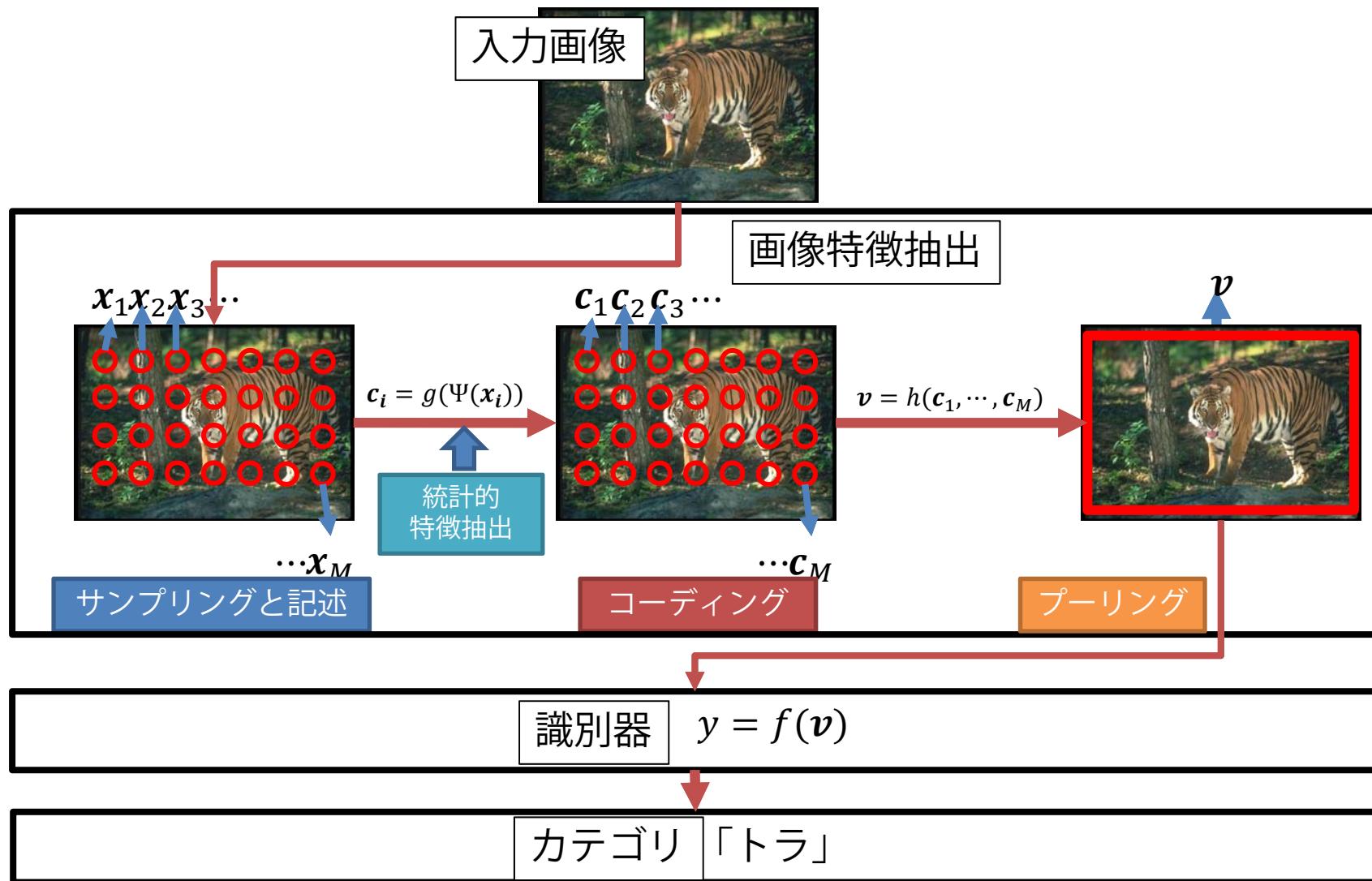


知能情報論 統計的特徴抽出

2016年5月18日

東京大学 大学院情報理工学系研究科
原田達也

カテゴリ認識の流れ



Schwartzらの人検出手法 (2009)



- W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis. Human Detection Using Partial Least Squares Analysis. In ICCV, 2009.
- 2009年までで、標準的な人検出データセットにおいて最高性能を出している手法。

Schwartzらの人検出手法デモ

Human Detection Using
Partial Least Squares Analysis

Paper ID: ICCV #350

Dalalらの人検出手法の問題点

- 人検出ではDalalらのHOG+SVMが有名であるが、HOGはエッジ情報しか利用していない。
- HOG特徴では人を検出するための重要な情報が欠けているのでFalse Positive（間違って人だと検出する）率が高くなる。



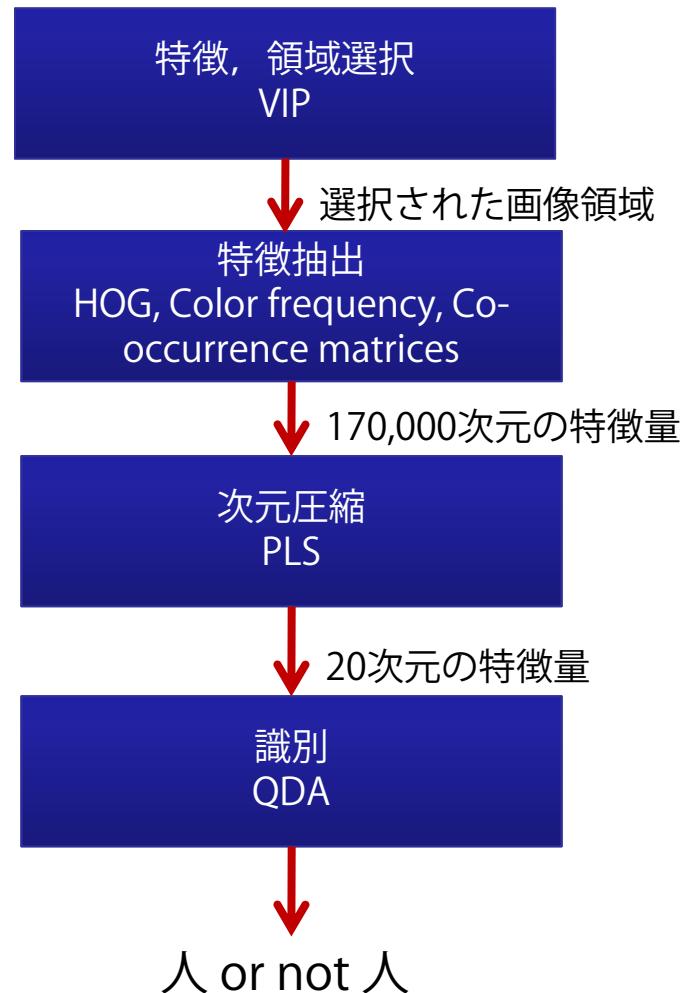
Figure 2. False positives obtained when only edge information (using HOG features) is considered.

HOGを利用して間違って人だと検出された例

Schwartzらの手法の概要

- HOGはエッジしか扱わないので、テクスチャやカラーを特徴量として活用する。
- いろんな特徴量を利用すると人とそれ以外との判別する特徴が含まれるので識別能力が上がるようと思われるが、特徴量の次元が膨大となり識別器を構成できなくなる。
 - Schwartzらの論文では約170,000次元の特徴量を扱っている！
- 膨大な次元を効率よく圧縮し、圧縮された空間で識別器を構成する。
 - ここでは次元圧縮にPartial Least Squareを利用して、170,000次元から20次元まで圧縮する。
- 圧縮された空間ではQDAを用いて識別器を構成する。
 - 性質の良い特徴量を用いているのでSVMを利用しなくとも、同等の性能かつ高速な識別器を構成できる。

Schwartzらの手法の概要



Schwartzら以前の人検出手法

- Dalalらの手法
 - N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In CVPR, 2005.
 - HOG+ Linear SVM
- A. Ess, B. Leibe, and L. V. Gool. Depth and appearance for mobile scene analysis. In ICCV, 2007.
 - 2007年時点で最も性能がよかつた手法.
 - Depth map, 地面推定, オクルージョンの推論などを利用しており手法は複雑

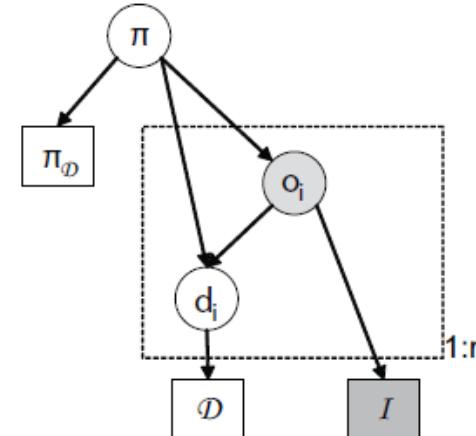
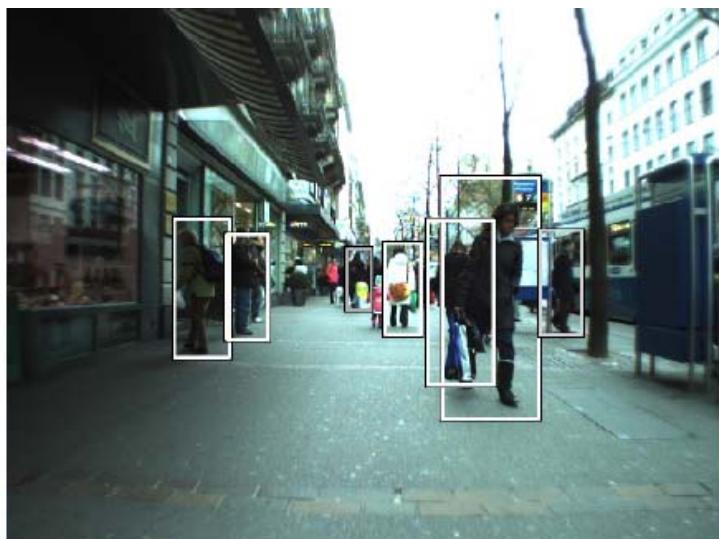
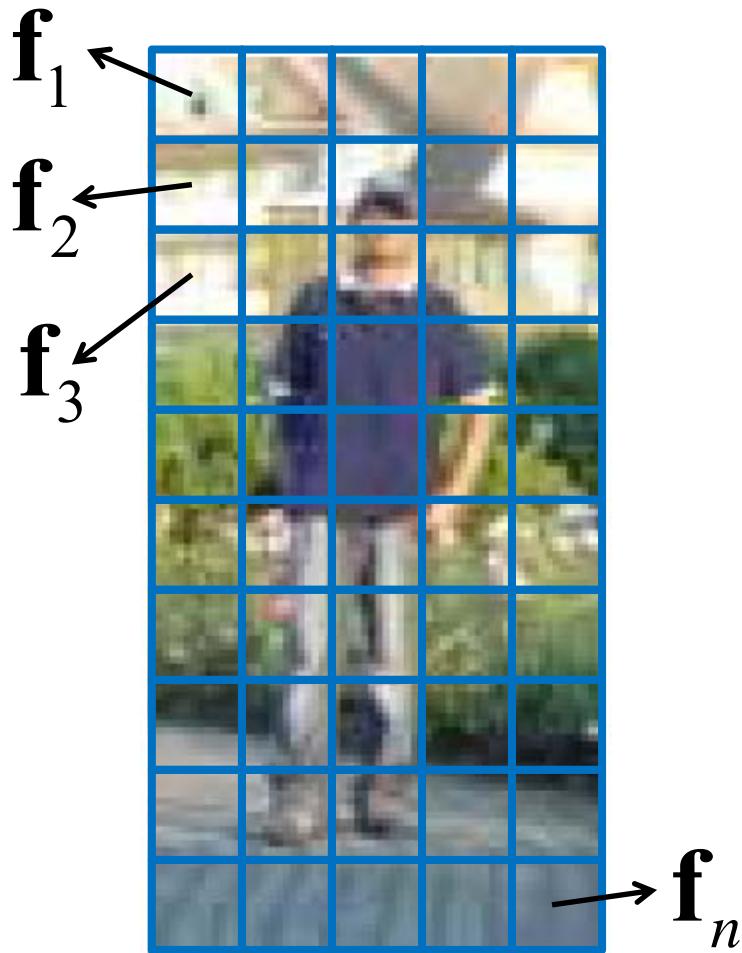


Figure 2. Graphical model for the fusion of object and ground plane detection with the aid of a depth map. The shaded areas indicate implicit loops that are solved in a novel two-stage approach.

利用している特徴量

$$\mathbf{f}_i = (\text{texture}, \text{edge}, \text{color})$$



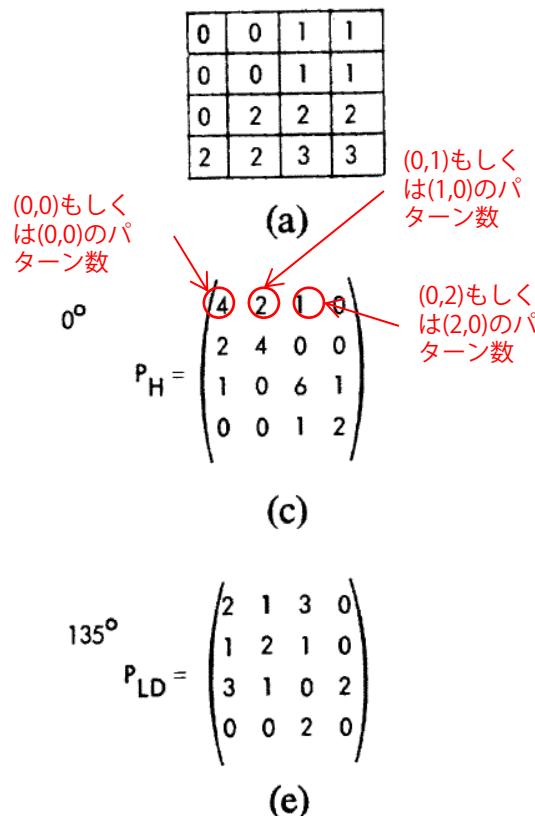
$$\mathbf{f} = (\mathbf{f}_1^T \ \mathbf{f}_2^T \ \mathbf{f}_3^T \ \dots \ \mathbf{f}_n^T)^T$$

- テクスチャ
 - Co-occurrence matrices
 - R. Haralick, K. Shanmugam, and I. Dinstein. Texture Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6), 1973.
 - 2次のテクスチャ情報を表現
 - ブロック内における隣り合うピクセルの結合確率密度分布
- エッジ
 - HOG
 - N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In CVPR, 2005.
- カラー
 - HOGはRGBの各カラーチャンネルで輝度勾配を計算→最も勾配の大きなカラーチャンネルの値を特徴量として採用
 - どのカラーチャンネルから勾配が得られたのか?をヒストグラムで表現し、カラー特徴とする。

Co-occurrence Matrices

- R. Haralick, K. Shanmugam, and I. Dinstein. Texture Features for Image Classification. IEEE Transactions on Systems, Man, and Cybernetics, 3(6), 1973.

4x4の画像、0から3のグレースケール



		Grey Tone			
		0	1	2	3
Grey Tone	0	#(0,0)	#(0,1)	#(0,2)	#(0,3)
	1	#(1,0)	#(1,1)	#(1,2)	#(1,3)
Grey Tone	2	#(2,0)	#(2,1)	#(2,2)	#(2,3)
	3	#(3,0)	#(3,1)	#(3,2)	#(3,3)

(a) $(0,1)$ もしくは $(1,0)$ のパターン数
 $(0,2)$ もしくは $(2,0)$ のパターン数
 $(0,0)$ もしくは $(0,0)$ のパターン数
 $(0,1)'$ もしくは $(0,0)'$ のパターン数
 $(0,2)'$ もしくは $(2,0)'$ のパターン数

Gray-tone spatial-dependence matricesを利用して多种のテクスチャ特徴を計算できる

Angular second moment

$$f_1 = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \left(\frac{P(i,j)}{R} \right)^2$$

Contrast

$$f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{|i-j|=n} \left(\frac{P(i,j)}{R} \right) \right\}$$

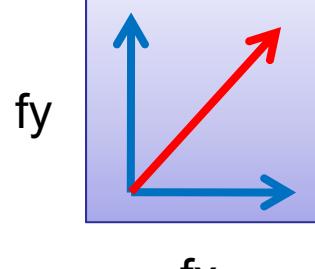
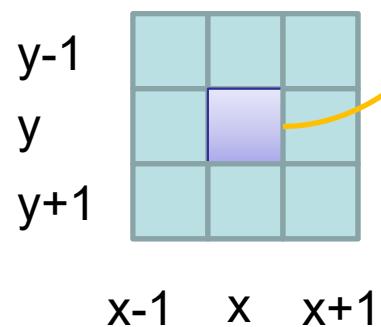
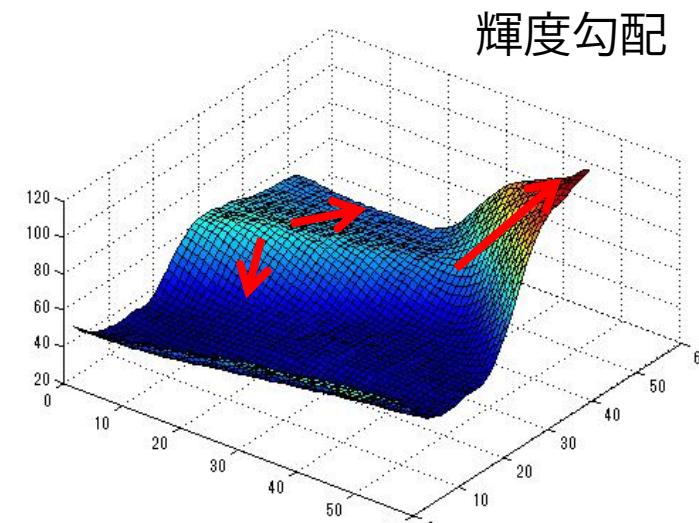
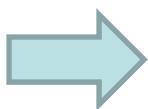
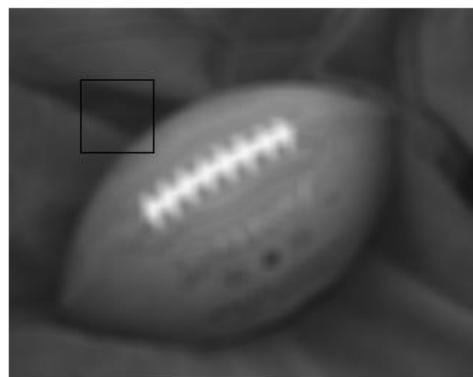
Correlation

$$f_3 = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [ijP(i,j)/R] - \mu_x \mu_y}{\sigma_x \sigma_y}$$

Fig. 3. (a) 4×4 image with four gray-tone values 0–3. (b) General form of any gray-tone spatial-dependence matrix for image with gray-tone values 0–3. $\#(i,j)$ stands for number of times gray tones i and j have been neighbors. (c)–(f) Calculation of all four distance 1 gray-tone spatial-dependence matrices.

P: Gray-tone spatial-dependence matrices

勾配の計算



輝度勾配

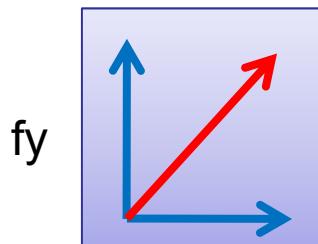
局所輝度勾配の計算

$$fx = I(x+1, y) - I(x-1, y)$$
$$fy = I(x, y+1) - I(x, y-1)$$

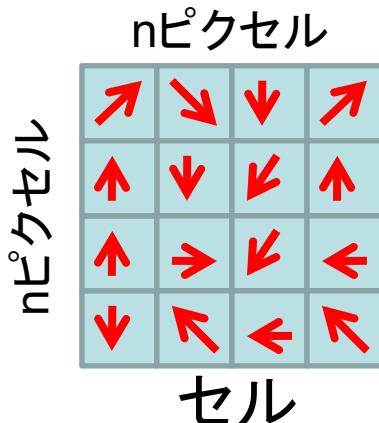
- 各ピクセル毎に勾配を計算する。
- 勾配計算前にガウシアンカーネルの平滑化を試したが性能は良くなかった。
- カラー画像の場合は各カラーチャンネルで勾配を計算する。

勾配のヒストグラム化

1. 全てのピクセルで輝度勾配を計算

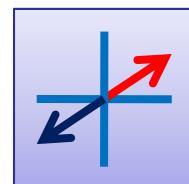


3. $n \times n$ ピクセルをひとまとまりとする → セル

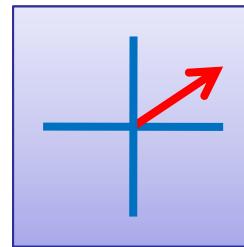


セル：1次元勾配ヒストグラムを計算する単位

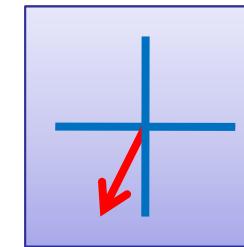
- 符号無し ($0^\circ \sim 180^\circ$) , 20° 每のヒストグラムが性能が高い。
- つまり1セルあたり9次元のベクトルで表現される。
- Dalalらの論文では 8×8 ピクセルを1セルとしている。



2. 勾配方向の量子化

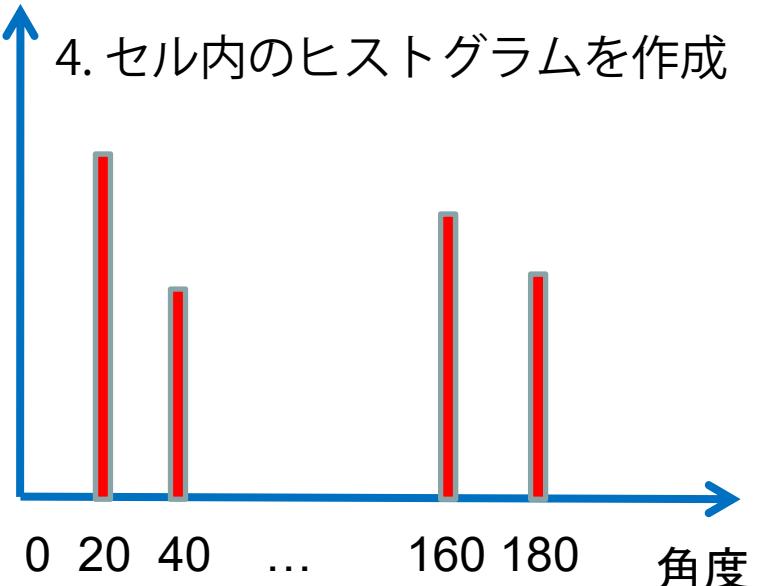


$37^\circ \rightarrow 40^\circ$



$256^\circ \rightarrow 260^\circ$

4. セル内のヒストグラムを作成



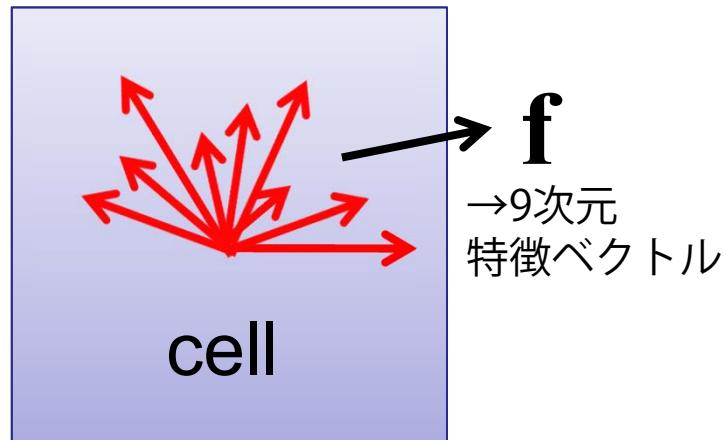
5. 特徴ベクトル

$\mathbf{f} \rightarrow 9$ 次元

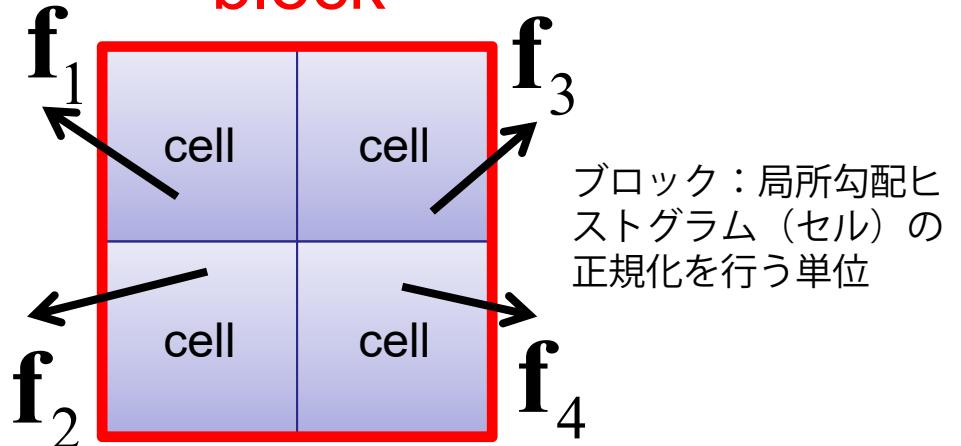
勾配を同じとみなす：符号なし
勾配を異なるとみなす：符号あり

正規化と記述子ブロック

1. 全てのセルの局所勾配ヒストグラムを作成



2. 複数の隣接するセルをまとめてブロック化
block



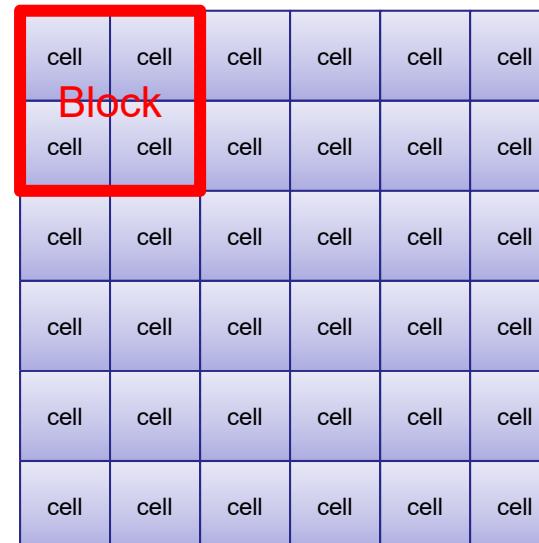
3. ブロックの特徴ベクトル

$$\mathbf{v} = (\mathbf{f}_1^T \ \mathbf{f}_2^T \ \mathbf{f}_3^T \ \mathbf{f}_4^T)^T$$

4. ブロック内での特徴ベクトルの正規化

$$\mathbf{v} = \frac{\mathbf{v}}{\sqrt{\|\mathbf{v}\|_2^2 + \epsilon^2}}$$

5. ブロックを移動させて正規化



つまり各セルは何度も正規化されることになる。

冗長だが性能が良かった。

次元圧縮

- Partial Least Squares (PLS)の概要
 - 教師ありの次元圧縮手法
 - (注) PCAは教師なしの次元圧縮手法
 - Fisher Discriminant Analysis (FDA)
 - FDAも教師あり次元圧縮手法
 - FDAの問題点
 - $C-1$ 次元しか潜在空間とれない (C はクラス数) . つまり2クラス問題では1次元の潜在空間しかとれない.
 - サンプル数より次元数が大きいと計算ができない. この論文では170,000次元.
 - PLSは $C-1$ 次元より大きな潜在空間の次元をとれる
 - サンプル数が次元数よりも少なくても計算可能.

PLSの詳細

m次元のベクトルをp次元に圧縮する。
サンプルはn個あるとする。

$$X = TP^T + E$$
$$y = Uq^T + f$$

nxm : 特徴ベクトルを並べた行列

nxp : 潜在空間における特徴ベクトルを並べた行列

mxp : 特徴ベクトルを潜在空間へ射影する行列

nxm : 誤差行列

nx1 : 教師信号を並べたベクトル

nxp : 潜在空間における教師信号を並べた行列

1xp : 教師信号を潜在空間へ射影するベクトル

nx1 : 誤差行列

$T = (\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_p)$

$U = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p)$

Nonlinear Iterative Partial Least Square (NIPALS)

- 逐次的アルゴリズム

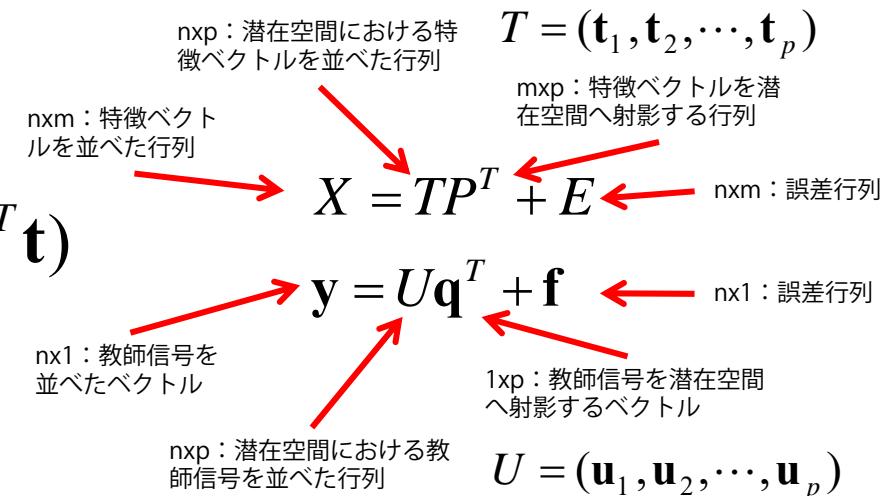
$$1) \mathbf{w} = X^T \mathbf{u} / (\mathbf{u}^T \mathbf{u}) \quad 4) \mathbf{c} = Y^T \mathbf{t} / (\mathbf{t}^T \mathbf{t})$$

$$2) \|\mathbf{w}\| \rightarrow 1$$

$$5) \|\mathbf{c}\| \rightarrow 1$$

$$3) \mathbf{t} = X\mathbf{w}$$

$$6) \mathbf{u} = Y\mathbf{c}$$



- PLSは潜在空間における共分散最大化と等価
 - 本当はPLSは固有値問題を解くだけでいい
 - CCAと異なり分散を考慮していない

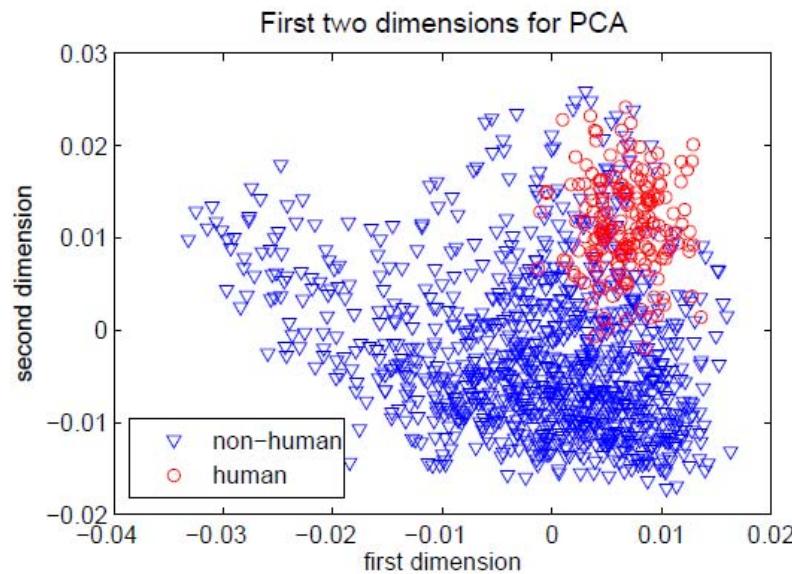
$$[cov(\mathbf{t}, \mathbf{u})]^2 = [cov(\mathbf{Xw}, \mathbf{Yc})]^2 = max_{|\mathbf{r}|=|\mathbf{s}|=1} [cov(\mathbf{Xr}, \mathbf{Ys})]^2$$

$$cov(\mathbf{t}, \mathbf{u}) = \mathbf{t}^T \mathbf{u} / n$$

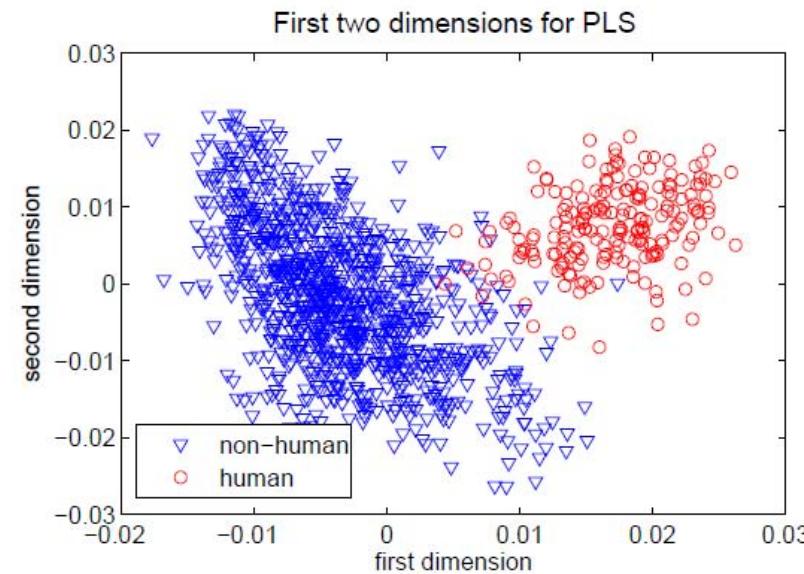
$$X^T Y Y^T X \mathbf{w} = \lambda \mathbf{w}$$

PLSの効果

- PCAとPLSとの比較
- human, non-humanを2次元に圧縮
- PLSの方がよく分離されている



(a) PCA - first two dimensions



(b) PLS - first two dimensions

検出の高速化

- 検出窓全てに170,000次元の特徴抽出は大変！
- 高速化
 - 計算する特徴量を絞り込む
 - 画像の小領域 (block) におけるデータ構造構築時間の削減 (Co-occurrence の計算時間)
- どのように少数の特徴を選択するのか？
 - PLSを基盤とした特徴選択手法
 - Variable Importance on Projection (VIP)
- データ構造構築時間の削減？
 - 各ブロック毎に重み付けをし、計算の優先順位をつける

Variable Importance on Projection (VIP)

$$X = TP^T + E$$

$T = (\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_p)$
 E

nxp : 潜在空間における特徴ベクトルを並べた行列
nxm : 特徴ベクトルを並べた行列
nxm : 誤差行列

 $y = U\mathbf{q}^T + \mathbf{f}$

nx1 : 教師信号を並べたベクトル
1xp : 教師信号を潜在空間へ射影するベクトル
nx1 : 誤差行列

 $U = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p)$

- 1) $\mathbf{w} = X^T \mathbf{u} / (\mathbf{u}^T \mathbf{u})$
- 2) $\|\mathbf{w}\| \rightarrow 1$
- 3) $\mathbf{t} = X\mathbf{w}$
- 4) $\mathbf{c} = Y^T \mathbf{t} / (\mathbf{t}^T \mathbf{t})$
- 5) $\|\mathbf{c}\| \rightarrow 1$
- 6) $\mathbf{u} = Y\mathbf{c}$

j番目の特徴の重要度

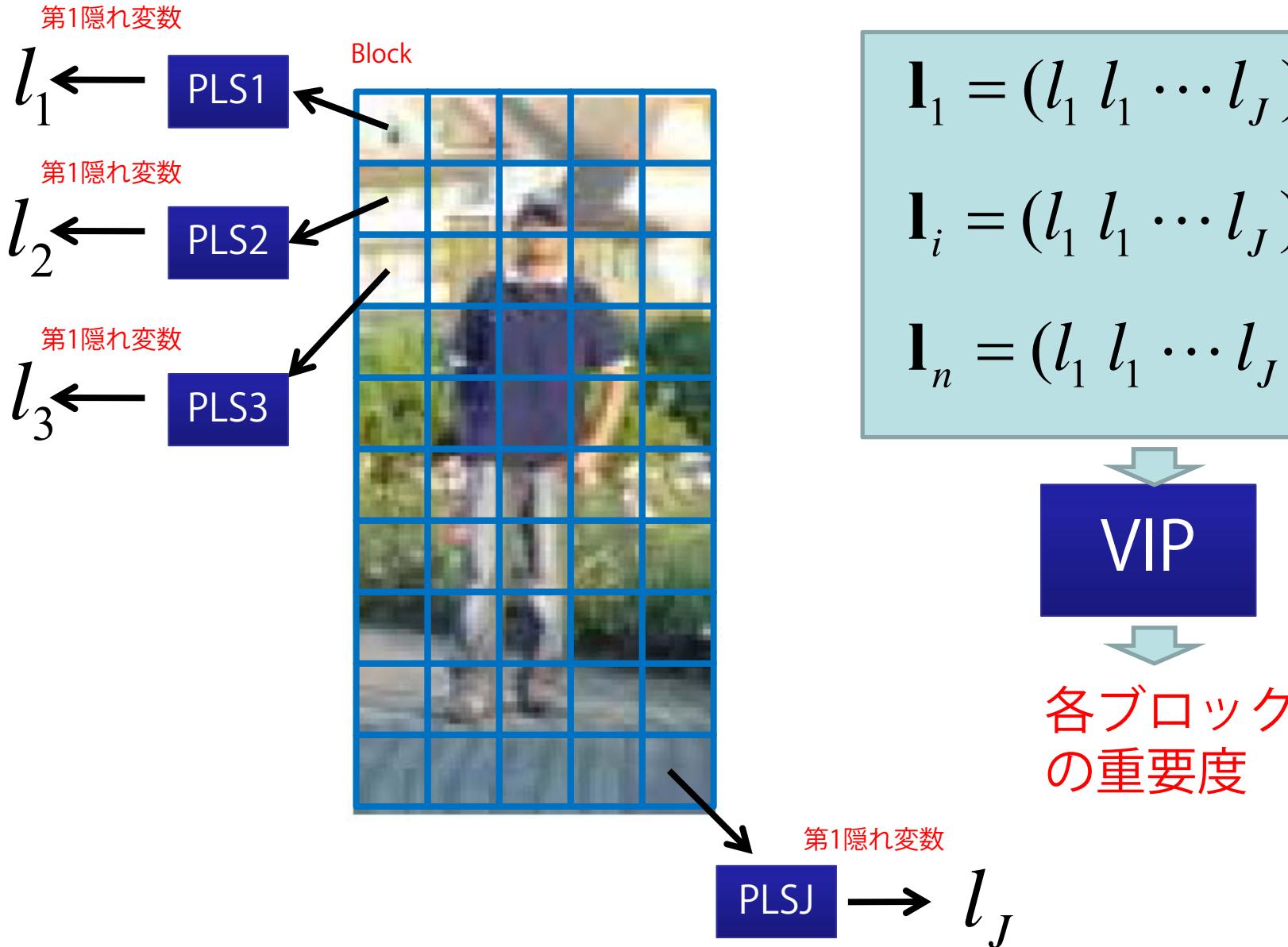
$$\text{VIP}_j = \sqrt{m \sum_{k=1}^p b_k^2 w_{jk}^2} / \sqrt{\sum_{k=1}^p b_k^2}$$

隠れ状態で分散が大きく、かつ射影の重みが大きければ重要な特徴！

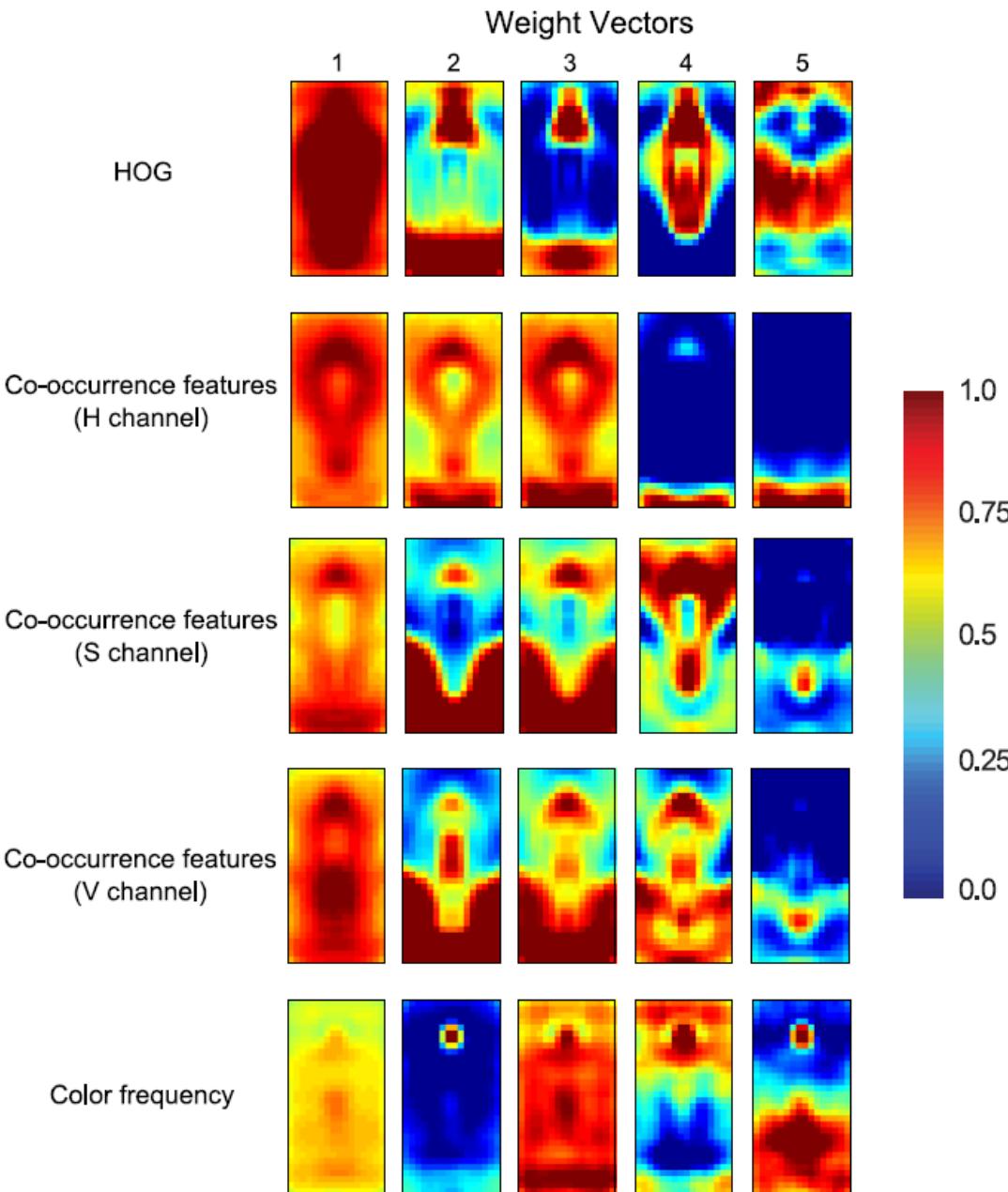
k番目の隠れ状態における共分散
 $b_k = \mathbf{u}_k^T \mathbf{t}_k$

j番目の特徴のk番目の隠れ状態における射影ベクトル
(重みベクトル)
 $\mathbf{t}_k = X\mathbf{w}_k$

各ブロック毎に重み付け



各特徴における領域の重要度



識別

- PLSで次元圧縮し、圧縮した空間で識別器を構成

- Quadratic Discriminant Analysis (QDA)
 - 各クラスkで平均と共分散行列を求める

$$\mu_k \quad \Sigma_k$$

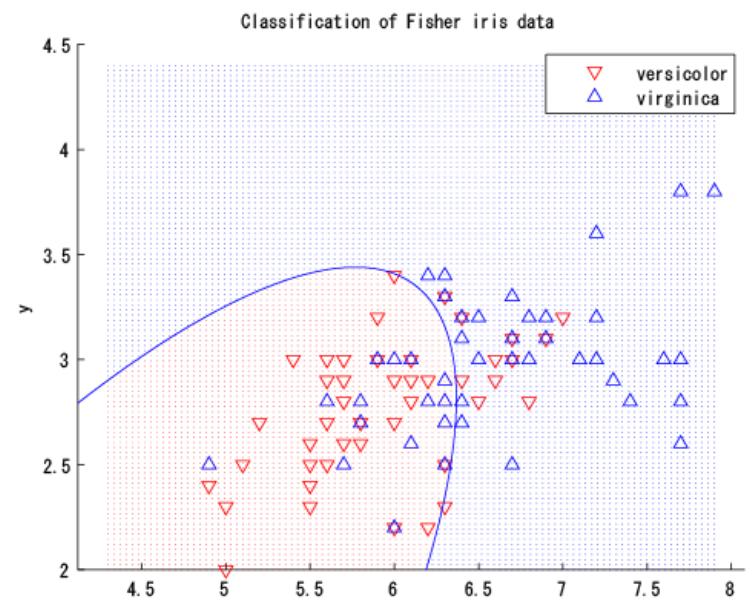
- Quadratic discriminant function

$$\delta_k(x) = -\frac{1}{2} \log |\Sigma_k| - \frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) + \log \pi_k$$

- 識別則

$$\arg \max_k \delta_k(x)$$

- 決定境界はxに対する2次関数



2次判別関数を使ったフィッシャーのアイリスデータの分類

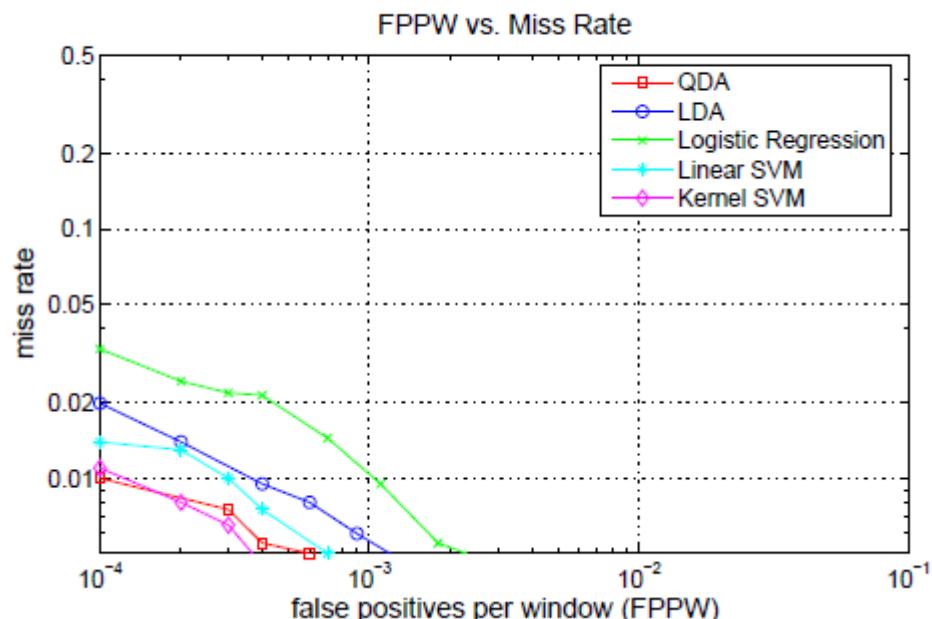
クラスkの事前確率

QDAと線形SVMの比較

- 計算速度
 - PLS+QDA vs 線形SVM
 - サンプル数が増えると線形SVMはメモリ不足で学習できない
 - Xeon 5160 8GB ram
- 識別性能
 - PLS+(QDA, LDA, 線形SVM, Kernel SVM, Logistic Regression)
 - Logistic Regression以外は性能は似たり寄ったり
 - 同じ性能ならPLS+QDAがシンプルでよい

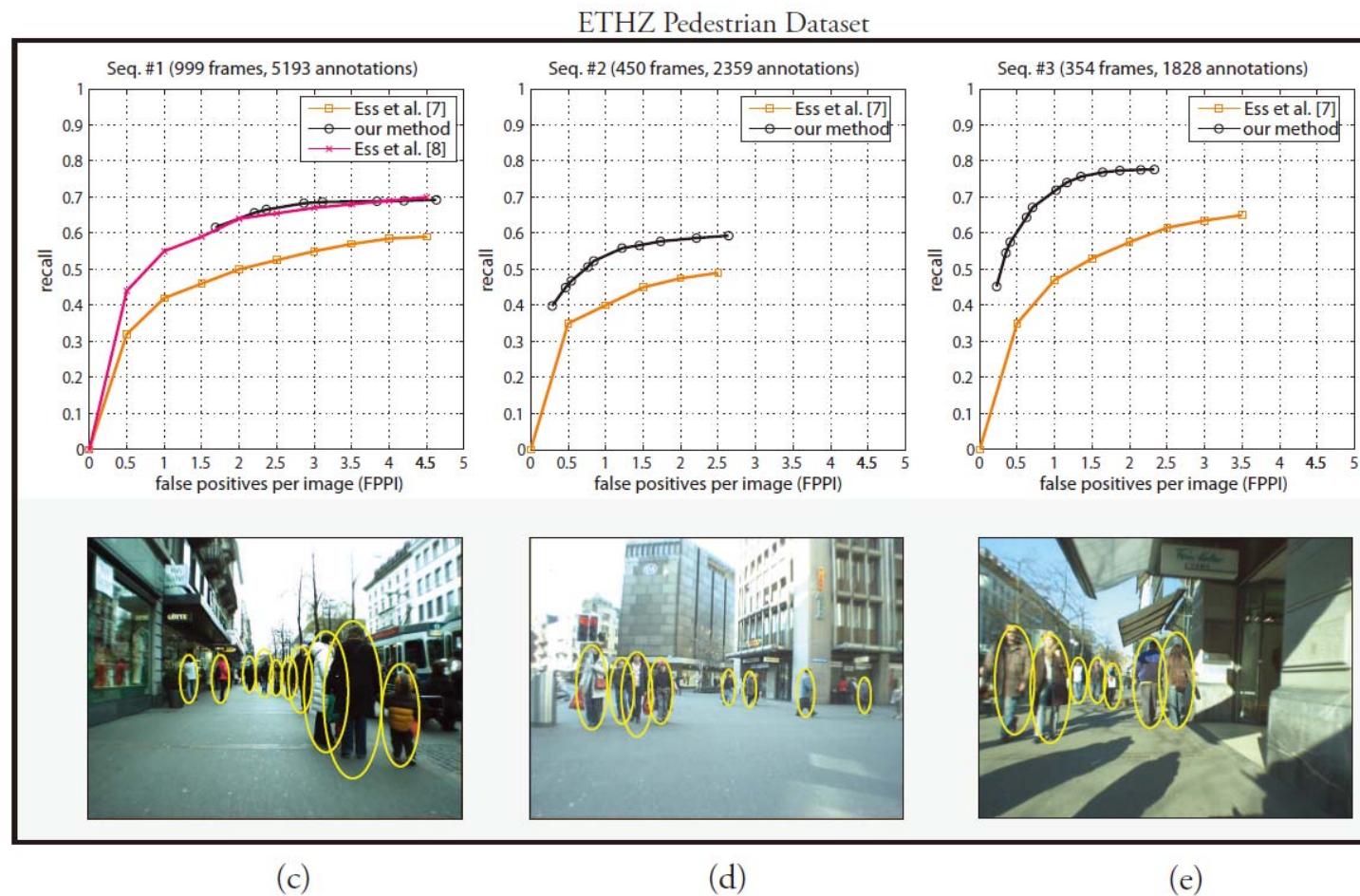
# samples	PLS + QDA	SVM
200	23.63	131.72
600	62.62	733.63
1000	97.38	1693.50
1400	135.81	2947.51
1800	174.57	4254.63
2200	213.93	-
11370	813.03	-

Table 1. Time, in seconds, to train SVM and PLS + QDA models. The number of features per sample is 170,820. The training time increases with an increase in the number of training samples.



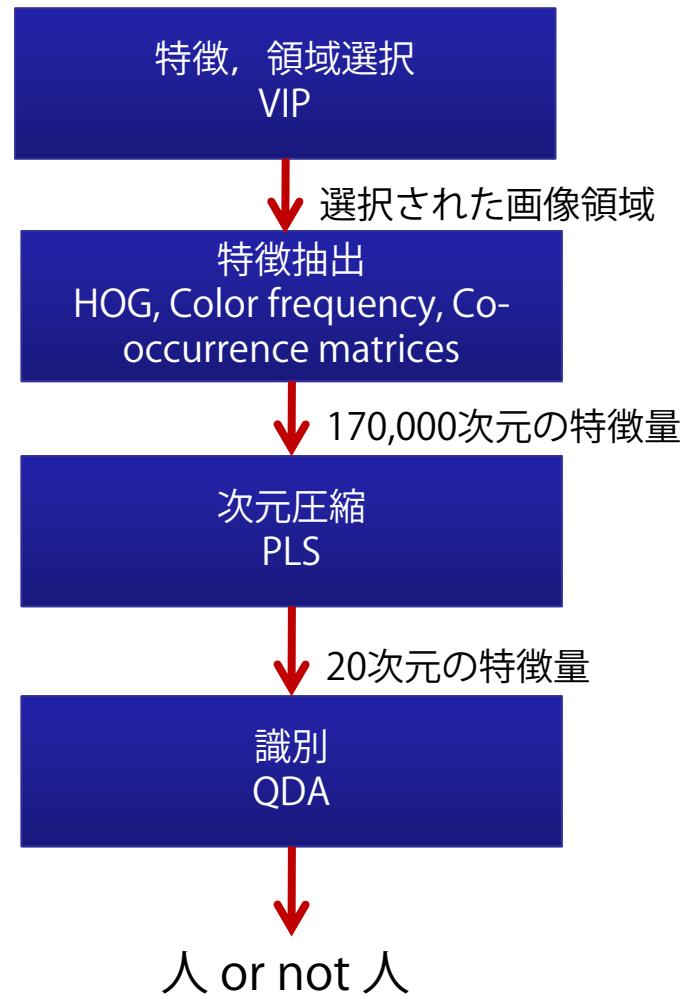
実験結果

- Essらの手法との比較
 - シンプルにもかかわらず高性能



Schwartzらの手法のまとめ

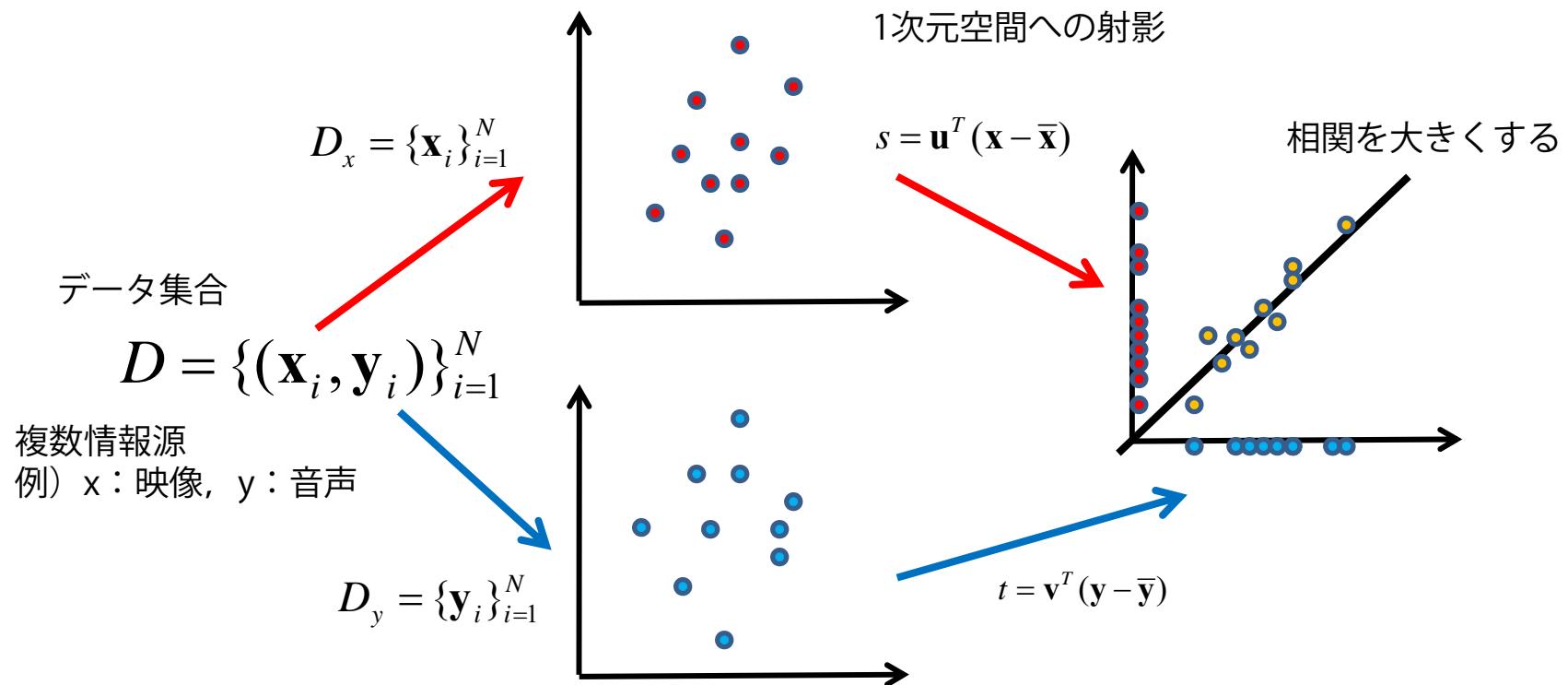
- エッジ (HOG) + テクスチャ + カラー特徴量の利用
- PLSによる効率的な次元圧縮
 - 170,000次元 → 20次元
- VIPによる特徴選択, ブロック選択
- 潜在空間におけるQDAを用いた識別器



次元圧縮手法の比較

- Principal Component Analysis (PCA)
- Linear Discriminant Analysis (LDA)
- Canonical Correlation Analysis (CCA)
- Partial Least Squares (PLS)

Canonical Correlation Analysis (CCA)



Canonical Correlation Analysis (CCA)

- 複数情報源に共通して含まれる情報抽出
 - データ集合

$$D = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N, \quad \mathbf{x}_i \in \mathbb{R}^{m_x}, \mathbf{y}_i \in \mathbb{R}^{m_y}$$

- 射影ベクトルによる1次元空間への射影

$$s = \mathbf{u}^T (\mathbf{x} - \bar{\mathbf{x}}) \quad t = \mathbf{v}^T (\mathbf{y} - \bar{\mathbf{y}})$$

平均 $\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$ 平均 $\bar{\mathbf{y}} = \frac{1}{N} \sum_{i=1}^N \mathbf{y}_i$

- 射影されたデータ間の相関

$$\begin{aligned}\text{cov}[s, t] &= \frac{1}{N} \sum_{i=1}^N s_i t_i \\ &= \frac{1}{N} \sum_{i=1}^N \mathbf{u}^T (\mathbf{x}_i - \bar{\mathbf{x}}) \mathbf{v}^T (\mathbf{y} - \bar{\mathbf{y}}) \\ &= \mathbf{u}^T \left(\frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{y} - \bar{\mathbf{y}})^T \right) \mathbf{v} \\ &= \mathbf{u}^T C_{xy} \mathbf{v}\end{aligned}$$

$$\begin{aligned}\text{cor}[s, t] &= \frac{\text{cov}[s, t]}{\sqrt{\text{var}[s]} \sqrt{\text{var}[t]}} \\ &= \frac{\mathbf{u}^T C_{xy} \mathbf{v}}{\sqrt{\mathbf{u}^T C_{xx} \mathbf{u}} \sqrt{\mathbf{v}^T C_{yy} \mathbf{v}}}\end{aligned}$$

$$\begin{aligned}\text{var}[s] &= \frac{1}{N} \sum_{i=1}^N s_i^2 \\ &= \frac{1}{N} \sum_{i=1}^N \mathbf{u}^T (\mathbf{x}_i - \bar{\mathbf{x}}) \mathbf{u}^T (\mathbf{x}_i - \bar{\mathbf{x}}) \\ &= \mathbf{u}^T \left(\frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \right) \mathbf{u} \\ &= \mathbf{u}^T C_{xx} \mathbf{u} \\ \text{var}[t] &= \mathbf{u}^T C_{yy} \mathbf{u}\end{aligned}$$

Canonical Correlation Analysis (CCA)

- 複数情報源に共通して含まれる情報抽出
 - 相関の最大化

$$\text{cor}[s, t] = \frac{\mathbf{u}^T C_{xy} \mathbf{v}}{\sqrt{\mathbf{u}^T C_{xx} \mathbf{u}} \sqrt{\mathbf{v}^T C_{yy} \mathbf{v}}}$$

\mathbf{u}, \mathbf{v} を定数倍しても
 $\text{cor}[s, t]$ の値は変わらない

$$\begin{aligned}\mathbf{u}^T C_{xx} \mathbf{u} &= 1 \\ \mathbf{v}^T C_{yy} \mathbf{v} &= 1\end{aligned}$$

- ラグランジュ関数

$$J = \mathbf{u}^T C_{xy} \mathbf{v} - \frac{\lambda_1}{2} (\mathbf{u}^T C_{xx} \mathbf{u} - 1) - \frac{\lambda_2}{2} (\mathbf{v}^T C_{yy} \mathbf{v} - 1)$$

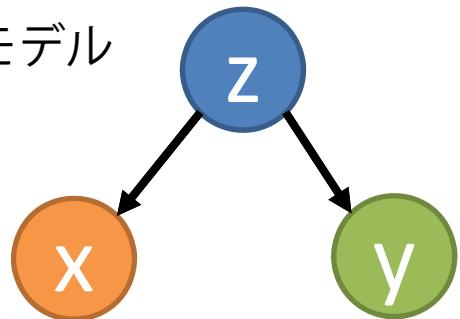
$$\frac{\partial J}{\partial \mathbf{u}} = 0, \quad \frac{\partial J}{\partial \mathbf{v}} = 0 \quad \rightarrow \quad \begin{aligned}C_{xy} C_{yy}^{-1} C_{yx} \mathbf{u} &= \rho^2 C_{xx} \mathbf{u} \\ C_{yx} C_{xx}^{-1} C_{xy} \mathbf{v} &= \rho^2 C_{yy} \mathbf{v}\end{aligned} \quad \begin{array}{l} \text{一般化固有値問題} \\ \rho^2 = \lambda_1 \lambda_2 \end{array}$$

$$\begin{pmatrix} 0 & C_{xy} \\ C_{xy} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}$$

確率的CCA

- 正準相関分析の確率的解釈

モデル



$$p(z) = \mathcal{N}(0, I_{d_z}), \min\{d_x, d_y\} \geq d_z \geq 1$$

$$p(\mathbf{x}|z) = \mathcal{N}(W_x z + \boldsymbol{\mu}_x, \Psi_x), W_x \in \mathbb{R}^{d_x \times d_z}, \Psi_x \succeq 0$$

$$p(\mathbf{y}|z) = \mathcal{N}(W_y z + \boldsymbol{\mu}_y, \Psi_y), W_y \in \mathbb{R}^{d_y \times d_z}, \Psi_y \succeq 0$$

パラメータの最尤推定

$$\hat{\boldsymbol{\mu}}_x = \bar{\mathbf{x}}, \quad \hat{\boldsymbol{\mu}}_y = \bar{\mathbf{y}}$$

$$\hat{W}_x = C_{xx} U_{xm_z} M_x, \quad \hat{W}_y = C_{yy} U_{ym_z} M_y$$

$$\hat{\Psi}_x = C_{xx} - \hat{W}_x \hat{W}_x^\top, \quad \hat{\Psi}_y = C_{yy} - \hat{W}_y \hat{W}_y^\top$$

Bach, Francis R and Jordan, Michael I. A probabilistic interpretation of canonical correlation analysis. Technical Report 688, Department of Statistics, University of California, Berkeley, 2005.

潜在変数 z の期待値と分散

$$E(z|\mathbf{x}) = M_x^\top U_{xd_z}^\top (\mathbf{x} - \bar{\mathbf{x}}), \quad E(z|\mathbf{y}) = M_y^\top U_{yd_z}^\top (\mathbf{y} - \bar{\mathbf{y}})$$

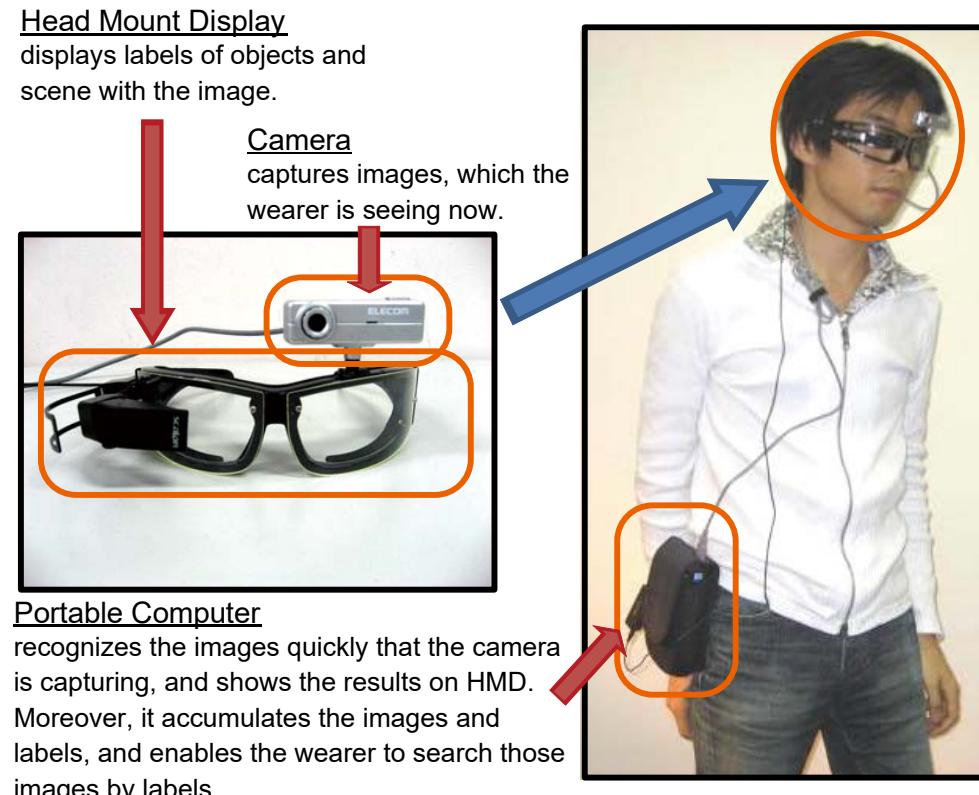
$$\text{var}(z|\mathbf{x}) = I - M_x M_x^\top, \quad \text{var}(z|\mathbf{y}) = I - M_y M_y^\top$$

$$E(z|\mathbf{x}, \mathbf{y}) = \begin{pmatrix} M_x \\ M_y \end{pmatrix}^\top L \begin{pmatrix} U_{xd_z}^\top (\mathbf{x} - \bar{\mathbf{x}}) \\ U_{yd_z}^\top (\mathbf{y} - \bar{\mathbf{y}}) \end{pmatrix}$$

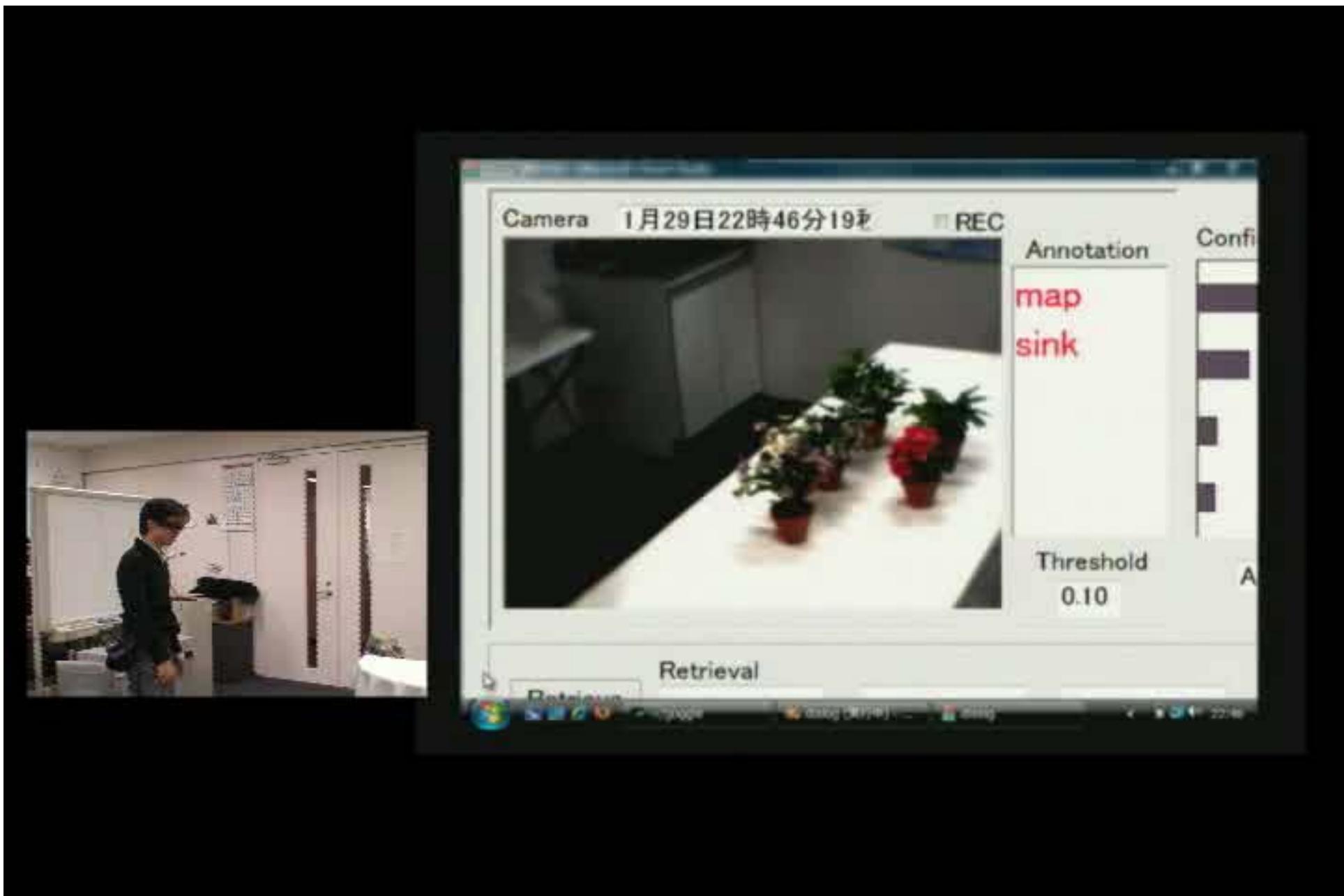
$$\text{var}(z|\mathbf{x}, \mathbf{y}) = I - \begin{pmatrix} M_x \\ M_y \end{pmatrix}^\top L \begin{pmatrix} M_x \\ M_y \end{pmatrix}$$

実世界応用 1 人工知能ゴーグルの開発

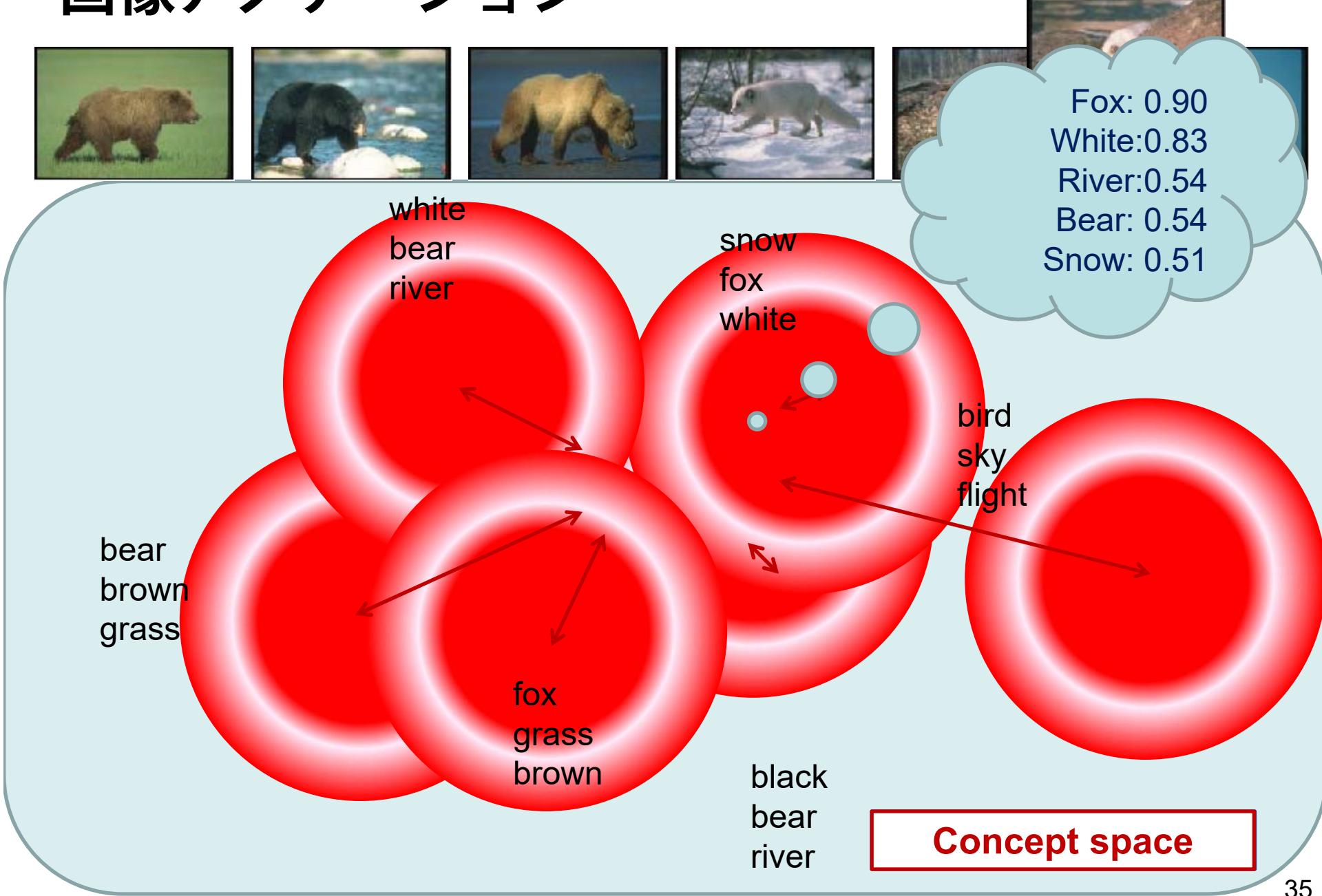
- 提案手法の実世界応用：人工知能ゴーグル
 - 身の回りの物体の素早い認識・検索を実現
 - HMDによる情報提示、記憶支援（忘れ物検索）



人工知能ゴーグルによる実世界リアルタイムアノテーション



画像アノテーション



確率的CCAを用いた計量

- カルバック・ライブラ一情報量 (KL 情報量)
 - 確率分布 P, Q 間の差異を計る尺度

$$D_{\text{KL}}(P||Q) = \int p(x) \log \frac{p(x)}{q(x)} d\nu(x)$$

$$\dot{\mathbf{z}} = E(\mathbf{z}|\mathbf{x}) = M_x^T U_x^T (\mathbf{x} - \bar{\mathbf{x}}),$$

$$\Phi_x = \text{var}(\mathbf{z}|\mathbf{x}) = I - M_x M_x^T.$$

$$\ddot{\mathbf{z}} = E(\mathbf{z}|\mathbf{x}, \mathbf{y}) = \begin{pmatrix} M_x \\ M_y \end{pmatrix}^T \begin{pmatrix} (I - \Lambda^2)^{-1} & -(I - \Lambda^2)^{-1} \Lambda \\ -(I - \Lambda^2)^{-1} \Lambda & (I - \Lambda^2)^{-1} \end{pmatrix} \begin{pmatrix} U_x^T (\mathbf{x} - \bar{\mathbf{x}}) \\ U_y^T (\mathbf{y} - \bar{\mathbf{y}}) \end{pmatrix},$$

$$\Phi_{xy} = \text{var}(\mathbf{z}|\mathbf{x}, \mathbf{y}) = I - \begin{pmatrix} M_x \\ M_y \end{pmatrix}^T \begin{pmatrix} (I - \Lambda^2)^{-1} & -(I - \Lambda^2)^{-1} \Lambda \\ -(I - \Lambda^2)^{-1} \Lambda & (I - \Lambda^2)^{-1} \end{pmatrix} \begin{pmatrix} M_x \\ M_y \end{pmatrix}.$$

1-view CCD (CCD1)

$$KL(p(\mathbf{z}|\mathbf{x}_q), p(\mathbf{z}|\mathbf{x}_t)) = (\dot{\mathbf{z}}_q - \dot{\mathbf{z}}_t)^T \Phi_x^{-1} (\dot{\mathbf{z}}_q - \dot{\mathbf{z}}_t).$$

2-view CCD (CCD2)

$$KL(p(\mathbf{z}|\mathbf{x}_q), p(\mathbf{z}|\mathbf{x}_t, \mathbf{y}_t)) = \frac{1}{2} \log \frac{|\Phi_{xy}|}{|\Phi_x|} - \frac{d}{2} + \frac{1}{2} \text{Tr}(\Phi_{xy}^{-1} \Phi_x) + (\dot{\mathbf{z}}_q - \ddot{\mathbf{z}}_t)^T \Phi_{xy}^{-1} (\dot{\mathbf{z}}_q - \ddot{\mathbf{z}}_t).$$

Partial Least Squares (PLS)

- 複数情報源に共通して含まれる情報抽出
 - データ集合

$$D = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N, \quad \mathbf{x}_i \in \mathbb{R}^{m_x}, \mathbf{y}_i \in \mathbb{R}^{m_y}$$

- 射影ベクトルによる1次元空間への射影

$$s = \mathbf{u}^T (\mathbf{x} - \bar{\mathbf{x}}) \quad t = \mathbf{v}^T (\mathbf{y} - \bar{\mathbf{y}})$$

平均 $\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$ 平均 $\bar{\mathbf{y}} = \frac{1}{N} \sum_{i=1}^N \mathbf{y}_i$

- 射影されたデータ間の共分散

$$\begin{aligned} \text{cov}[s, t] &= \frac{1}{N} \sum_{i=1}^N s_i t_i \\ &= \frac{1}{N} \sum_{i=1}^N \mathbf{u}^T (\mathbf{x}_i - \bar{\mathbf{x}}) \mathbf{v}^T (\mathbf{y} - \bar{\mathbf{y}}) \\ &= \mathbf{u}^T \left(\frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{y} - \bar{\mathbf{y}})^T \right) \mathbf{v} \\ &= \mathbf{u}^T C_{xy} \mathbf{v} \end{aligned}$$

Partial Least Squares (PLS)

- 複数情報源に共通して含まれる情報抽出
 - 共分散の最大化

$$\text{cov}[s, t] = \mathbf{u}^T C_{xy} \mathbf{v}$$

- ラグランジュ関数

$$\begin{array}{l} \|\mathbf{u}\| \rightarrow \infty \\ \|\mathbf{v}\| \rightarrow \infty \end{array} \xrightarrow{\text{を防ぐ}} \begin{array}{l} \mathbf{u}^T \mathbf{u} = 1 \\ \mathbf{v}^T \mathbf{v} = 1 \end{array} \text{を仮定}$$

$$J = \mathbf{u}^T C_{xy} \mathbf{v} - \frac{\lambda_1}{2} (\mathbf{u}^T \mathbf{u} - 1) - \frac{\lambda_2}{2} (\mathbf{v}^T \mathbf{v} - 1)$$

$$\frac{\partial J}{\partial \mathbf{u}} = 0, \quad \frac{\partial J}{\partial \mathbf{v}} = 0 \xrightarrow{\text{ }} \begin{array}{l} C_{xy} C_{yx} \mathbf{u} = \lambda_1 \lambda_2 \mathbf{u} \\ C_{yx} C_{xy} \mathbf{v} = \lambda_1 \lambda_2 \mathbf{v} \end{array} \quad \text{固有値問題}$$

$$\begin{pmatrix} 0 & C_{xy} \\ C_{xy} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}$$

PLSの逐次的解法との関係

NIPALS

$$\begin{aligned}\mathbf{w}_n &= X^\top \mathbf{u}_{n-1} / (\mathbf{u}_{n-1}^\top \mathbf{u}_{n-1}) \\ \|\mathbf{w}_n\| &\rightarrow 1 \\ \mathbf{t}_n &= X \mathbf{w}_n \\ \mathbf{c}_n &= Y^\top \mathbf{t}_n / (\mathbf{t}_n^\top \mathbf{t}_n) \\ \|\mathbf{c}_n\| &\rightarrow 1 \\ \mathbf{u}_n &= Y \mathbf{c}_n\end{aligned}$$

$$\begin{aligned}\mathbf{w}_n &= X^\top \mathbf{u}_{n-1} / \|\mathbf{u}_{n-1}\|^2 \\ &= X^\top Y \mathbf{c}_{n-1} / \|\mathbf{u}_{n-1}\|^2 \|\mathbf{c}_{n-1}\|^2 \\ &= X^\top Y Y^\top \mathbf{t}_{n-1} / \|\mathbf{u}_{n-1}\|^2 \|\mathbf{c}_{n-1}\|^2 \|\mathbf{t}_{n-1}\|^2 \\ &= X^\top Y Y^\top X \mathbf{w}_{n-1} / \|\mathbf{u}_{n-1}\|^2 \|\mathbf{c}_{n-1}\|^2 \|\mathbf{t}_{n-1}\|^2 \|\mathbf{w}_{n-1}\|^2\end{aligned}$$

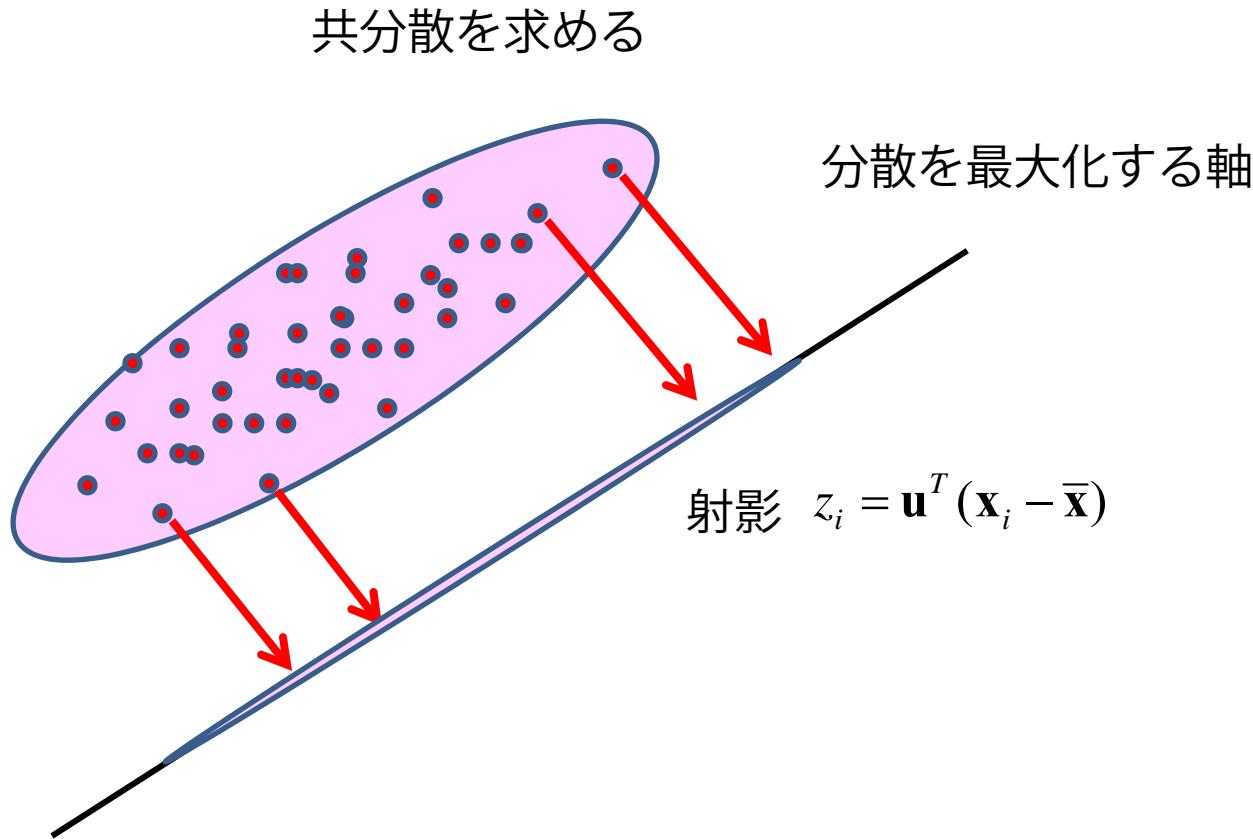
収束

$$\xrightarrow{} X^\top Y Y^\top X \mathbf{w} = \alpha \mathbf{w} \quad \alpha = \|\mathbf{t}\|^2 \|\mathbf{w}\|^2 \|\mathbf{u}\|^2 \|\mathbf{c}\|^2$$



$$\boxed{C_{xy} C_{yx} \mathbf{w} = \hat{\alpha} \mathbf{w}}$$

Principal Component Analysis (PCA)



Principal Component Analysis (PCA)

- 分散最大化基準
 - データ集合
 - 射影ベクトルによる1次元空間への射影

$$D = \{\mathbf{x}_i\}_{i=1}^N, \quad \mathbf{x}_i \in \mathbb{R}^m$$

– 射影されたデータの分散

$$z_i = \mathbf{u}^T (\mathbf{x}_i - \bar{\mathbf{x}})$$

平均 $\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$

$$\text{var}[z] = \frac{1}{N} \sum_{i=1}^N z_i^2 = \frac{1}{N} \sum_{i=1}^N \mathbf{u}^T (\mathbf{x}_i - \bar{\mathbf{x}}) \mathbf{u}^T (\mathbf{x}_i - \bar{\mathbf{x}}) = \mathbf{u}^T C_{xx} \mathbf{u}$$

共分散行列

$$C_{xx} = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

– 射影された分散の最大化

$\|\mathbf{u}\| \rightarrow \infty$ を防ぐ $\Rightarrow \mathbf{u}^T \mathbf{u} = 1$ を仮定

$$J = \mathbf{u}^T C_{xx} \mathbf{u} - \lambda(\mathbf{u}^T \mathbf{u} - 1)$$

$$\frac{\partial J}{\partial \mathbf{u}} = 0 \Rightarrow C_{xx} \mathbf{u} = \lambda \mathbf{u}$$



共分散行列の
固有値問題

$$C_{xx} \mathbf{U} = \mathbf{U} \Lambda$$

$$\mathbf{U} = (\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_M)$$

$$\Lambda = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_M \end{pmatrix}$$

Eigenface

Turk and A. Pentland. Eigenfaces for recognition.
Journal of Cognitive Neuroscience 3 (1): 71–86,
1991.



Training data



Figure 1. (b) The average face Ψ .

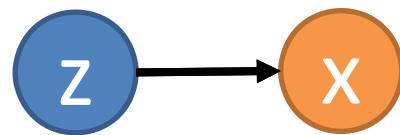


Figure 2. Seven of the eigenfaces calculated from the input images of Figure 1. 42

確率的PCA

- 主成分分析の確率的解釈

モデル



$$p(z) = \mathcal{N}(0, I_{d_z})$$
$$p(x|z) = \mathcal{N}(Wz + \mu, \sigma^2 I_{d_x})$$

パラメータの最尤推定

$$\mu = \bar{x}$$
$$\hat{W} = U_{d_z}(\Lambda_{d_z} - \sigma^2 I)^{1/2}R,$$
$$\hat{\sigma}^2 = \frac{1}{d_x - d_z} \sum_{i=d_z+1}^{d_x} \lambda_i$$

潜在変数zの期待値

$$E(z|x) = R^\top (\Lambda_{d_z} - \sigma^2 I)^{1/2} \Lambda_{d_z}^{-1} U_{d_z}^\top (x - \bar{x}).$$

Bach, Francis R and Jordan, Michael I. A probabilistic interpretation of canonical correlation analysis.
Technical Report 688, Department of Statistics,
University of California, Berkeley, 2005.

PCAによる白色化

- 白色化 (whitening)
 - 球状化 (sphereing)
 - データの平均を0, 共分散を単位行列とする正規化

$$U = (\mathbf{u}_1 \ \mathbf{u}_2 \cdots \mathbf{u}_M)$$

$$C_{xx} U = U \Lambda$$
$$\Lambda = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_M \end{pmatrix}$$


$$z_i = \Lambda^{-1/2} U^T (\mathbf{x}_i - \bar{\mathbf{x}})$$

白色化とk-meansを利用した局所特徴学習

1. Normalize inputs:

$$x^{(i)} := \frac{x^{(i)} - \text{mean}(x^{(i)})}{\sqrt{\text{var}(x^{(i)}) + \epsilon_{\text{norm}}}}, \forall i$$

2. Whiten inputs:

$$\begin{aligned} [V, D] &:= \text{eig}(\text{cov}(x)); // \text{ So } VDV^\top = \text{cov}(x) \\ x^{(i)} &:= V(D + \epsilon_{\text{zca}}I)^{-1/2}V^\top x^{(i)}, \forall i \end{aligned}$$

3. Loop until convergence (typically 10 iterations is enough):

$$\begin{aligned} s_j^{(i)} &:= \begin{cases} \mathcal{D}^{(j)\top} x^{(i)} & \text{if } j == \arg \max_l |\mathcal{D}^{(l)\top} x^{(i)}| \\ 0 & \text{otherwise.} \end{cases} \quad \forall j, i \\ \mathcal{D} &:= XS^\top + \mathcal{D} \\ \mathcal{D}^{(j)} &:= \mathcal{D}^{(j)} / \|D^{(j)}\|_2 \quad \forall j \end{aligned}$$

$$\begin{aligned} \mathcal{D}_{\text{new}} &:= \arg \min_{\mathcal{D}} \|\mathcal{D}S - X\|_2^2 + \|\mathcal{D} - \mathcal{D}_{\text{old}}\|_2^2 \\ &= (SS^\top + I)^{-1}(XS^\top + \mathcal{D}_{\text{old}}) \\ &\propto XS^\top + \mathcal{D}_{\text{old}} \\ \mathcal{D}_{\text{new}} &:= \text{normalize}(\mathcal{D}_{\text{new}}). \end{aligned}$$

白色化とk-meansを利用した局所特徴学習

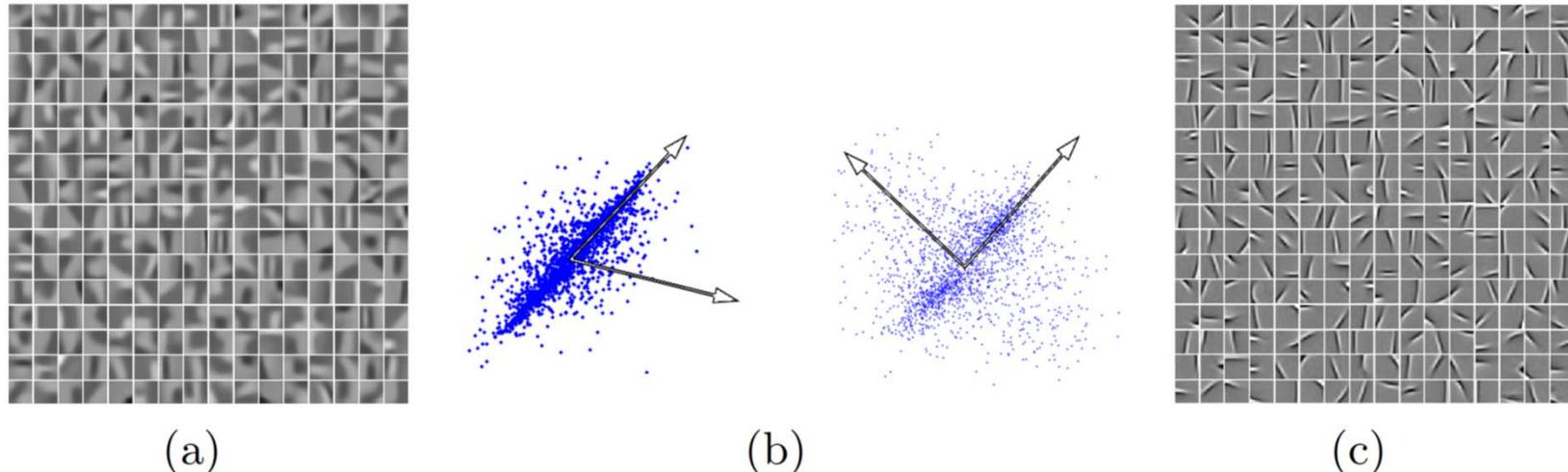


Fig. 22.1. (a) Centroids learned by K-means from natural images without whitening. (b) A cartoon depicting the effect of whitening on the K-means solution. Left: unwhitened data, where the centroids tend to be biased by the correlated data. Right: whitened data, where centroids are more orthogonal. (c) Centroids learned from whitened image patches.

白色化とk-meansを利用した局所特徴学習

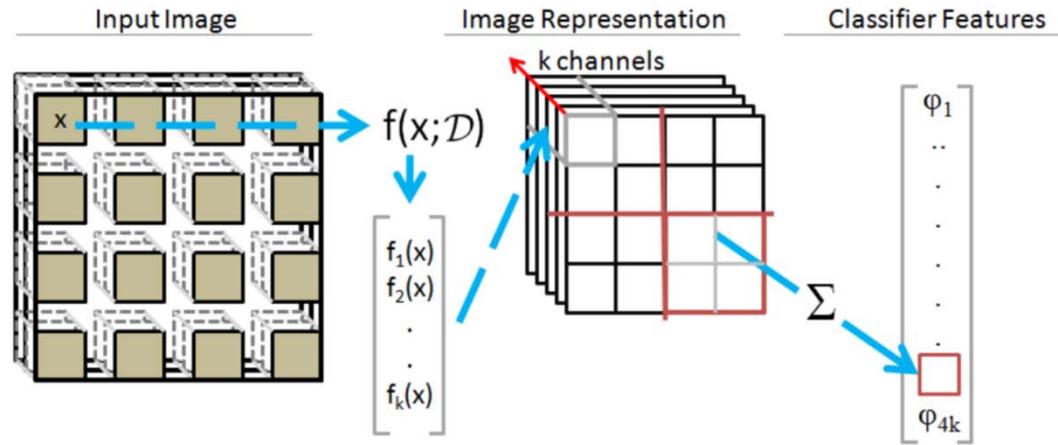


Fig. 5: A standard image recognition pipeline used in conjunction with K-means dictionary learning.

Table 1: Results on CIFAR-10 (full)

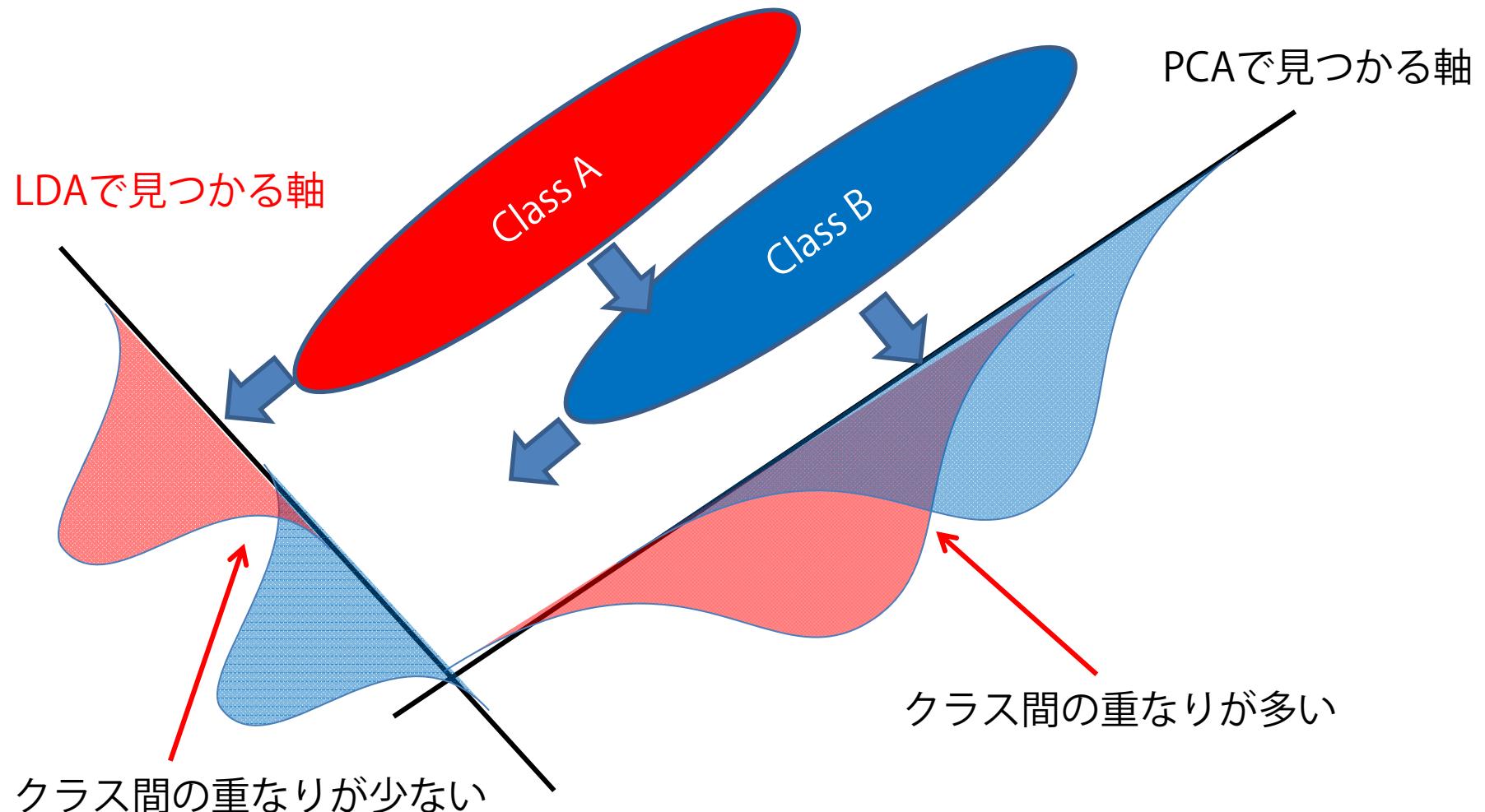
Architecture	Accuracy (%)
1 Layer	78.3%
1 Layer (4800 maps)	80.6%
2 Layers (Single RF)	77.4%
2 Layers (Random RF)	77.6%
2 Layers (Learned RF)	81.2%
3 Layers (Learned RF)	82.0%
VQ (6000 maps) [12]	81.5%
Conv. DBN [26]	78.9%
Deep NN [8]	80.49%
Multi-column Deep NN [9]	88.79%

Table 3: Classification Results on STL-10

Architecture	Accuracy (%)
1 Layer	54.5% ($\pm 0.8\%$)
1 Layer (4800 maps)	53.8% ($\pm 1.6\%$)
2 Layers (Single RF)	55.0% ($\pm 0.8\%$)
2 Layers (Random RF)	54.4% ($\pm 1.2\%$)
2 Layers (Learned RF)	58.9% ($\pm 1.1\%$)
3 Layers (Learned RF)	60.1% ($\pm 1.0\%$)
Sparse coding (1 layer) [12]	59.0% ($\pm 0.8\%$)
VQ (1 layer) [12]	54.9% ($\pm 0.4\%$)

Coates, Adam; Ng, Andrew Y. (2012). "Learning feature representations with k-means". In G. Montavon, G. B. Orr and K.-R. Müller. Neural Networks: Tricks of the Trade. Springer.

Linear Discriminant Analysis (LDA)



Linear Discriminant Analysis (LDA)

- クラス内分散に対するクラス間分散最大化
 - データ集合

$$D = \{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=1}^N, \mathbf{x}_i \in \mathbb{R}^m, \mathbf{t}_i \in \mathbb{R}^K$$

教師信号

- 射影ベクトルによる1次元空間への射影

$$z_i = \mathbf{u}^T (\mathbf{x}_i - \bar{\mathbf{x}}) \quad \text{平均 } \bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$$

- 射影された空間での平均

$$\text{クラス } k \text{ の平均 } \mu_k = \frac{1}{n_k} \sum_{i \in C_k} z_i$$

クラス k に属するサンプル数

$$\text{全データの平均 } \mu = \frac{1}{N} \sum_{k=1}^K n_k \mu_k$$

$$N = \sum_{k=1}^K n_k$$

- クラス内分散, クラス間分散

クラス内分散

$$S_W = \sum_{k=1}^K \frac{1}{n_k} \sum_{i \in C_k} (z_i - \mu_k)^2 = \mathbf{u}^T S_W \mathbf{u}$$

クラス間分散

$$S_B = \frac{1}{N} \sum_{k=1}^K n_k (\mu_k - \mu)^2 = \mathbf{u}^T S_B \mathbf{u}$$

$$S_W = \sum_{k=1}^K \frac{1}{n_k} \sum_{i \in C_k} (\mathbf{x}_i - \bar{\mathbf{x}}_k)(\mathbf{x}_i - \bar{\mathbf{x}}_k)^T$$

$$S_B = \frac{1}{N} \sum_{k=1}^K n_k (\bar{\mathbf{x}}_k - \bar{\mathbf{x}})(\bar{\mathbf{x}}_k - \bar{\mathbf{x}})^T$$

$$S_T = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T = S_B + S_W$$

Linear Discriminant Analysis (LDA)

- クラス内分散に対するクラス間分散最大化
 - フィッシャーの評価基準

$$J_s = \frac{\mathbf{u}^T S_B \mathbf{u}}{\mathbf{u}^T S_W \mathbf{u}}$$

クラス内分散とクラス間分散の比
u を定数倍しても
 J_s の値は変わらない

- フィッシャーの評価基準の最大化

$$J = \mathbf{u}^T S_B \mathbf{u} - \lambda (\mathbf{u}^T S_W \mathbf{u} - 1)$$

$$\frac{\partial J}{\partial \mathbf{u}} = 0 \rightarrow S_B \mathbf{u} = \lambda S_W \mathbf{u}$$

一般化固有値問題

$$U = (\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_M)$$

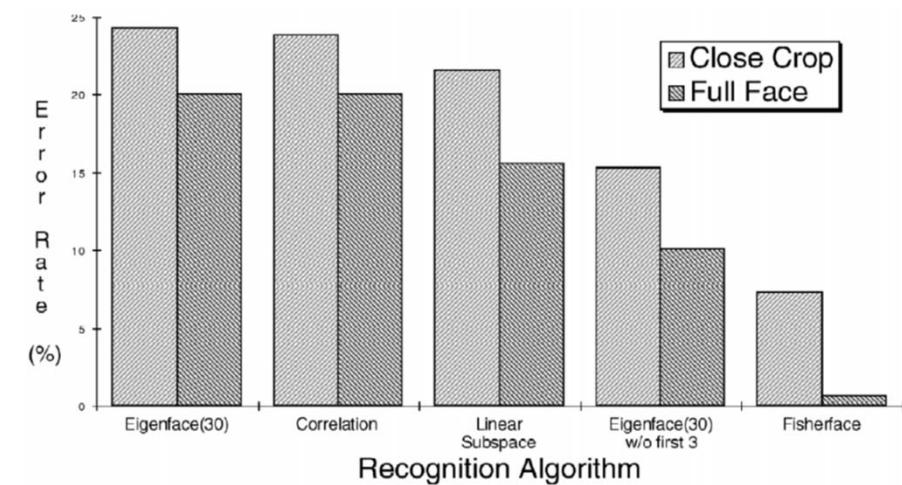
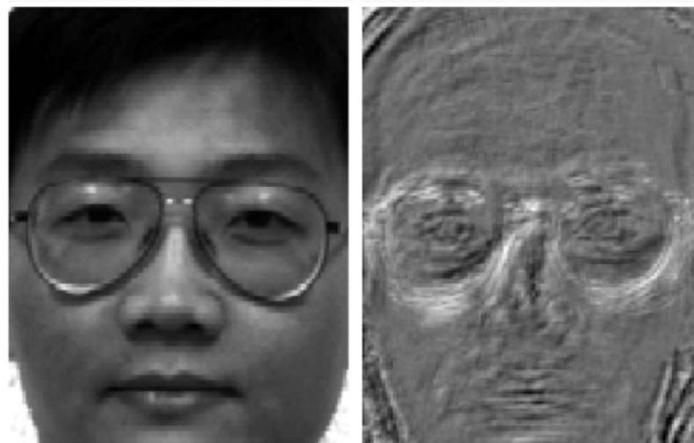
$$S_B U = S_W U \Lambda$$
$$\Lambda = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_M \end{pmatrix} \quad M \leq \max \{K-1, m\}$$

Fisherface

Peter N. Belhumeur, Joao P. Hespanha, and David J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. TPAMI, vol. 19, no. 7, pp.711-720, 1997.



Yale database



"Leaving-One-Out" of Yale Database			
Method	Reduced Space	Error Rate (%)	
		Close Crop	Full Face
Eigenface	30	24.4	19.4
Eigenface w/o 1st 3	30	15.3	10.8
Correlation	160	23.9	20.0
Linear Subspace	48	21.6	15.6
Fisherface	15	7.3	0.6

Fig. 10. The left image is an image from the Yale Database of a person wearing glasses. The right image is the Fisherface used for determining if a person is wearing glasses.

PCAとLDAの関係

- クラス内共分散行列の白色化（正規化）

$${U_1}^T S_W U_1 = I_m \quad \begin{array}{l} \text{クラス内共分散行列が平均的} \\ \text{に単位行列となるように変換} \end{array}$$

- 正規化空間におけるクラス間共分散行列

$${U_1}^T S_B U_1 \quad \begin{array}{l} \text{非負正定値対象行列} \\ S_B = \frac{1}{N} \sum_{k=1}^K n_k (\bar{\mathbf{x}}_k - \bar{\mathbf{x}})(\bar{\mathbf{x}}_k - \bar{\mathbf{x}})^T \end{array}$$

- 固有値問題を解く

$$(U_1^T S_B U_1) U_2 = U_2 \Lambda \quad \Lambda = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_M \end{pmatrix}$$
$$U_2^T U_2 = I_M$$

正規化された空間における
クラス内平均の主成分分析

$$U_2 = (\mathbf{u}_1^{(2)} \ \mathbf{u}_2^{(2)} \cdots \mathbf{u}_M^{(2)})$$

PCAとLDAの関係

- LDA?

$$S_B U = S_W U \Lambda$$

クラス間共分散行列と
クラス内共分散行列の
一般化固有値問題

- クラス内共分散行列の白色化

$$U_1^T S_W U_1 = I_m \quad \Rightarrow \quad (U_1^T)^{-1} = S_W U_1$$

- 正規化空間におけるクラス内平均の主成分分析

$$(U_1^T)^{-1} = S_W U_1$$

$$(U_1^T)^{-1} = S_W U_1$$

$$(U_1^T S_B U_1) U_2 = U_2 \Lambda$$

$$\Rightarrow (U_1^T)^{-1} (U_1^T S_B U_1) U_2 = S_W U_1 U_2 \Lambda$$

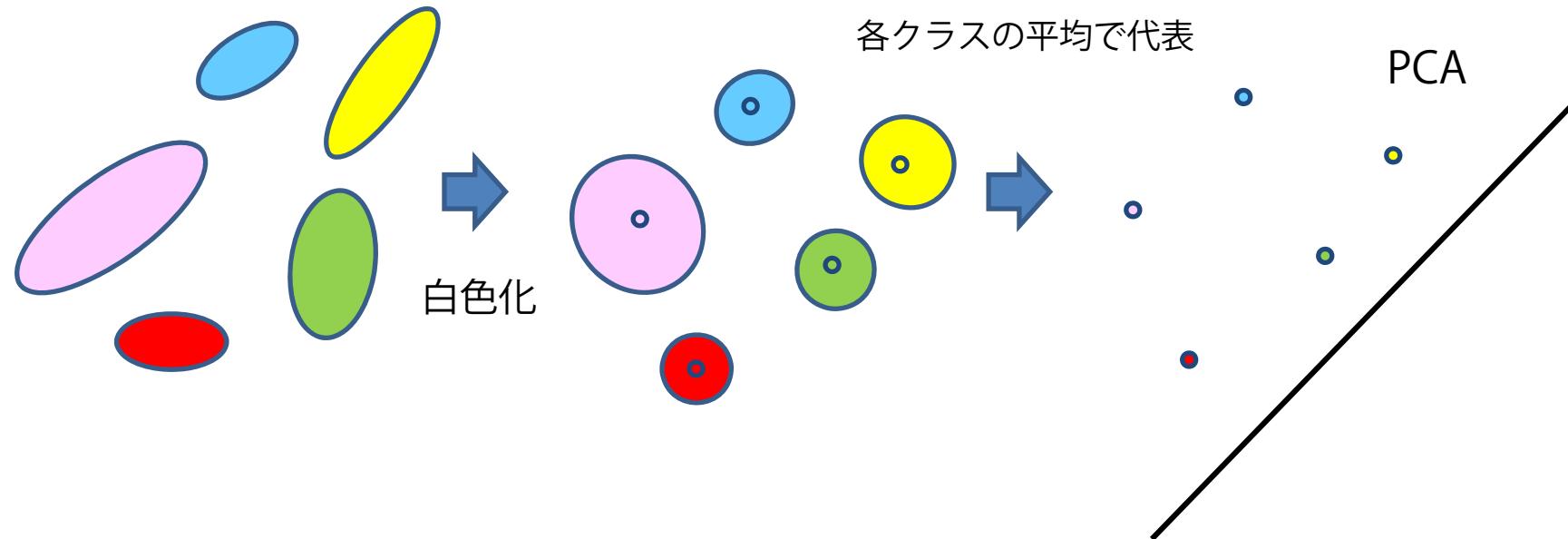
$$\Rightarrow S_B U_1 U_2 = S_W U_1 U_2 \Lambda$$

$$\Rightarrow S_B U = S_W U \Lambda \quad \text{LDAと一致!}$$

$U = U_1 U_2$ とおく

PCAとLDAの関係

- LDAのPCAの観点からの解釈
 - クラス内共分散行列が平均的に単位行列となるように変換
 - クラス内共分散行列の白色化（PCAの利用）
 - 各クラスを各クラスの平均で代表
 - 正規化された空間におけるクラス内平均の主成分分析（PCA）



CCAとLDAの関係

1-of-K表記法
クラスCjに属する時, j番目の要素を1, その他を0とする

- データ集合

$$D = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N, \quad \mathbf{x}_i \in \mathbb{R}^m, \mathbf{y}_i \in \mathbb{R}^K$$

$$(0, \dots, 1, \dots, 0)^T$$

↓ クラス数

$$X = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_N^T \end{pmatrix} \quad X \in \mathbb{R}^{N \times m} \quad Y = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \cdots & \mathbf{0}_{n_2} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0}_{n_K} & \mathbf{0}_{n_K} & \cdots & \mathbf{1}_{n_K} \end{pmatrix} \quad Y \in \mathbb{R}^{N \times K} \quad \mathbf{1}_{n_j} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

x を並べた行列

データをC1, C2, ..., CKに順番に属するようにソート

- 平均

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i = \frac{1}{N} X^T \mathbf{1}_N \quad \bar{\mathbf{y}} = \frac{1}{N} \sum_{i=1}^N \mathbf{y}_i = \frac{1}{N} (n_1 \cdots n_K) = \frac{1}{N} Y^T \mathbf{1}_N$$

- Xの共分散行列

$$C_{xx} = \frac{1}{N} X^T X - \bar{\mathbf{x}} \bar{\mathbf{x}}^T = \frac{1}{N} X^T X - \left(\frac{1}{N} X^T \mathbf{1}_N \right) \left(\frac{1}{N} X^T \mathbf{1}_N \right)^T$$

分散 = 二乗平均 - 平均の2乗

$$= \frac{1}{N} X^T P_c X \quad P_c = \left(I_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^T \right)$$

CCAとLDAの関係

- Xの共分散行列

$$\begin{aligned}
 C_{xx} &= \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T = \sum_{k=1}^K \frac{n_k}{N} \frac{1}{n_k} \sum_{i \in C_k} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \\
 &= \sum_{k=1}^K \frac{n_k}{N} \frac{1}{n_k} \sum_{i \in C_k} (\mathbf{x}_i - \bar{\mathbf{x}}_k + \bar{\mathbf{x}}_k - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}}_k + \bar{\mathbf{x}}_k - \bar{\mathbf{x}})^T \\
 &= \sum_{k=1}^K \frac{n_k}{N} \frac{1}{n_k} \sum_{i \in C_k} \left\{ (\mathbf{x}_i - \bar{\mathbf{x}}_k)(\mathbf{x}_i - \bar{\mathbf{x}}_k)^T + (\bar{\mathbf{x}}_k - \bar{\mathbf{x}})(\bar{\mathbf{x}}_k - \bar{\mathbf{x}})^T + (\mathbf{x}_i - \bar{\mathbf{x}}_k)(\bar{\mathbf{x}}_k - \bar{\mathbf{x}})^T + (\bar{\mathbf{x}}_k - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}}_k)^T \right\} \\
 &= \sum_{k=1}^K \frac{n_k}{N} \frac{1}{n_k} \sum_{i \in C_k} (\mathbf{x}_i - \bar{\mathbf{x}}_k)(\mathbf{x}_i - \bar{\mathbf{x}}_k)^T + \sum_{k=1}^K \frac{n_k}{N} \frac{1}{n_k} \sum_{i \in C_k} (\bar{\mathbf{x}}_k - \bar{\mathbf{x}})(\bar{\mathbf{x}}_k - \bar{\mathbf{x}})^T + \sum_{k=1}^K \frac{n_k}{N} \frac{1}{n_k} \sum_{i \in C_k} \left\{ (\mathbf{x}_i - \bar{\mathbf{x}}_k)(\bar{\mathbf{x}}_k - \bar{\mathbf{x}})^T + (\bar{\mathbf{x}}_k - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}}_k)^T \right\} \\
 &= \sum_{k=1}^K \frac{n_k}{N} S_k + \sum_{k=1}^K \frac{n_k}{N} \frac{1}{n_k} n_k (\bar{\mathbf{x}}_k - \bar{\mathbf{x}})(\bar{\mathbf{x}}_k - \bar{\mathbf{x}})^T + \sum_{k=1}^K \frac{n_k}{N} \left\{ \sum_{i \in C_k} \frac{1}{n_k} (\mathbf{x}_i - \bar{\mathbf{x}}_k) \right\} (\bar{\mathbf{x}}_k - \bar{\mathbf{x}})^T + \sum_{k=1}^K \frac{n_k}{N} (\bar{\mathbf{x}}_k - \bar{\mathbf{x}}) \left\{ \sum_{i \in C_k} \frac{1}{n_k} (\mathbf{x}_i - \bar{\mathbf{x}}_k) \right\}^T \\
 &\quad S_k = \frac{1}{n_k} \sum_{i \in C_k} (\mathbf{x}_i - \bar{\mathbf{x}}_k)(\mathbf{x}_i - \bar{\mathbf{x}}_k)^T \\
 &= S_W + S_B + \sum_{k=1}^K \frac{n_k}{N} (\bar{\mathbf{x}}_k - \bar{\mathbf{x}})(\bar{\mathbf{x}}_k - \bar{\mathbf{x}})^T + \sum_{k=1}^K \frac{n_k}{N} (\bar{\mathbf{x}}_k - \bar{\mathbf{x}})(\bar{\mathbf{x}}_k - \bar{\mathbf{x}})^T \\
 S_W &= \frac{1}{N} \sum_{k=1}^K n_k S_k \quad S_B = \frac{1}{N} \sum_{k=1}^K n_k (\bar{\mathbf{x}}_k - \bar{\mathbf{x}})(\bar{\mathbf{x}}_k - \bar{\mathbf{x}})^T \\
 &= S_W + S_B
 \end{aligned}$$

CCAとLDAの関係

- Yの共分散行列

$$C_{yy} = \frac{1}{N} Y^T Y - \bar{\mathbf{y}}\bar{\mathbf{y}}^T = \frac{1}{N} \begin{pmatrix} n_1 & & 0 \\ & \ddots & \\ 0 & & n_K \end{pmatrix} - \frac{1}{N} \begin{pmatrix} n_1 \\ \vdots \\ n_K \end{pmatrix} \frac{1}{N} (n_1 \quad \cdots \quad n_K) = \text{diag}(\bar{\mathbf{y}}) - \bar{\mathbf{y}}\bar{\mathbf{y}}^T$$

- XとYの共分散行列

$$\begin{aligned} C_{xy} &= \frac{1}{N} X^T Y - \bar{\mathbf{x}}\bar{\mathbf{y}}^T = \frac{1}{N} (\mathbf{x}_1 \cdots \mathbf{x}_N) \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & & \mathbf{0}_{n_2} \\ \vdots & & \ddots & \vdots \\ \mathbf{0}_{n_K} & \mathbf{0}_{n_K} & \cdots & \mathbf{1}_{n_K} \end{pmatrix} - \bar{\mathbf{x}}\bar{\mathbf{y}}^T \\ &= \frac{1}{N} (n_1 \bar{\mathbf{x}}_1 \cdots n_K \bar{\mathbf{x}}_K) - \frac{1}{N} (n_1 \bar{\mathbf{x}}_1 + \cdots + n_K \bar{\mathbf{x}}_K) \bar{\mathbf{y}}^T \\ &= (\bar{\mathbf{x}}_1 \cdots \bar{\mathbf{x}}_K) \frac{1}{N} \begin{pmatrix} n_1 & & 0 \\ & \ddots & \\ 0 & & n_K \end{pmatrix} - (\bar{\mathbf{x}}_1 \cdots \bar{\mathbf{x}}_K) \frac{1}{N} \begin{pmatrix} n_1 \\ \vdots \\ n_K \end{pmatrix} \bar{\mathbf{y}}^T \\ &= M \left(\text{diag}(\bar{\mathbf{y}}) - \bar{\mathbf{y}}\bar{\mathbf{y}}^T \right) \quad M = (\bar{\mathbf{x}}_1 \cdots \bar{\mathbf{x}}_K) \end{aligned}$$

↑
 $\bar{\mathbf{x}} = \frac{1}{N} (n_1 \bar{\mathbf{x}}_1 + \cdots + n_K \bar{\mathbf{x}}_K)$

CCAとLDAの関係

- CCA? $C_{xy}C_{yy}^{-1}C_{yx}\mathbf{u} = \rho^2 C_{xx}\mathbf{u}$ \rightarrow 一般化固有値問題

$$C_{yx}C_{xx}^{-1}C_{xy}\mathbf{v} = \rho^2 C_{yy}\mathbf{v}$$

$$\begin{aligned} & C_{xy}C_{yy}^{-1}C_{yx}\mathbf{u} = \rho^2 C_{xx}\mathbf{u} \\ & M\left(\text{diag}(\bar{\mathbf{y}}) - \bar{\mathbf{y}}\bar{\mathbf{y}}^T\right) \quad \left\{M\left(\text{diag}(\bar{\mathbf{y}}) - \bar{\mathbf{y}}\bar{\mathbf{y}}^T\right)\right\}^T \\ & \text{diag}(\bar{\mathbf{y}}) - \bar{\mathbf{y}}\bar{\mathbf{y}}^T \end{aligned}$$

$$M\left(\text{diag}(\bar{\mathbf{y}}) - \bar{\mathbf{y}}\bar{\mathbf{y}}^T\right)\left(\text{diag}(\bar{\mathbf{y}}) - \bar{\mathbf{y}}\bar{\mathbf{y}}^T\right)^{-1}\left(\text{diag}(\bar{\mathbf{y}}) - \bar{\mathbf{y}}\bar{\mathbf{y}}^T\right)M^T\mathbf{u} = \rho^2(S_W + S_B)\mathbf{u}$$

$$\rightarrow M\left(\text{diag}(\bar{\mathbf{y}}) - \bar{\mathbf{y}}\bar{\mathbf{y}}^T\right)M^T\mathbf{u} = \rho^2(S_W + S_B)\mathbf{u}$$

$$\rightarrow S_B\mathbf{u} = \rho^2(S_W + S_B)\mathbf{u}$$

$$\rightarrow S_B\mathbf{u} = \frac{\rho^2}{1-\rho^2}S_W\mathbf{u}$$

\rightarrow LDA!

$$\begin{aligned} S_B &= \frac{1}{N} \sum_{k=1}^K n_k (\bar{\mathbf{x}}_k - \bar{\mathbf{x}})(\bar{\mathbf{x}}_k - \bar{\mathbf{x}})^T \\ &= \frac{1}{N} \sum_{k=1}^K n_k (\bar{\mathbf{x}}_k \bar{\mathbf{x}}_k^T + \bar{\mathbf{x}} \bar{\mathbf{x}}^T - \bar{\mathbf{x}}_k \bar{\mathbf{x}}^T - \bar{\mathbf{x}} \bar{\mathbf{x}}_k^T) \\ &= \frac{1}{N} \sum_{k=1}^K n_k \bar{\mathbf{x}}_k \bar{\mathbf{x}}_k^T + \frac{1}{N} \sum_{k=1}^K n_k \bar{\mathbf{x}} \bar{\mathbf{x}}^T - \frac{1}{N} \sum_{k=1}^K n_k \bar{\mathbf{x}}_k \bar{\mathbf{x}}^T - \frac{1}{N} \sum_{k=1}^K n_k \bar{\mathbf{x}} \bar{\mathbf{x}}_k^T \\ &= \frac{1}{N} \sum_{k=1}^K n_k \bar{\mathbf{x}}_k \bar{\mathbf{x}}_k^T - \bar{\mathbf{x}} \bar{\mathbf{x}}^T \\ &= M\left(\text{diag}(\bar{\mathbf{y}}) - \bar{\mathbf{y}}\bar{\mathbf{y}}^T\right)M^T \end{aligned}$$

$$\begin{aligned} \bar{\mathbf{x}} &= \frac{1}{N}(n_1 \bar{\mathbf{x}}_1 + \dots + n_K \bar{\mathbf{x}}_K) \\ &= M\bar{\mathbf{y}} \end{aligned}$$

統計的特徴抽出のまとめ

- Canonical Correlation Analysis (CCA)

$$\begin{pmatrix} 0 & C_{xy} \\ C_{xy} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}$$

- Partial Least Squares (PLS)

$$\begin{pmatrix} 0 & C_{xy} \\ C_{xy} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}$$

- Principal Component Analysis (PCA)

$$C_{xx} \mathbf{u} = \lambda \mathbf{u}$$

- Linear Discriminant Analysis (LDA)

$$\begin{pmatrix} 0 & C_{xy} \\ C_{xy} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix}$$

← $Y = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & & \mathbf{0}_{n_2} \\ \vdots & & \ddots & \vdots \\ \mathbf{0}_{n_K} & \mathbf{0}_{n_K} & \cdots & \mathbf{1}_{n_K} \end{pmatrix} \quad Y \in R^{N \times K}$

統計的特徴抽出のまとめ

