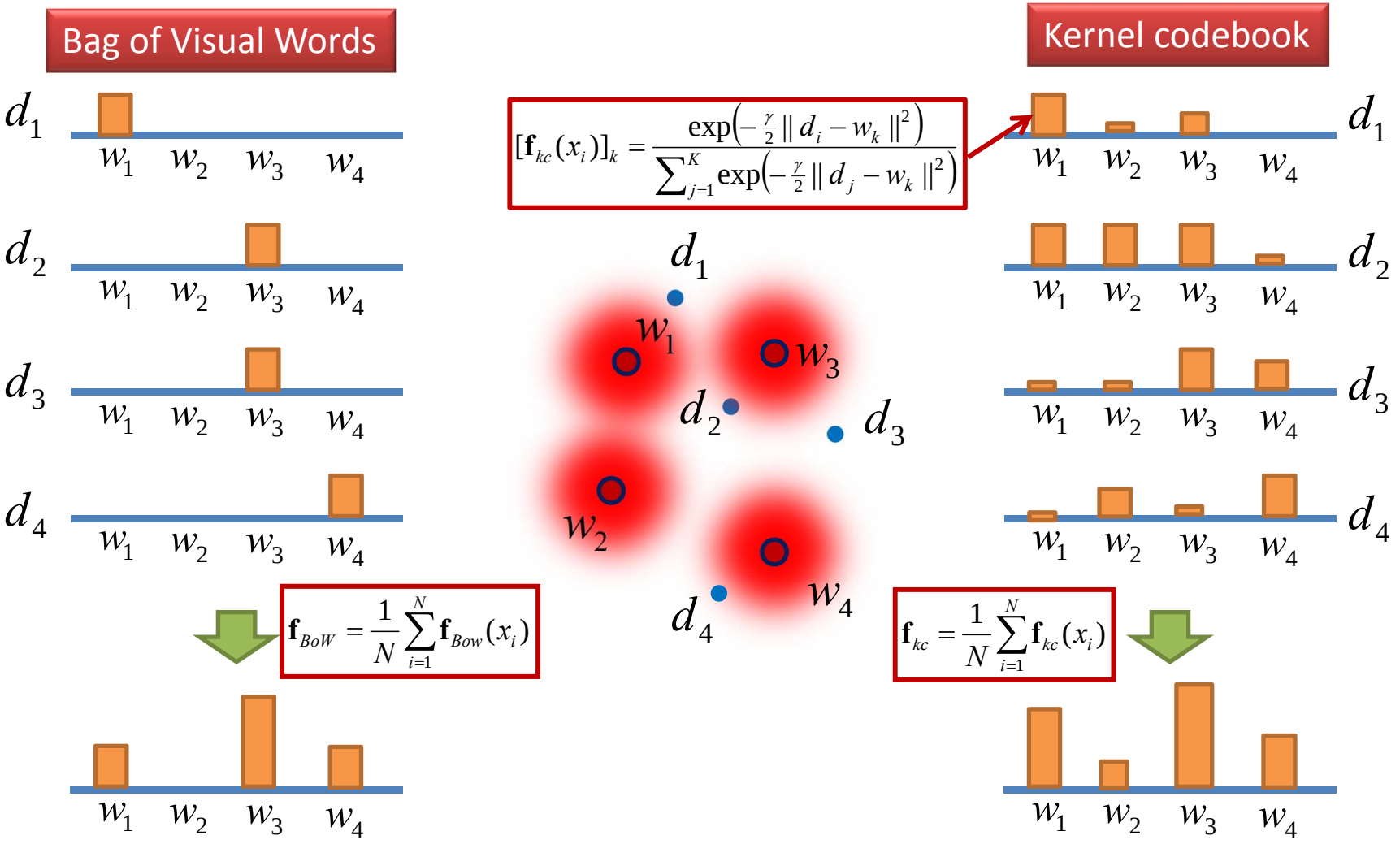
Code wordsの生成：clustering



- ベクトル量子化と呼ばれるプロセス
- 一般的にk-meansによるクラスタリング
  - 階層的クラスタリング：Vocabulary Tree
- 局所記述子にはSIFTがよく用いられる
  - もちろんSURFやRGB，Self Similarityでもよい

Kernel codebook

- 局所記述子を一つのコードワードに割り付けるのではなく，距離に応じた重み付けで全てのコードワードと関連づける．
- Jan C. van Gemert, Jan-Mark Geusebroek, Cor J. Veenman, and Arnold W.M. Smeulders. Kernel Codebooks for Scene Categorization. ECCV, 2008.



BoFのGMM利用による改善

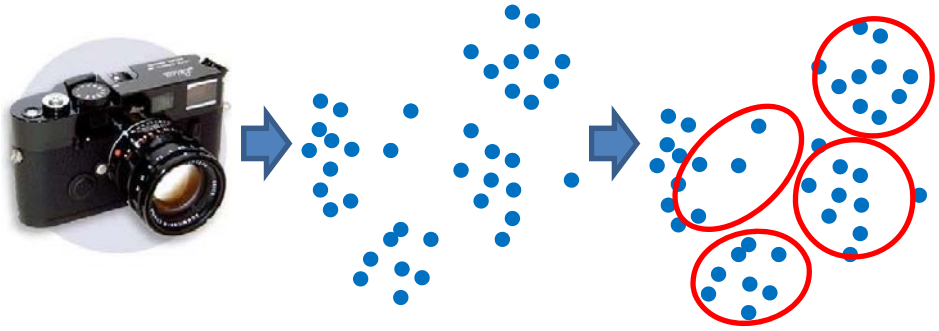


Image　　Local descriptors in feature space　　PDF estimation

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \sum_{k=1}^{K} \pi_k p_k(\boldsymbol{x})$$

$$\gamma_n(k) = p(k|\boldsymbol{x}_n, \boldsymbol{\theta}^{(t)}) = \frac{\pi_k p_k(\boldsymbol{x}_n)}{\sum_{j=1}^{K} \pi_j p_j(\boldsymbol{x}_n)}$$

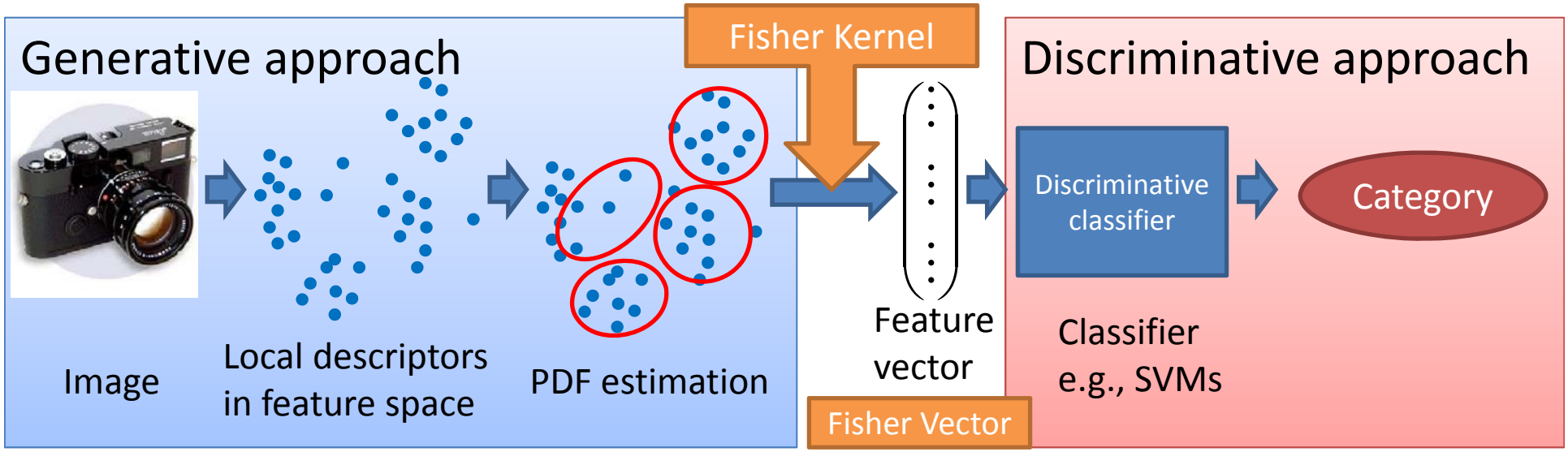
$$\boldsymbol{f} = \frac{1}{N} \sum_{n=1}^{N} [\gamma_n(1), \cdots, \gamma_n(K)]^\top \in R^K$$

- メリット
  - 混合ガウス分布を構成する各ガウス分布がそれぞれ共分散を持つため，共分散を考慮した距離計量を利用できる
  - 混合ガウス分布では局所特徴と多くのコードワードとの関係を表現できるので，特徴空間における局所特徴の位置に関する情報をエンコードできる

- デメリット
  - 混合ガウス分布表現はBoF と比較してパラメータが多い
    - 混合ガウス分布：O(K(D^2/2 + D))，BoF：O(KD)
  - 混合ガウス分布は訓練データに対して過剰適合する可能性があり，学習時に正則化を行う必要

フィッシャーベクトル

F. Perronnin and C. Dance. Fisher kernels on visual vocabularies for image categorization. CVPR, 2007.



- 混合ガウス分布を用いた確率密度分布推定によるBoF の改良
  - 生成モデル（generative model）
- 生成モデルを識別的なアプローチに適応可能なより洗練された手法があれば識別性能の改善につながる.
- フィッシャーカーネル（Fisher Kernel)
  - 生成的アプローチ（generative approach）と識別的アプローチ（discriminative approach）を結合させる強力な枠組み→確率分布の空間に適切な距離計量を埋め込む
  - 確率分布のなす空間は，Fisher 情報行列を計量とするリーマン空間
  - 手順
  1. 局所特徴を生成する確率密度分布から導出される勾配ベクトルの計算
  2. 画像を表現する一つの特徴ベクトルの計算
     →フィッシャーベクトル（Fisher Vector）
  3. 得られた特徴ベクトルを識別的分類機に入力する.

フィッシャーベクトルのメリット

- 豊かな特徴ベクトル表現
  - BoF と比較してフィッシャーカーネルを利用するメリットは，コードブックサイズが同じであればより要素数の多い特徴ベクトルが得られる.
    - コードブックサイズ：K，局所特徴の次元：d
    - BoFの次元：K
    - フィッシャーベクトル：　(2d+1)K-1
  - 特徴ベクトルの表現する情報が多いため計算コストの高いカーネル法を利用して高次元空間へ射影する必要がなく，線形識別機でも十分な識別性能を出すことが可能となる.

大規模データに
最も重要な要素

フィッシャーベクトル詳細

- 局所特徴群

$$\mathcal{X} = \{\boldsymbol{x}_n \in R^D\}_{n=1}^N$$

- あらゆる画像内容を表現する局所特徴の確率密度分布

$$u_\theta$$

- 対数尤度の勾配

$$G_\theta^{\mathcal{X}} = \frac{1}{N}\nabla_\theta \log u_\theta(\mathcal{X}|\boldsymbol{\theta})$$

  - データに最も適合するように確率密度関数のパラメータが修正すべき方向を表現
  - 異なるデータサイズ集合をパラメータ数に依存した特定の長さの特徴ベクトルに変換
  - 内積を利用する識別機には適切な計量が必要！！

- フィッシャー情報行列

$$F_\theta = E_X[\nabla_\theta \log u_\theta(\mathcal{X}|\boldsymbol{\theta})\nabla_\theta \log u_\theta(\mathcal{X}|\boldsymbol{\theta})^\top]$$

- フィッシャーベクトル（Fisher Vector）

$$\mathcal{G}_\theta^X = F_\theta^{-1/2}\nabla_\theta \log u_\theta(\mathcal{X}|\boldsymbol{\theta})$$

フィッシャー情報行列による対数尤度の勾配の正規化

混合ガウス分布におけるフィッシャーベクトル

- 確率密度分布を混合ガウス分布とする
  - 共分散行列は対角行列と仮定

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \sum_{k=1}^{K} \pi_k p_k(\boldsymbol{x})$$

- 対数尤度の微分

$$\mathcal{L}(\mathcal{X}|\boldsymbol{\theta}) = \log u_\theta(\mathcal{X}|\boldsymbol{\theta})$$

あらゆる画像を生成する確率密度分布

画像1枚から得られる局所特徴の集合

負担率：局所特徴x_nがGMMのコンポーネントkに属する確率

$$\frac{\partial \mathcal{L}(\mathcal{X}|\boldsymbol{\theta})}{\partial \pi_k} = \sum_{n=1}^{N} \left[\frac{\gamma_n(k)}{\pi_k} - \frac{\gamma_n(1)}{\pi_1} \right]$$

$$\frac{\partial \mathcal{L}(\mathcal{X}|\boldsymbol{\theta})}{\partial \boldsymbol{\mu}_k^d} = \sum_{n=1}^{N} \gamma_n(k) \left[\frac{\boldsymbol{x}_n^d - \boldsymbol{\mu}_k^d}{(\boldsymbol{\sigma}_k^d)^2} \right]$$

$$\frac{\partial \mathcal{L}(\mathcal{X}|\boldsymbol{\theta})}{\partial \boldsymbol{\sigma}_k^d} = \sum_{n=1}^{N} \gamma_n(k) \left[\frac{(\boldsymbol{x}_n^d - \boldsymbol{\mu}_k^d)^2}{(\boldsymbol{\sigma}_k^d)^3} - \frac{1}{\boldsymbol{\sigma}_k^d} \right]$$

GMMのBoFとほぼ同じ

$$\boldsymbol{f} = \frac{1}{N} \sum_{n=1}^{N} [\gamma_n(1), \cdots, \gamma_n(K)]^\top \in R^K$$

局所特徴x_nとGMMの各コンポーネントkの平均との差分

- 混合比：BoFとほぼ同じ
- 平均，分散：あらゆる画像を表現するpdfの平均との差分
- BoFは0次，Fisher Vectorは1次，2次の統計量を含む
- 分散の表現は平均の表現とあまり差がない？本来は各コンポーネント間の相関が必要

フィッシャー情報行列

- フィッシャー情報行列

$$F_\theta = E_X[\nabla_\theta \log u_\theta(\mathcal{X}|\boldsymbol{\theta}) \nabla_\theta \log u_\theta(\mathcal{X}|\boldsymbol{\theta})^\top]$$

- 混合ガウス分布において近似的に閉じた解が得られる
- 仮定
  - フィッシャー情報行列は対角行列
  - 共分散行列は対角行列
  - 負担率はピーキー
  - 一枚の画像から得られる局所特徴数は一定

$$\frac{\partial \mathcal{L}(\mathcal{X}|\theta)}{\partial \pi_k} \Rightarrow f_{\pi_k} = N\left(\frac{1}{\pi_k} + \frac{1}{\pi_1}\right)$$

$$\frac{\partial \mathcal{L}(\mathcal{X}|\theta)}{\partial \mu_k^d} \Rightarrow f_{\mu_k^d} = \frac{N\pi_k}{(\boldsymbol{\sigma}_k^d)^2}$$

$$\frac{\partial \mathcal{L}(\mathcal{X}|\theta)}{\partial \sigma_k^d} \Rightarrow f_{\sigma_k^d} = \frac{2N\pi_k}{(\boldsymbol{\sigma}_k^d)^2}$$

フィッシャー情報行列の要素

フィッシャーベクトルの性能

http://www.image-net.org/challenges/LSVRC/2010/ILSVRC2010_XRCE.pdf

- Pascal VOC 2007
- 改良されたフィッシャーベクトルを利用
- 識別機：線形SVM

PN	L2	SP	SIFT	Col	S+C
-	-	-	47.9	34.2	45.9
✓	-	-	54.2	45.9	57.6
-	✓	-	51.8	40.6	53.9
-	-	✓	50.3	37.5	49.0
✓	✓	✓	58.3	50.9	60.3

パワー正規化＞L2正規化＞空間ピラミッド，の順で改善の効果が高い

GMM Supervectors

- W. M. Campbell and D. E. Sturim and D. A. Reynolds. Support vector machines using GMM supervectors for speaker verification. IEEE Signal Processing Letters, Vol.13, pp.308-311, 2006.
- もともと音声認識で利用されていたもの.

$$p(x;\theta) = \sum_{k=1}^{K} w_k N(x;\mu_k,\Sigma_k)$$

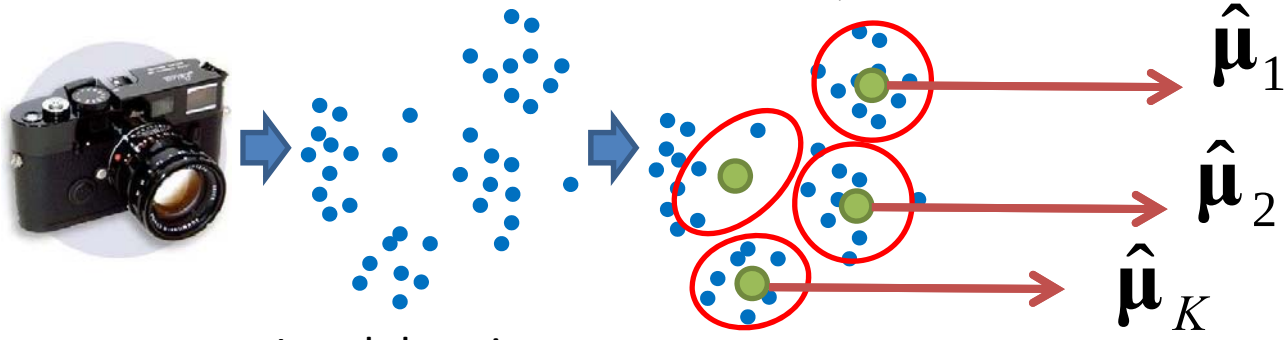


Image　　Local descriptors in feature space　　GMM　　Means of components

•訓練画像全体からユニバーサルな分布を作成
•各画像に適合させる

$$\hat{\mu}_k = \frac{\tau\hat{\mu}_k^{(U)} + \sum_{i=1}^{N}\gamma_i(k)x_i}{\tau + \sum_{i=1}^{N}\gamma_i(k)}$$

MAP適合

$$\tilde{\mu}_k = \sqrt{w_k^{(U)}}\,(\Sigma_k^{(U)})^{-1/2}\,\hat{\mu}_k$$

コンポーネントkの混合比　　コンポーネントkの共分散行列

$$\phi(X) = \begin{pmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \\ \vdots \\ \tilde{\mu}_K \end{pmatrix}$$

GMM supervectors

GMM Supervectors

- W. M. Campbell and D. E. Sturim and D. A. Reynolds. Support vector machines using GMM supervectors for speaker verification. IEEE Signal Processing Letters, Vol.13, pp.308-311, 2006.

GMM supervectors

$$\hat{\boldsymbol{\mu}}_k = \frac{\tau\hat{\mu}_k^{(U)} + \sum_{i=1}^{N}\gamma_i(k)x_i}{\tau + \sum_{i=1}^{N}\gamma_i(k)}$$

$\tau = 0$

$$\begin{aligned}\tilde{\boldsymbol{\mu}}_k &= \sqrt{w_k^{(U)}}(\Sigma_k^{(U)})^{-1/2}\hat{\boldsymbol{\mu}}_k \\ &\approx \frac{\sqrt{w_k^{(U)}}}{\sum_{i=1}^{N}\gamma_k(i)}(\Sigma_k^{(U)})^{-1/2}\sum_{i=1}^{N}\gamma_i(k)x_i \\ &\approx \frac{1}{N\sqrt{w_k^{(U)}}}(\Sigma_k^{(U)})^{-1/2}\sum_{i=1}^{N}\gamma_i(k)x_i\end{aligned}$$

$$Nw_k = \sum_{i=1}^{N}\gamma_i(k)$$

GMM supervectorとFisher Vectorの平均成分はほぼ同一

Fisher Vectorの平均成分

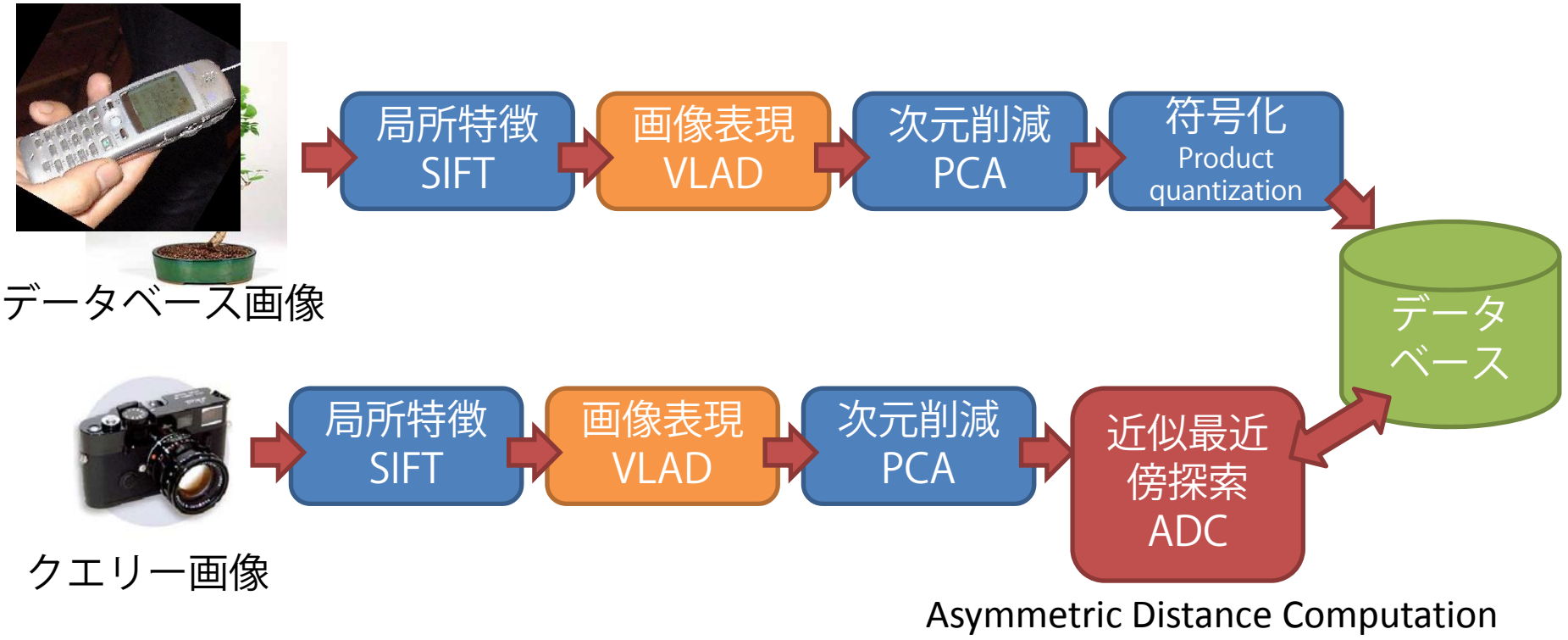
$$g_{\mu,i} = \frac{1}{N\sqrt{w_k}}\sum_{i=1}^{N}\gamma_i(k)(\Sigma_k)^{-1/2}(\mathbf{x}_i - \boldsymbol{\mu}_k)$$

TRECVID 2011ではGMM supervectorの局所特徴の各コンポーネントへの割り付けを高速化させることで第一位の性能を上げている.

N. Inoue and K. Shinoda. A Fast MAP Adaptation Technique for GMMsupervector-based Video Semantic Indexing. ACM Multimedia, 2011.

フィッシャーベクトルの画像検索への応用例

- H. Jegou, M. Douze, C. Schmid, and P. Perez. Aggregating local descriptors into a compact image representation. CVPR, 2010.
- 20bitに画像表現しても，生のBoFを使った検索と同じ検索性能
- パイプライン



VLAD

H. Jegou, M. Douze, C. Schmid, and P. Perez. Aggregating local descriptors into a compact image representation. CVPR, 2010.

- Vector of Locally Aggregated Descriptors

VLADのd番目要素

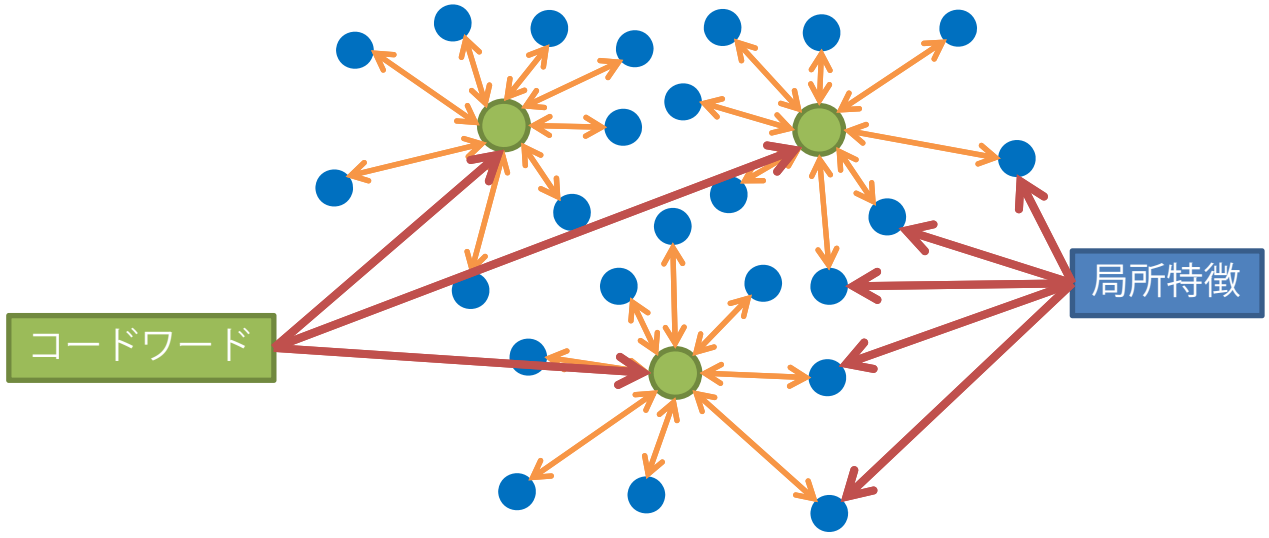
局所特徴のd番目要素

局所特徴が割り当てられたコードワードiのベクトル

$$z_i^d = \sum_{x \in \mathcal{X}_i} \left(x^d - v_i^d \right)$$

この後L2正規化

コードワードiに属する局所特徴集合



VLADとフィッシャーベクトル

- フィッシャーベクトル

$$\frac{\partial \mathcal{L}(\mathcal{X}|\boldsymbol{\theta})}{\partial \pi_k} = \sum_{n=1}^{N} \left[\frac{\gamma_n(k)}{\pi_k} - \frac{\gamma_n(1)}{\pi_1} \right]$$

$$\frac{\partial \mathcal{L}(\mathcal{X}|\boldsymbol{\theta})}{\partial \boldsymbol{\mu}_k^d} = \sum_{n=1}^{N} \gamma_n(k) \left[\frac{\boldsymbol{x}_n^d - \boldsymbol{\mu}_k^d}{(\boldsymbol{\sigma}_k^d)^2} \right]$$

$$\frac{\partial \mathcal{L}(\mathcal{X}|\boldsymbol{\theta})}{\partial \boldsymbol{\sigma}_k^d} = \sum_{n=1}^{N} \gamma_n(k) \left[\frac{(\boldsymbol{x}_n^d - \boldsymbol{\mu}_k^d)^2}{(\boldsymbol{\sigma}_k^d)^3} - \frac{1}{\boldsymbol{\sigma}_k^d} \right]$$

GMMのBoFとほぼ同じ

局所特徴x_nとGMMの各コンポーネントkの平均との差分

- VLAD

$$\boldsymbol{z}_i^d = \sum_{x \in \mathcal{X}_i} \left(\boldsymbol{x}^d - \boldsymbol{v}_i^d \right)$$

VLADのd番目要素

局所特徴のd番目要素

局所特徴が割り当てられたコードワードiのベクトル

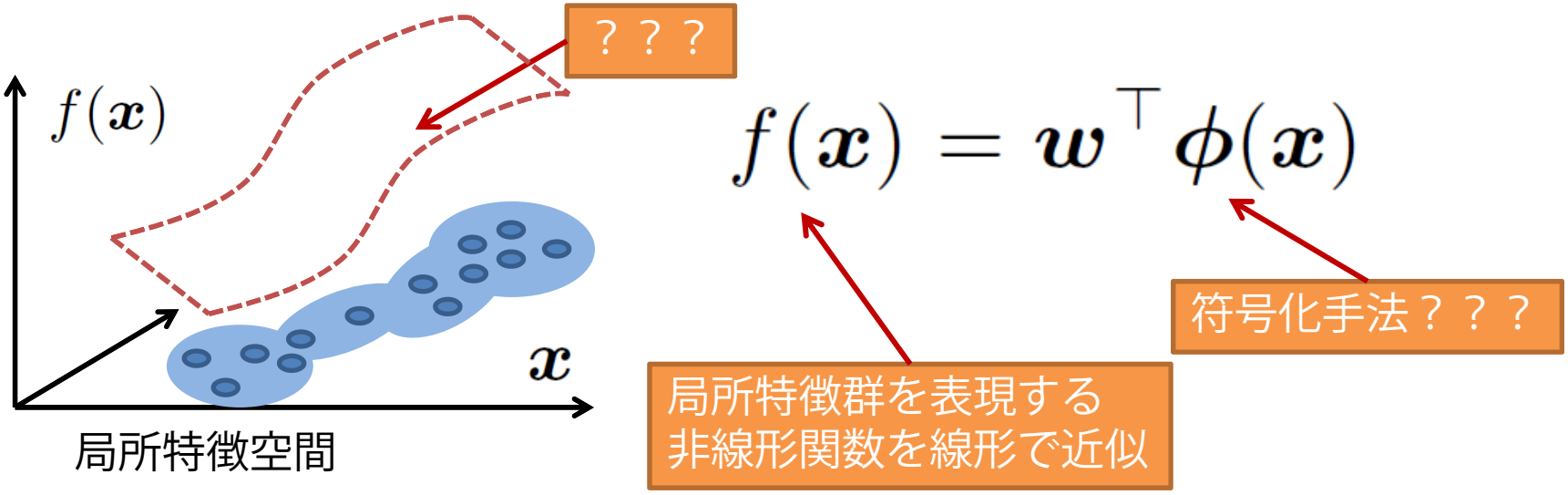
コードワードiに属する局所特徴集合

•負担率：ハードな割り当て
•分散：全てのコンポーネントで同じ

•→VLADはフィッシャーベクトルの平均に関する要素と同じ.
•（注）分散を考えていないのでフィッシャーとは言い難い.

スーパーベクトル符号化
Super-Vector Coding

- X. Zhou, K. Yu, T. Zhang, and T.S. Huang. Image classification using super-vector coding of local image descriptors. ECCV, 2010.
- BoF や混合ガウス分布を用いたBoF の改善手法
  - 特徴空間における局所特徴の分布の表現を得るプロセスと解釈できた.
- ここでも高次元空間における局所特徴分布を表現する，なめらかな非線形関数 *f(x)* の学習について考える.
- 非線形関数 *f(x)* を線形表現可能な符号化手法 *Φ(x)* を求める.



スーパーベクトル符号化の導出

- 局所特徴をコードブックを利用して近似

$$\boldsymbol{x} \approx \sum_{k=1}^{K} \gamma_x(k)\boldsymbol{v}_k \qquad \boldsymbol{\gamma}_x = [\gamma_x(1), \cdots, \gamma_x(K)], \sum_{k=1}^{K} \gamma_x(k) = 1$$

負担率のようなもの

コードワードk

- β Lipschitz derivative smooth

関数f(x)の1次近似のUpper boundに関する式

$$|f(\boldsymbol{x}) - f(\boldsymbol{x}') - \nabla f(\boldsymbol{x}')^\top (\boldsymbol{x} - \boldsymbol{x}')| \leqq \frac{\beta}{2}||\boldsymbol{x} - \boldsymbol{x}'||^2$$

コードワードの代入 $\boldsymbol{x}' = \boldsymbol{v}^x$

$$|f(\boldsymbol{x}) - f(\boldsymbol{v}^x) - \nabla f(\boldsymbol{v}^x)^\top (\boldsymbol{x} - \boldsymbol{v}^x)| \leqq \frac{\beta}{2}||\boldsymbol{x} - \boldsymbol{v}^x||^2$$

$$f(\boldsymbol{x}) = f(\boldsymbol{v}^x) + \nabla f(\boldsymbol{v}^x)^\top (\boldsymbol{x} - \boldsymbol{v}^x) \cdots (\star)$$

||x-v||が小さければ近似精度が向上

- スーパーベクトル符号化

$$f(\boldsymbol{x}) \approx \boldsymbol{w}^\top \boldsymbol{\phi}(\boldsymbol{x})$$

式(☆)を分解！

Super Vector Coding

$$\boldsymbol{\phi}(\boldsymbol{x}) = \left[s\gamma_x(k), \gamma_x(k)(\boldsymbol{x} - \boldsymbol{v}_k)^\top \right]^\top_{v_k \in \mathcal{V}}$$

$$\boldsymbol{w} = \left[\frac{1}{s} f(\boldsymbol{v}_k), (\nabla f(\boldsymbol{v}_k))^\top \right]^\top_{v_k \in \mathcal{V}}$$

スーパーベクトル符号化の解釈

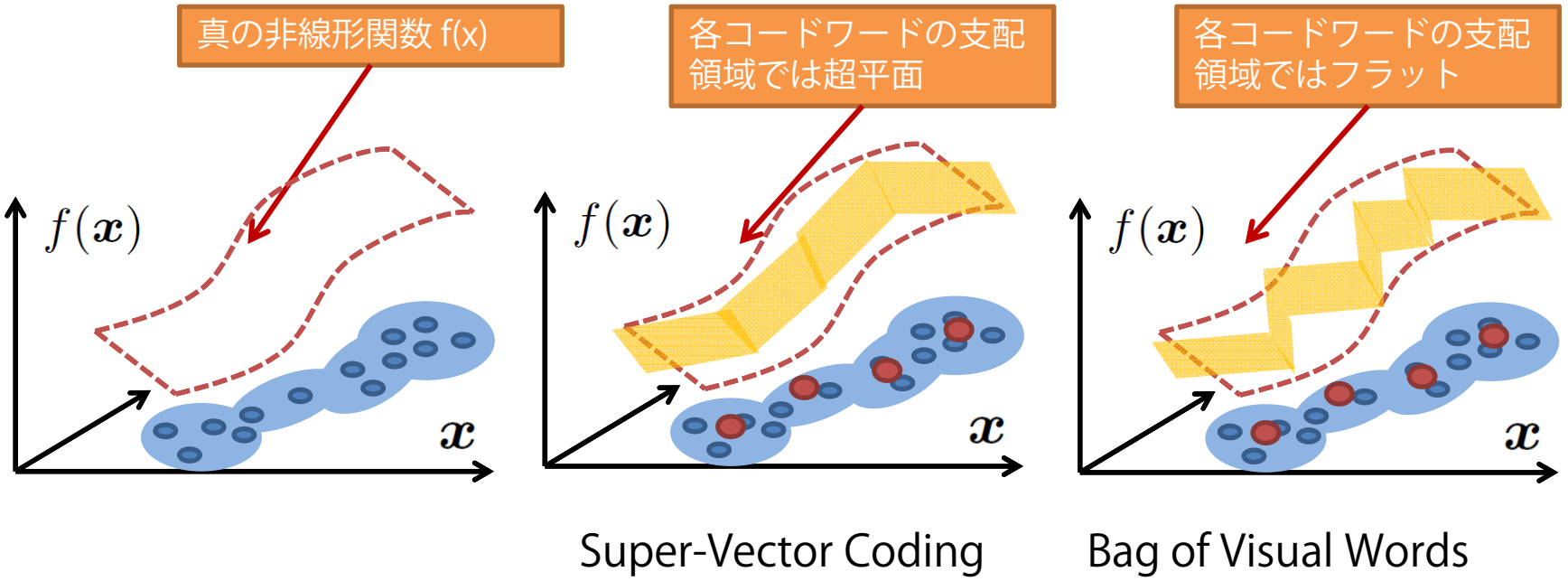
- スーパーベクトル符号化の例
  - コードワード数：3，　γ =[0 1 0]'

Super Vector Coding

$$\phi(\boldsymbol{x}) = \left[s\gamma_x(k), \gamma_x(k)(\boldsymbol{x} - \boldsymbol{v}_k)^\top \right]^\top_{v_k \in \mathcal{V}}$$

$$\Rightarrow \phi(x) = \left[\underbrace{0,\cdots,0}_{d+1\text{dim}}, \underbrace{s,(x-v)^T}_{d+1\text{dim}}, \underbrace{0,\cdots,0}_{d+1\text{dim}} \right]^T$$

- スーパーベクトル符号化とBoF



スーパーベクトル符号化とフィッシャーベクトル

- フィッシャーベクトル

$$\frac{\partial \mathcal{L}(\mathcal{X}|\boldsymbol{\theta})}{\partial \pi_k} = \sum_{n=1}^{N} \left[\frac{\gamma_n(k)}{\pi_k} - \frac{\gamma_n(1)}{\pi_1} \right]$$

$$\frac{\partial \mathcal{L}(\mathcal{X}|\boldsymbol{\theta})}{\partial \boldsymbol{\mu}_k^d} = \sum_{n=1}^{N} \gamma_n(k) \left[\frac{\boldsymbol{x}_n^d - \boldsymbol{\mu}_k^d}{(\boldsymbol{\sigma}_k^d)^2} \right]$$

$$\frac{\partial \mathcal{L}(\mathcal{X}|\boldsymbol{\theta})}{\partial \boldsymbol{\sigma}_k^d} = \sum_{n=1}^{N} \gamma_n(k) \left[\frac{(\boldsymbol{x}_n^d - \boldsymbol{\mu}_k^d)^2}{(\boldsymbol{\sigma}_k^d)^3} - \frac{1}{\boldsymbol{\sigma}_k^d} \right]$$

GMMのBoFとほぼ同じ

局所特徴x_nとGMMの各コンポーネントkの平均との差分

- スーパーベクトル符号化

$$\boldsymbol{\phi}(\boldsymbol{x}) = \left[s\gamma_x(k), \gamma_x(k)(\boldsymbol{x} - \boldsymbol{v}_k)^\top \right]^\top_{v_k \in \mathcal{V}}$$

負担率

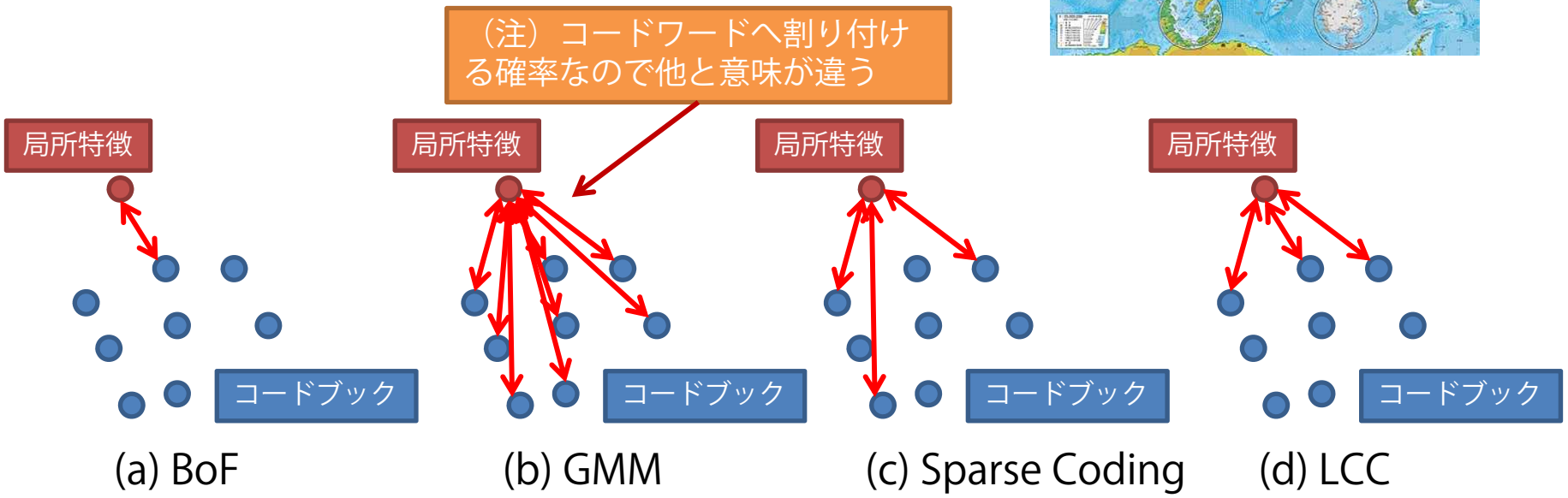
局所特徴x_nとコードワードとの差分

- 混合比：一定
- 分散：一定

- →スーパーベクトル符号化はフィッシャーベクトルの混合比と平均に関する要素と同じ
- （注）分散を考えていないのでフィッシャーとは言い難い.

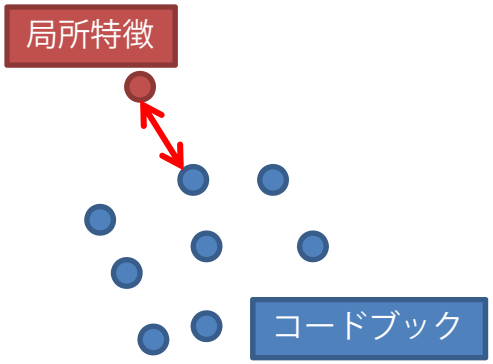
スパース符号化の比較

- BoF
  - 局所特徴が一つのコードワードに割り当てられる
- BoFのGMMによる表現
  - 局所特徴が全てのコードワードと関係を持つ
- スパース符号化
  - 局所特徴が少数のコードワードと関係を持つ
- 局所線形制約符号化
  - 局所特徴が局所の少数コードワードと関係を持つ



スパース符号化の定式化

- Bag of Visual Words
  - ベクトル量子化（VQ）



(a) BoF

$$\min_{U,V} \sum_{n=1}^{N} ||\boldsymbol{x}_n - \boldsymbol{V}\boldsymbol{u}_n||^2$$

$$\text{s.t. } \text{Card}(\boldsymbol{u}_n) = 1, |\boldsymbol{u}_n| = 1, \boldsymbol{u}_n \succeq 0, \forall n$$

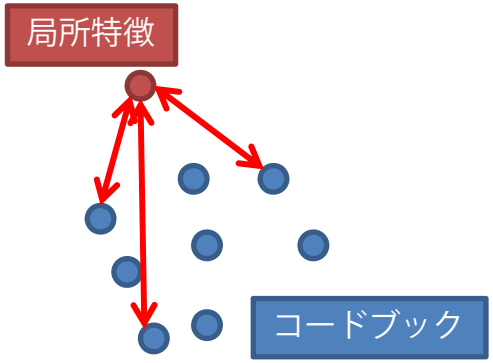
局所特徴

コードブック

局所特徴がどのコードワードに所属するかを示す指標

一つのコードワードに属する制約→厳しすぎる！！！

- スパース符号化（Sparse Coding）



(c) Sparse Coding

$$\min_{U,V} \sum_{n=1}^{N} ||\boldsymbol{x}_n - \boldsymbol{V}\boldsymbol{u}_n||^2 + \lambda|\boldsymbol{u}_n|$$

$$\text{s.t. } ||\boldsymbol{v}_k|| \leqq 1, \forall k$$

L1ノルム正則化項
→少数のコードワードへの所属を許容

L1正則化の役割

- コードブックは局所特徴の次元数よりも多く，過剰（K > D）なため，under determinedな系である．つまり情報が不足して解を定められない状況にある．そのためL1 正則化により解を定めることが可能となる．

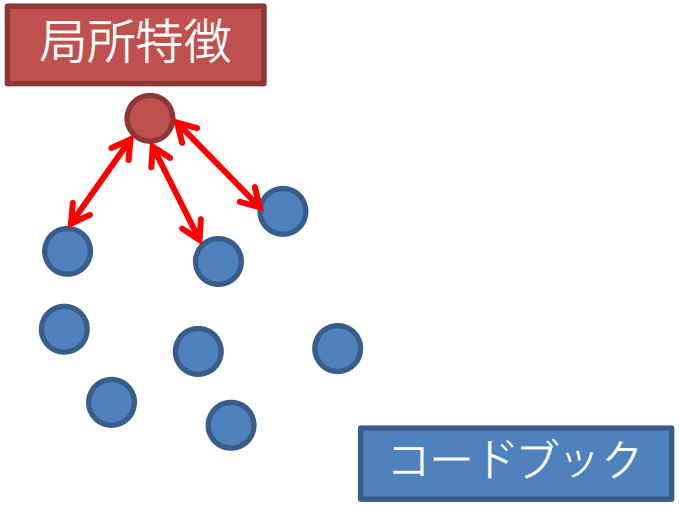
- *スパース性の事前知識を用いることによって*局所特徴の顕著なパターンを捉えることができる．

- *ベクトル量子化よりもスパース符号化の方が*量子化誤差を低減させられる．

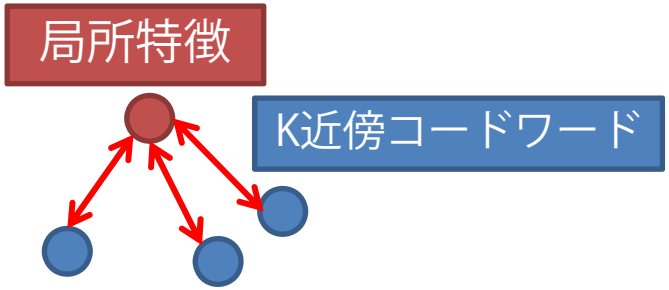
局所制約線形符号化

- 局所制約線形符号化
  - Locality-constrained Linear Coding (LLC)
  - J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. CVPR, 2010.
  - 局所座標符号化Local Coordinate Coding (LCC)の高速な実装
  - K. Yu, T. Zhang, and Y. Gong. Nonlinear learning using local coordinate coding. NIPS, 2009.

1) 局所特徴のK近傍のコードワードを探索



2) 局所特徴をK近傍コードワードを用いて再構築



$$\min_{\tilde{U}} \sum_{n=1}^{N} ||\boldsymbol{x}_n - \boldsymbol{V}_n \tilde{\boldsymbol{u}}_n||^2$$

$$\text{s.t. } \mathbf{1}^\top \tilde{\boldsymbol{u}}_n = 1, \forall n$$

局所線形埋込み（Local Linear Embedding, LLE）と比較して，局所制約線形符号化はコードブックの学習が入る点で異なる．

スパース符号化空間ピラミッド

- 空間ピラミッド
  - 符号化された局所特徴群 $\boldsymbol{U}$ から一つの特徴ベクトル $\boldsymbol{f}$ を得る手段

- プーリング（pooling）

$$\boldsymbol{f} = \mathcal{F}(\boldsymbol{U})$$

局所特徴集合

プーリング関数

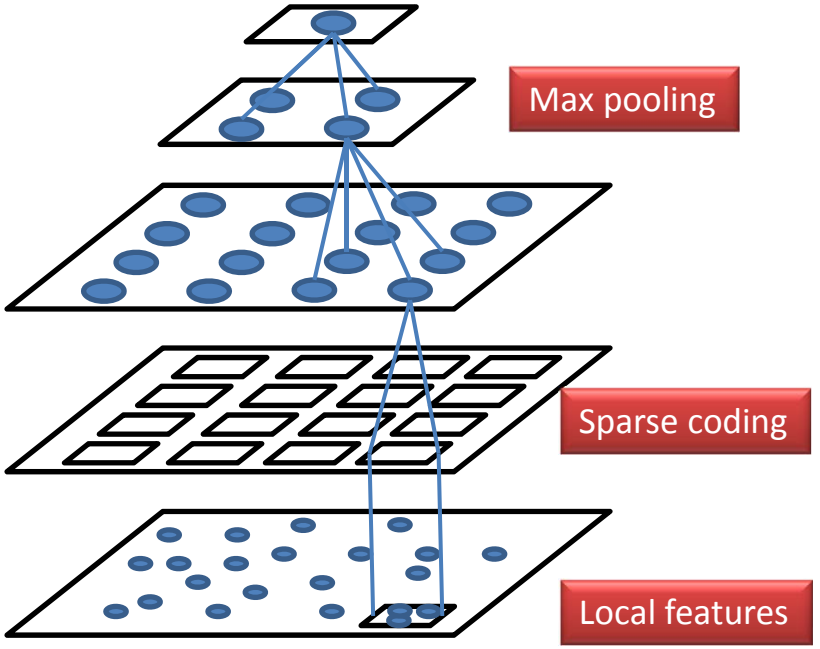
  - 平均プーリング
    average pooling

$$\boldsymbol{f} = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{u}_n$$

BoFはこれを利用

  - 最大値プーリング
    max pooling

$$\boldsymbol{f}^d = \max\{|\boldsymbol{u}_1^d|, |\boldsymbol{u}_2^d|, \cdots, |\boldsymbol{u}_N^d|\}$$



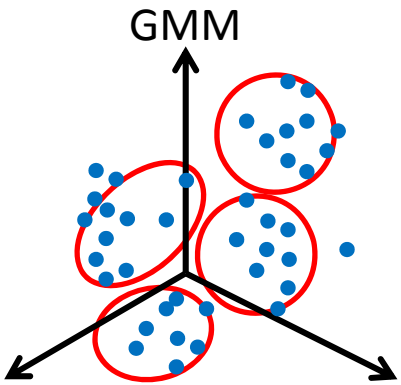
J. Yang, K. Yu, Y. Gong, and T. Huang. Linear spatial pyramid matching using sparse coding for image classification. CVPR, 2009.

画像表現
大域特徴

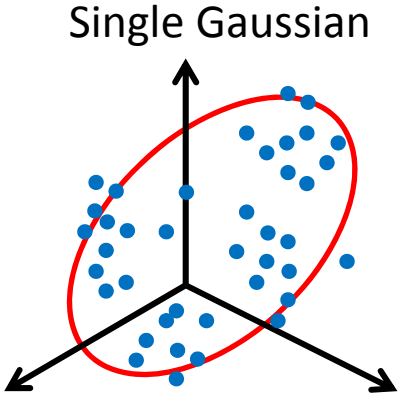
Generalized Local Correlation (GLC)

H. Nakayama, T. Harada, and Y. Kuniyoshi. Dense Sampling Low-Level Statistics of Local Features. In CIVR, 2009.

- GMM
  - 表現能力が高い
  - GMMはパラメータが多いので，共分散行列の非対角成分を0とする場合が多い．
  - 計算コストが高い



- Single Gaussian
  - 表現能力に限界有り
  - パラメータが少ないので，共分散行列推定可能
  - 共分散行列の非対角成分を有効活用
  - 計算コストが低い



局所記述子

平均 $\boldsymbol{\mu}^{(j)} = \frac{1}{p^{(j)}} \sum_k^{p^{(j)}} \boldsymbol{v}_k^{(j)}$

自己相関行列 $R^{(j)} = \frac{1}{p^{(j)}} \sum_k^{p^{(j)}} \boldsymbol{v}_k^{(j)} \boldsymbol{v}_k^{(j)T}$

GLC

$$\boldsymbol{x}^{(j)} = \begin{pmatrix} \boldsymbol{\mu}^{(j)} \\ upper(R^{(j)}) \end{pmatrix}$$

Generalized Local Correlation (GLC)

H. Nakayama, T. Harada, and Y. Kuniyoshi. Dense Sampling Low-Level Statistics of Local Features. In CIVR, 2009.

- GLCは単純であるが結構いける

Table 2: Comparison of the performance in two scene datasets and Caltech-101 (%). (*)approximate value read from the graph.

Dataset	GLC + PLDA			Previous	
	L1	L2	L3	no SI	with SI
OT8	88.8	90.5	91.1	82.3 [19]	90.2 [19]
				82.5 [3]	87.8 [3]
LSP15	80.0	83.2	84.1	72.7 [3]	83.7 [3]
				74.8 [11]	81.4 [11]
Caltech-101	55.0	63.3	64.8		72.0* [1]
					67.7 [3]
					66.2 [20]
				41.2 [11]	64.6 [11]
				58.2 [7]	
				39.6 [8]	

SPM+SVM

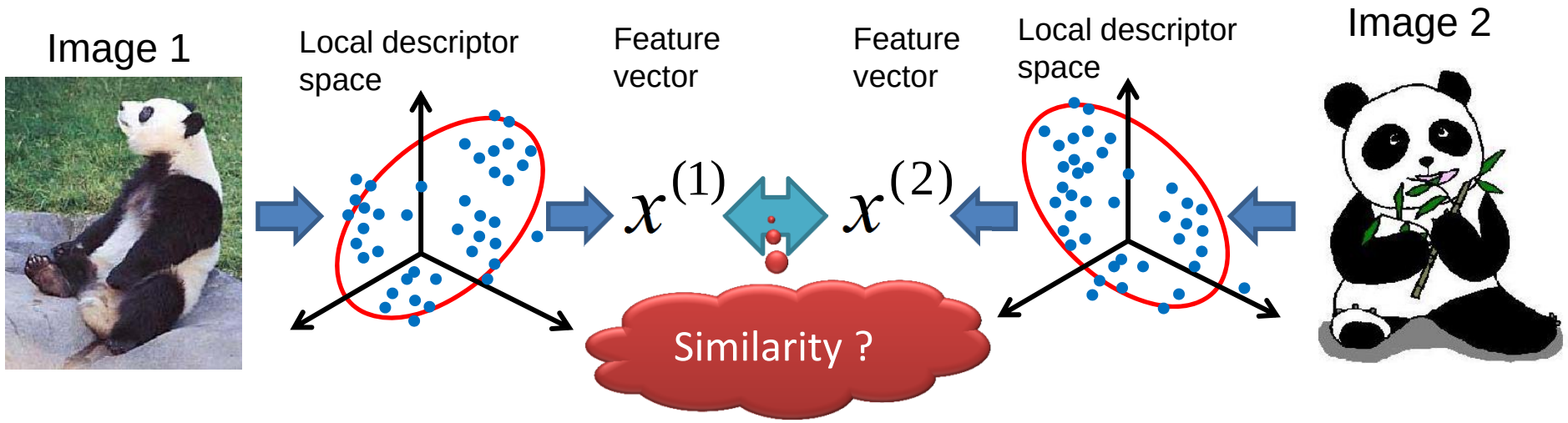
Figure 1: Sample images from the OT8 dataset.

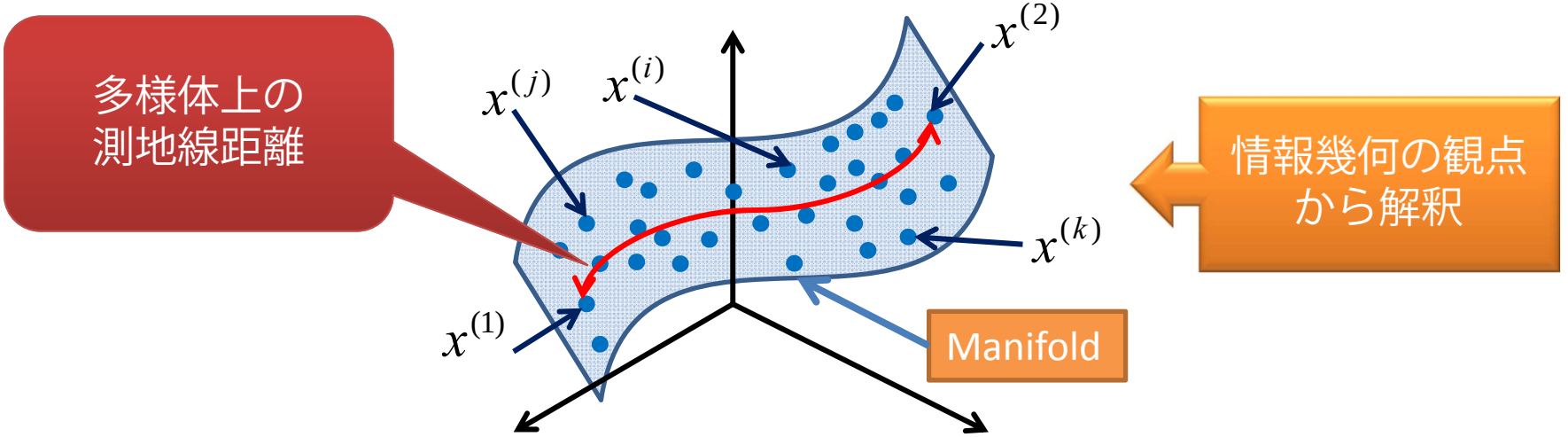
Figure 2: Additional seven classes in the LSP15 dataset.

Global Gaussian (GG)

H. Nakayama, T. Harada, and Y. Kuniyoshi. Global Gaussian Approach for Scene Categorization Using Information Geometry. In CVPR, 2010.

- 平均と分散を並べたGLCの表現は適切か？
- GLC間の距離計量は適切か？





Global Gaussian

H. Nakayama, T. Harada, and Y. Kuniyoshi. Global Gaussian Approach for Scene Categorization Using Information Geometry. In CVPR, 2010.

- 確率密度分布間の距離計量を正しく設定
- 情報幾何の考え方からGLCが自然に出てくる

η座標系におけるsingle Gaussianの表現

$$\begin{aligned}\boldsymbol{\eta} &= \sum_{1\le i\le d}\eta_i \boldsymbol{e}_i + \sum_{1\le i\le j\le d}\eta_{ij}\boldsymbol{e}_{ij} \\ &= (\eta_1,...,\eta_d,\eta_{11},...,\eta_{1d},\eta_{22},...\eta_{2d},...,\eta_{dd})^T \\ &= \big(\hat{\mu}_1,...,\hat{\mu}_d,\hat{\Sigma}_{11}+\hat{\mu}_1^2,...,\hat{\Sigma}_{1d}+\hat{\mu}_1\hat{\mu}_d, \\ &\qquad \hat{\Sigma}_{22}+\hat{\mu}_2^2,...,\hat{\Sigma}_{dd}+\hat{\mu}_d^2\big)^T.\end{aligned}$$

GLC

$$\boldsymbol{x}^{(j)} = \begin{pmatrix} \boldsymbol{\mu}^{(j)} \\ upper(R^{(j)}) \end{pmatrix}$$

GLCの厳密な距離

$$\begin{aligned}dist\big(\boldsymbol{\eta}(P),\boldsymbol{\eta}(Q)\big) &= tr(\Sigma_P\Sigma_Q^{-1}) + tr(\Sigma_Q\Sigma_P^{-1}) - 2d + \\ &\quad tr\left((\Sigma_P^{-1}+\Sigma_Q^{-1})(\boldsymbol{\mu}_P-\boldsymbol{\mu}_Q)(\boldsymbol{\mu}_P-\boldsymbol{\mu}_Q)^T\right)\end{aligned}$$

$$K_{kl}(P,Q) = \exp\big(-a\, dist\big(\boldsymbol{\eta}(P),\boldsymbol{\eta}(Q)\big)\big)$$

Gauss分布間のsymmetric KL-divergence

GLCの近似的な距離

Fisher Information Matrix

$$K_{ct}(P,Q) = \boldsymbol{\eta}(P)^T G^{\eta}(\boldsymbol{\eta}_c)\boldsymbol{\eta}(Q) \quad \Rightarrow \quad \boldsymbol{\zeta} = \big(G^{\eta}(\boldsymbol{\eta}_c)\big)^{1/2}\boldsymbol{\eta}$$

Linear-SVMにそのまま利用可

Global Gaussian (GG)

H. Nakayama, T. Harada, and Y. Kuniyoshi. Global Gaussian Approach for Scene Categorization Using Information Geometry. In CVPR, 2010.

- 性能評価

Table 5. Performances of global Gaussian, BoK, and combined approach (%). $L = 2$ spatial pyramid is implemented. Kernel PDA is used for classification. SURF descriptor is used for LSP15 and SIFT descriptor is used for 8-sports.

	LSP15	8-sports
GG (KL)	86.1±0.5	84.4±1.4
GG (ct-linear)	82.3±0.4	82.9±1.0
BoK200	81.1±0.7	79.6±1.1
BoK1000	82.5±0.7	81.5±1.7
GG (ct-linear) + BoK200	85.0±0.5	83.2±0.9
GG (ct-linear) + BoK1000	85.3±0.5	83.4±0.7

提案手法のスコア

Table 6. Performance comparison with previous work (%). For our method, $L = 2$ spatial pyramid is implemented, and kernel PDA is used for classification. We use the SURF descriptor for LSP15 and Indoor67, and the SIFT descriptor for 8-sports.

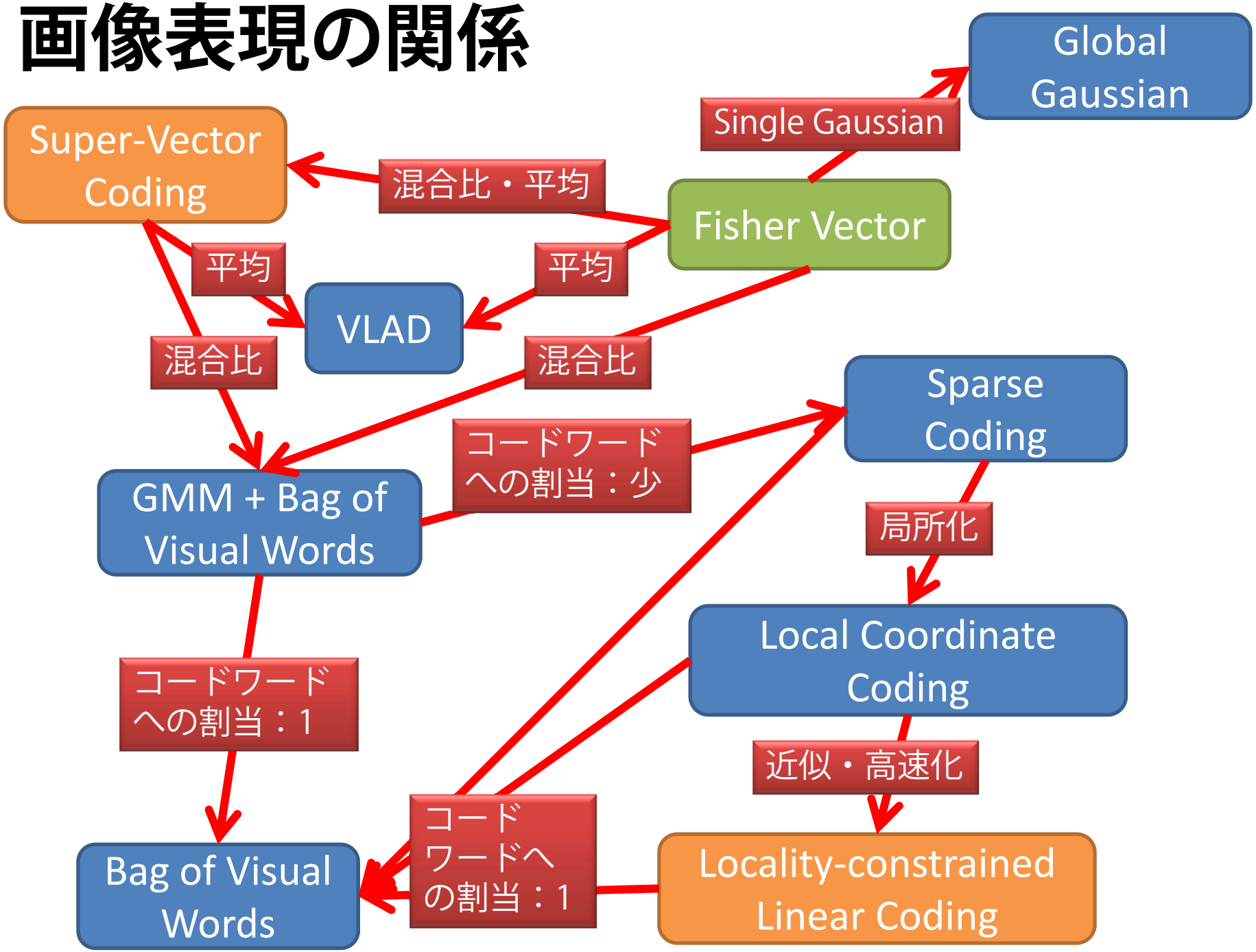
Method	LSP15	8-sports	Indoor67
GG (KL-div.)	86.1±0.5	84.4±1.4	45.5±1.1
GG (ct-linear) + BoK1000	85.3±0.5	83.4±0.7	44.9±1.3
Previous	85.2 [30]	84.2 [29]	25.0 [23]
	84.1 [29]	73.4 [14]	
	83.7 [6]		

従来手法のスコア

局所特徴の空間的関係性考えなくていいの？
→MIRU2012にて！

いずれのデータセットにおいても従来手法を上回る性能
平均，共分散といったパラメータのみで計算可能

画像表現の関係


Global Gaussian
Single Gaussian
Super-Vector Coding
混合比・平均
Fisher Vector
平均
平均
VLAD
混合比
混合比
Sparse Coding
コードワードへの割当：少
GMM + Bag of Visual Words
局所化
Local Coordinate Coding
コードワードへの割当：1
近似・高速化
コードワードへの割当：1
Bag of Visual Words
Locality-constrained Linear Coding

画像表現の性能比較

- K. Chatfield, V. Lempitsky, A. Vedaldi and A. Zisserman. The devil is in the details: an evaluation of recent feature encoding methods. BMVC, 2011.

Method				mAP								
(a) FK	Lin	ss3	256	**61.69**	78.97	67.43	51.94	70.92	30.79	72.18	79.94	61.35
(b) SV	Lin	ss3	1024	**58.16**	74.32	63.79	47.02	69.44	29.06	66.46	77.31	60.18
(c) LLC	Lin	ss2	25k	**57.60**	71.05	62.85	47.40	67.67	25.21	62.70	77.02	59.59
(d) LLC-F	Lin	ss2	25k	**59.32**	74.10	64.92	51.48	68.33	27.18	62.89	78.44	61.39
(e) VQ	Chi	ss2	25k	**56.07**	70.00	58.90	42.86	66.75	26.59	62.27	75.67	57.09
(f) LLC	Lin	ss3	25k	**57.27**	71.35	62.65	46.12	68.98	26.04	63.92	76.98	59.71
(g) LLC	Sqr	ss3	25k	**56.71**	71.24	61.75	42.73	68.21	25.85	62.33	76.40	59.31
(h) LLC	Chi	ss3	25k	**57.66**	72.41	62.19	47.30	68.91	25.78	63.95	77.27	59.83
(i) LLC-F	Lin	ss3	25k	**59.74**	74.17	65.39	51.15	69.69	28.67	64.40	78.48	63.00
(j) VQ	Chi	ss3	25k	**55.30**	70.10	59.24	44.14	66.34	26.79	60.88	75.62	55.42
(k) KCB	Chi	ss3	25k	**56.26**	70.83	60.60	44.50	66.52	27.02	62.07	76.29	57.61
(l) LLC	Lin	ss5	25k	**56.96**	69.82	61.63	46.71	68.27	25.66	63.78	76.32	59.83
(m) LLC-F	Lin	ss5	25k	**58.70**	73.44	62.90	50.22	67.90	27.85	64.35	77.91	62.44
(n) VQ	Chi	ss5	25k	**53.87**	68.74	57.14	41.24	64.54	25.20	61.12	74.06	53.22
(o) LLC	Lin	ss3	14k	**56.18**	70.71	59.67	44.81	67.20	26.03	60.99	76.25	58.54
(p) VQ	Chi	ss3	14k	**54.82**	69.09	58.61	41.27	66.30	26.49	61.46	75.42	55.77
(q) LLC	Lin	ss3	10k	**56.01**	69.66	60.44	44.21	67.78	24.66	61.84	75.42	57.70
(r) VQ	Chi	ss3	10k	**54.98**	69.56	57.97	42.86	65.84	23.52	61.06	75.89	55.55
(s) LLC	Lin	ss3	4k	**53.79**	69.83	57.63	42.04	66.46	22.44	55.62	72.77	56.98
(t) LLC	Sqr	ss3	4k	**52.07**	68.52	54.62	40.14	65.34	21.53	51.89	71.54	55.19
(u) LLC	Chi	ss3	4k	**53.47**	70.17	56.20	42.73	65.27	22.23	55.18	72.78	56.95
(v) LLC-1	Lin	ss3	4k	**36.06**	53.39	43.20	22.47	46.32	11.40	29.48	64.66	45.41
(w) LLC-F	Lin	ss3	4k	**55.87**	72.27	61.41	44.08	67.85	24.97	57.92	75.40	59.44
(x) VQ	Sqr	ss3	4k	**51.97**	67.29	55.22	36.58	64.42	21.89	56.31	72.90	52.11
(y) VQ	Chi	ss3	4k	**53.42**	68.65	57.04	39.86	64.59	21.96	58.79	73.89	53.77
(z) VQ	Lin	ss3	4k	**46.54**	60.63	48.80	32.76	58.54	16.26	50.44	68.42	45.97
(α) KCB	Chi	ss3	4k	**54.60**	69.82	59.20	41.97	64.85	23.90	59.02	74.98	54.63

Fisher Vectorは結構いい.

【データセット】
- Pascal VOC 2007

【比較した画像表現】
- VQ: Bag of Words
- FK: Fisher Vector
- SV: Super Vector
- LLC: Locality-constrained Linear Coding
- KCB: Kernel Codebook

FVいいけど計算大変，，，
→GGいいよ！MIRU2012

効率的な識別機

膨大なクラス識別における識別機

- 特徴ベクトルの次元を高次元に保つ

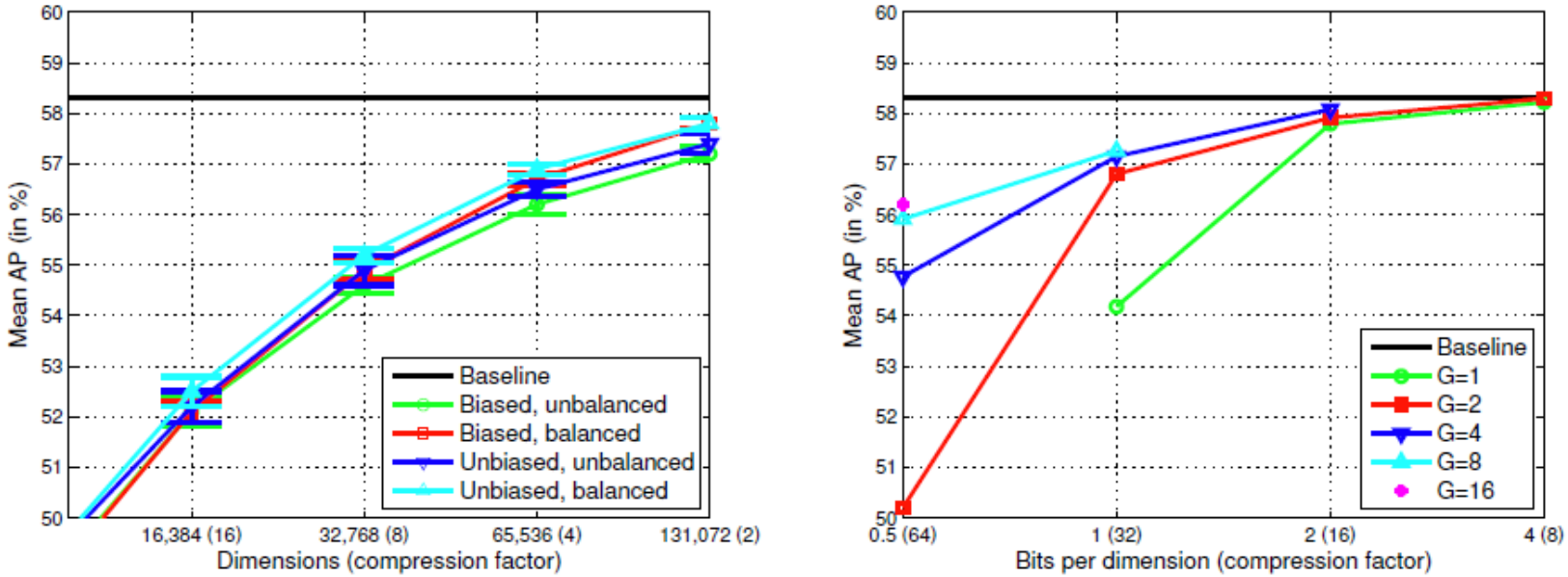


Figure 2. Compression results on VOC 2007. Left: HK results as a function of the number of dimensions. Right: PQ results as a function of the number b of bits per dimension and the group size G (without sparsity encoding). The baseline corresponds to the uncompressed signature (262,144 dimensions). For a given compression factor, PQ performs much better than HK.

J. Sanchez, and F. Perronnin. High-Dimensional Signature Compression for Large-Scale Image Classification. In CVPR, 2011.

高次元の特徴であっても破綻しない識別機が必要→SVMの利用が多い

SVMの逐次学習

本当に出力結果の大小関係いらないの？
→大幅に性能改善！MIRU2012

- 多クラスSVM
  - One-vs-all SVMがほとんど
  - 各クラスのSVM出力値の大小関係が適切である保証はない．しかしながら十分高い識別率が得られることが知られている．
  - 各クラス毎に独立に学習，識別ができる→容易に並列化でき，大規模データに有効

- SVMのオンライン学習
  - 訓練サンプルを1個メモリにロードする．
  - 現状の識別機で識別し，誤っていたら識別機を更新する．
  - →訓練サンプルを全てメモリに展開する必要がないの，大規模データに適する．

- 確率的勾配降下法（Stochastic Gradient Decent）によるSVM

評価関数

$$L = \sum_{t=1}^{T} L(\boldsymbol{w}, b, \boldsymbol{x}_t, y_t) = \sum_{t=1}^{T} \frac{\lambda}{2} ||\boldsymbol{w}||^2 + \max\left[0, 1 - y_t(\boldsymbol{w}^\top \boldsymbol{x}_t + b)\right],$$

パラメータの更新

$$\boldsymbol{w}^t = \boldsymbol{w}^{t-1} - \eta \nabla_w L(\boldsymbol{w}, b, \boldsymbol{x}_t, y_t), \quad b^t = b^{t-1} - \eta \nabla_b L(\boldsymbol{w}, b, \boldsymbol{x}_t, y_t).$$

平均化：高速化

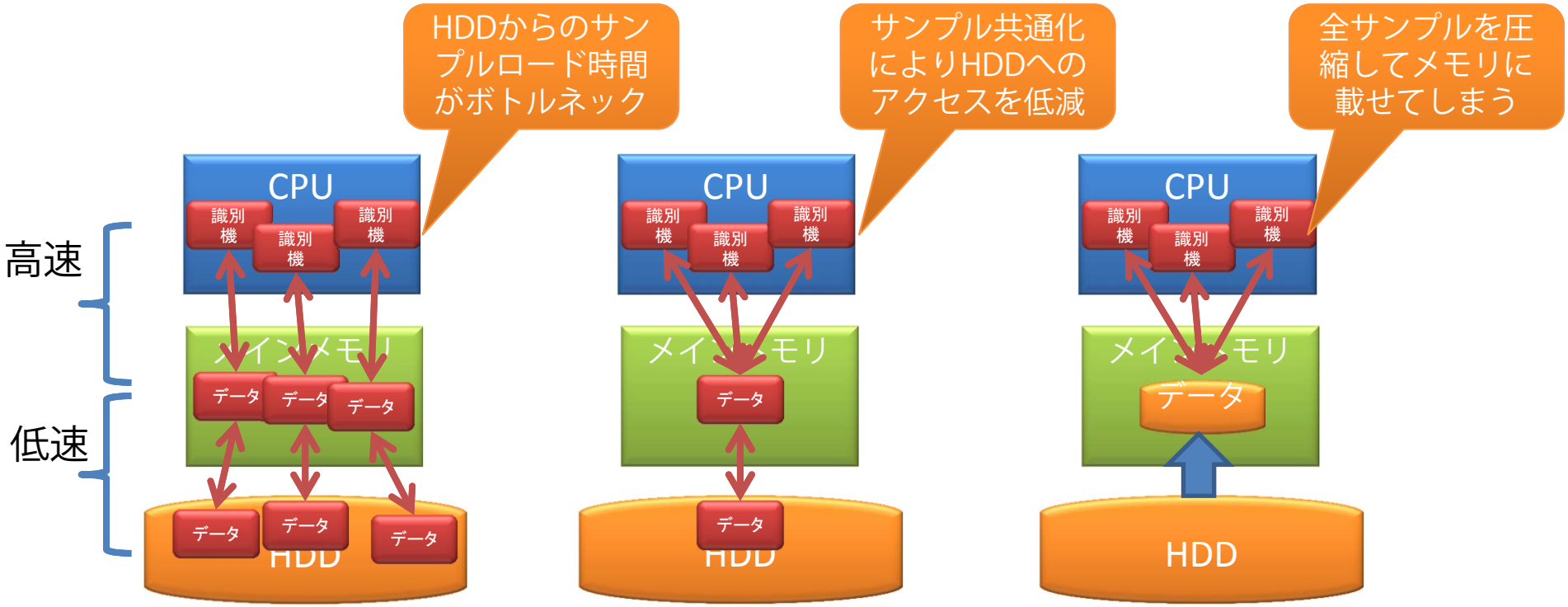
$$\bar{\boldsymbol{w}}^t = (1 - 1/t)\bar{\boldsymbol{w}}^{t-1} + \boldsymbol{w}^t/t, \quad \bar{b}^t = (1 - 1/t)\bar{b}^{t-1} + b^t/t.$$

I/Oのボトルネック対策

- 計算速度 （高速）>> HDDのアクセス速度（低速）
  - ボトルネックは補助記憶装置への頻繁なアクセス
  - データが膨大なので全てのデータをロードするだけで一苦労

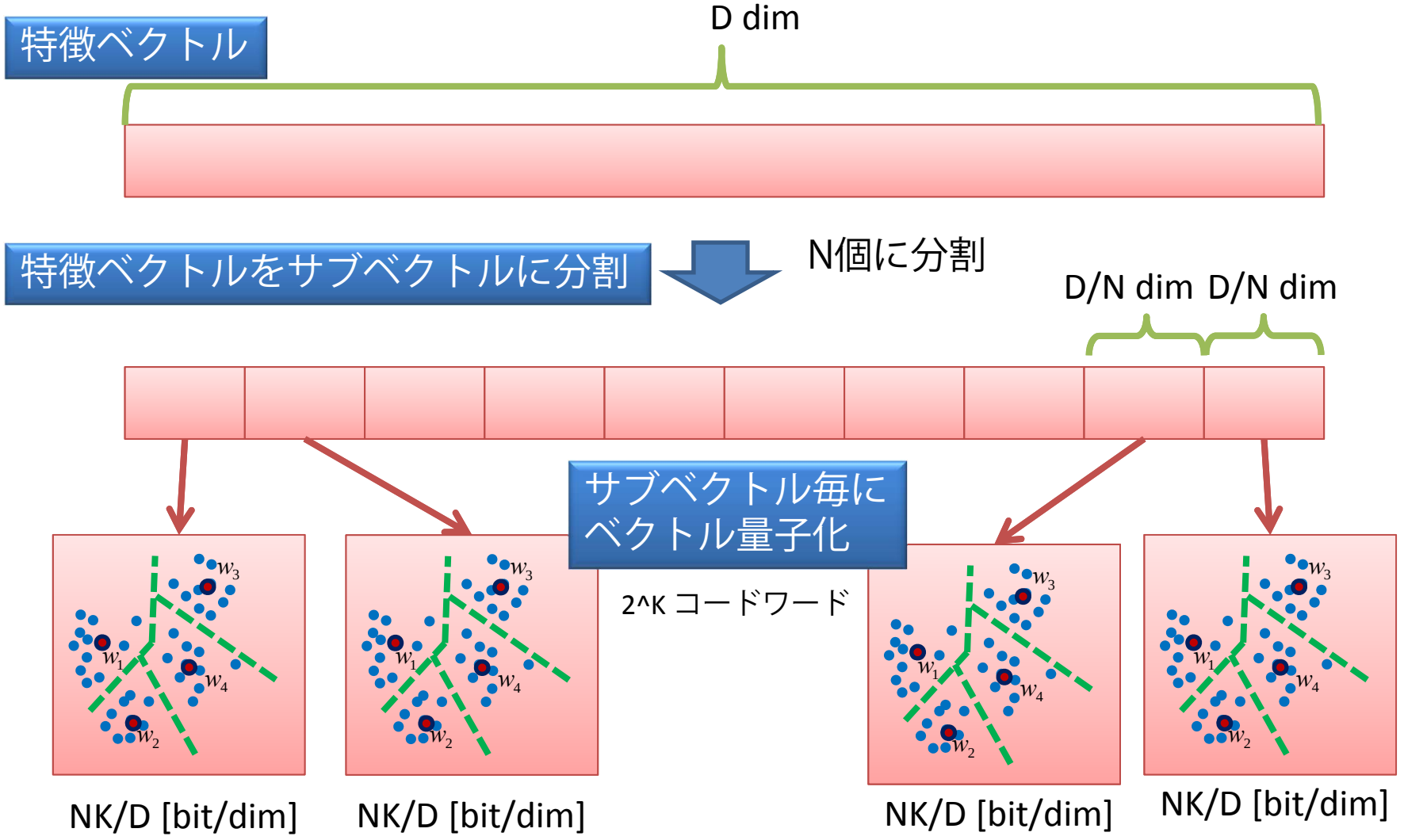
- NEC system
  - Hadoopによる並列化
  - One-vs-all SVMの識別機を並列に学習する際に，学習サンプルを共通化.

- XRCE system
  - 学習データを圧縮し，全ての学習データをメインメモリに押し込む.

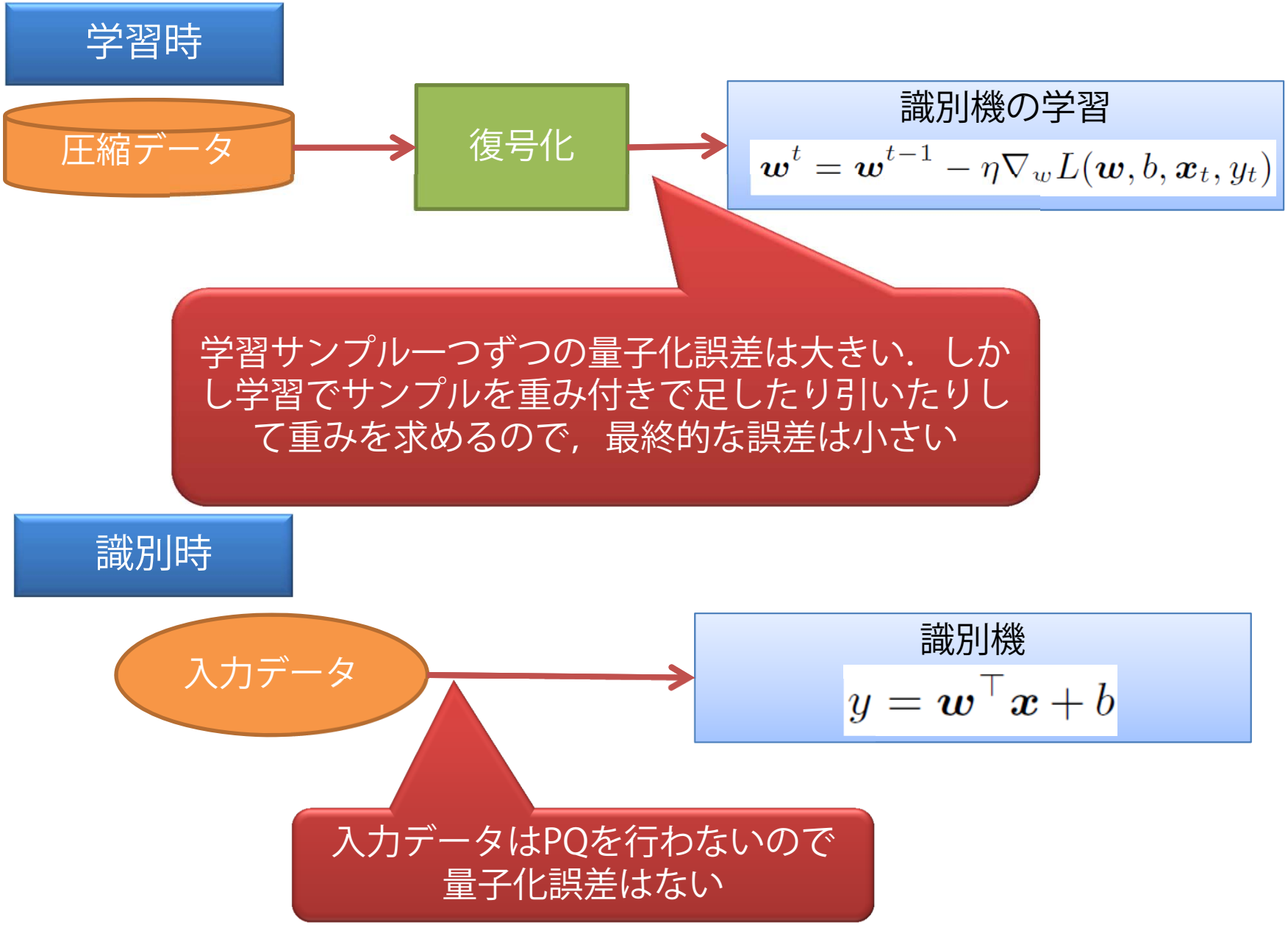


プロダクト量子化

- H. Jegou, M. Douze, and C. Schmid. Product Quantization for Nearest Neighbor Search. IEEE Trans. on PAMI, Vol.33, pp.117-128, 2011.

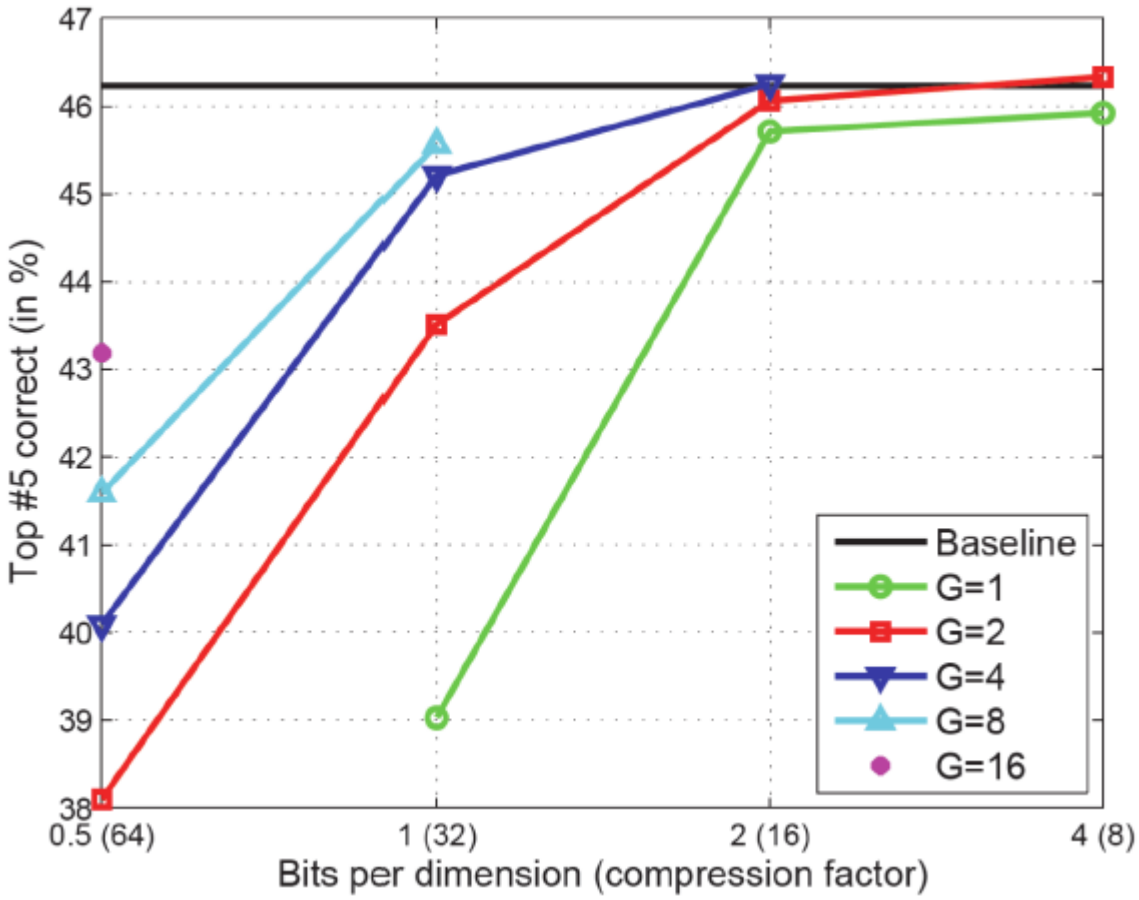


SGD-SVM + プロダクト量子化



PQ + SGD-SVMの実験結果

- 1次元1bitにしても性能低下は少ない



http://www.image-net.org/challenges/LSVRC/2011/ilsvrc11.pdf

まとめ

- 大規模画像データを用いた一般画像認識に関して概観した.
- 近年，大規模画像識別に用いられている画像表現を紹介し，それらの体系化の試みを解説した.
- 画像認識性能向上には，データ，特徴抽出，モデルの順に高い質が必要であることを述べた.
- データの大規模化，リッチかつ高次元の画像特徴，モデルのスケーラビリティの重要性を述べた.