具体例に基づく画像の幾何学的不変性の実現

堀田政二 (東京農工大学)
@seiji_hotta

第 59 回 STARC アドバンストセミナー

2014 年 9 月 19 日
URL は 2014 年 8 月 22 日に閲覧済み

Object Recognition (物体認識) とは何か

Object Recognition

実環境で撮影された画像（映像）に含まれる物体の名称を計算機が推定し出力すること

image　　objects　　class

©Visual Object Classes Challenge [1]

- **汎用的過ぎるので制約を加えて簡略化するのが一般的** [2, 3]
- **映像認識にも多くの共通点がある**

Object Recognition の種類

1. verification (**物体照合**)
   画像中のある物体に着目し，それが対象物体のクラスであるかを照合する問題
2. detection (**物体検出**)
   顔検出などの所望の物体が画像中のどこにあるか (localization) **を答える問題**
3. identification (**特定物体認識**)
   画像中の物体の固有名詞を答える問題，入力と同じものを探す問題
4. object categorization (**画像分類**)
   画像中の物体をクラスに分類する問題
5. scene and context categorization (**シーン認識**)
   場所や天気などの画像が表す状況を認識する問題

Detection の例

顔検出の例

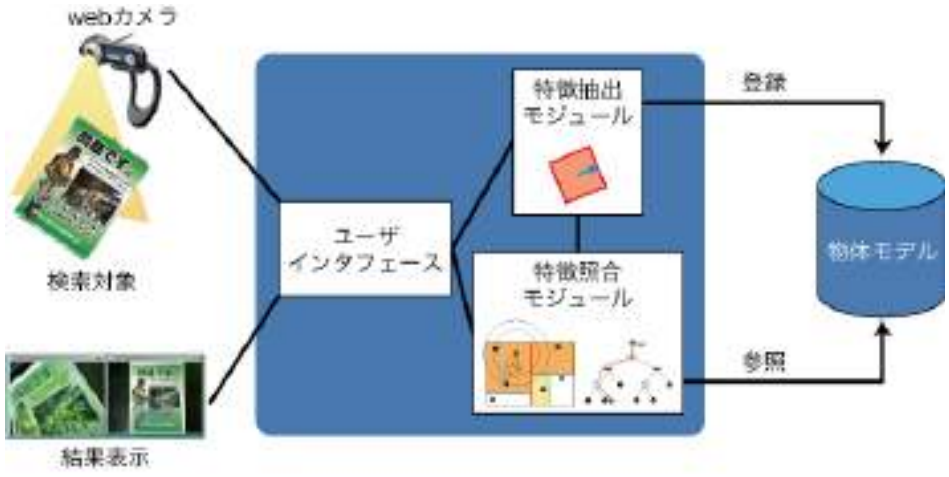
OpenCV [4] のサンプルプログラムを利用して任意の検出器が作成可能

OpenCV で学ぶ画像認識

http://gihyo.jp/dev/feature/01/opencv

identification (特定物体認識) の例

カメラで撮影した画像と同じ画像，もしくは非常に類似した画像をデータベースから高速に検索する



ⓒ3 日で作る高速特定物体認識システム (岩村，黄瀬)

3 日で作る高速特定物体認識システムのサイト

http://www.m.cs.osakafu-u.ac.jp/IPSJ_3days/

Object Categorization (Classification)

dog
chair

man
bike

airplane

- **画像中の物体** (objects) **の一般名称（**category, class**）を答える**
- **物体は一つとは限らないし，ある物体はさらにサブクラス** (subclass) **に分けられるかもしれない (車のクラスに属するものはバスや軽自動車などのサブクラスに分けられる)**

Segmentation を併用した Object Categorization の例

画像の領域分割 (segmentation) **と領域間の関係** (context) **を利用した物体認識** [5]

Jamie Shotton

`http://jamie.shotton.org/work/`

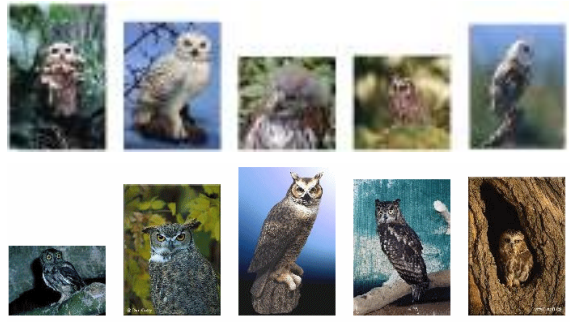
Object Recognition は何が難しいのか？

画像の見えの変動に対する不変性の必要性

アフィン変換 (Affine transform) **などの画像の見え方** (appearance) **の変動に対する不変性** (invariance) **を必要とする**

- **拡大縮小** (scale)，**回転** (rotation)，**平行移動** (translation)
- **照明変動** (illumination variation)，**隠れ** (occlusion)

同じクラスに属していても種の違いにより見えが変化する $\rightarrow$ **詳細画像識別** [6]



種の違いの例

Viewpoint の違いに基づく見えの変化 [7]

認識対象のクラス数が多い

©Caltech **データセット** [8]

認識対象となるクラス数が多い (人間は数万クラスを分類できると言われている)

代表的なデータベース・タスク

- TREC Video Retrieval Evaluation: TRECVID http://trecvid.nist.gov/
- Caltech 256 http://www.vision.caltech.edu/Image_Datasets/Caltech256/
- ImageNet (1400 **万画像，**22k **クラス**): http://www.image-net.org/
- Large Scale Visual Recognition Challenge 2014 (ILSVRC2014)
  http://www.image-net.org/challenges/LSVRC/2014/
- Tiny Image Dataset (8000 **万画像．ノイズ多し**)
  http://horatio.cs.nyu.edu/mit/tiny/data/index.html
- CIFAR-10 and CIFAR-100 (tiny image **から作った綺麗なデータ**)
  http://www.cs.utoronto.ca/~kriz/cifar.html
- SUN dataset (**シーン認識**) http://people.csail.mit.edu/jxiao/SUN/
- The Chars74K dataset (**環境文字データ**)
  http://www.ee.surrey.ac.uk/CVSSP/demos/chars74k/

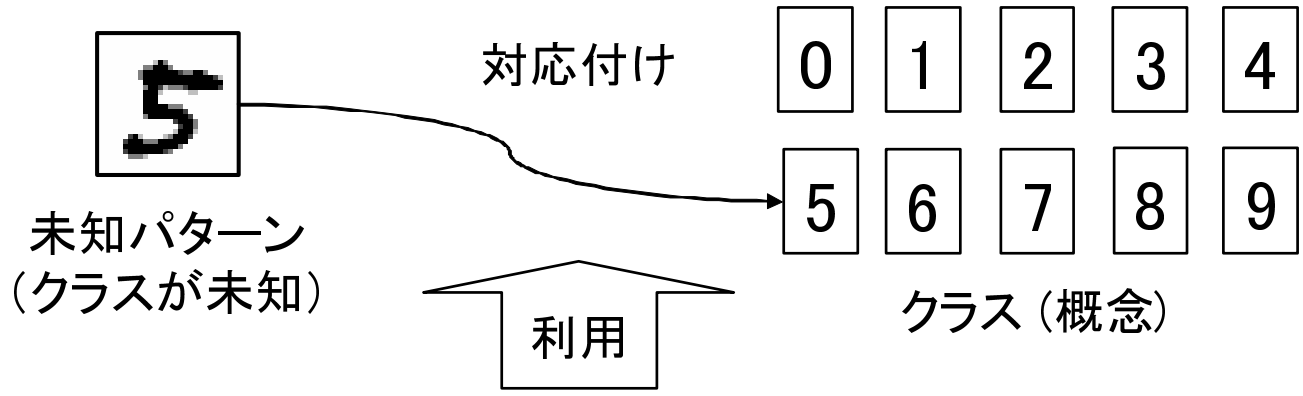
Object Recognition の発展

- 70 年代：線画解釈（画像処理が中心）
- 80 年代前半：知識ベース型システム
- 80 年代後半：3 次元の復元， モデルベース
- 90 年代：顔認識，固有空間法
- 90 年代後半：局所特徴量 (local feature) と機械学習 (machine learning) の登場
- 200X 年代：データベースの充実，Bag of Features の登場
- 2010 年前後：Big Data 時代の到来，Super Vector，Deep learning の登場

特徴量 · 機械学習の発展，およびデータベース構築が Object Recognition の研究の発展に大きく貢献している

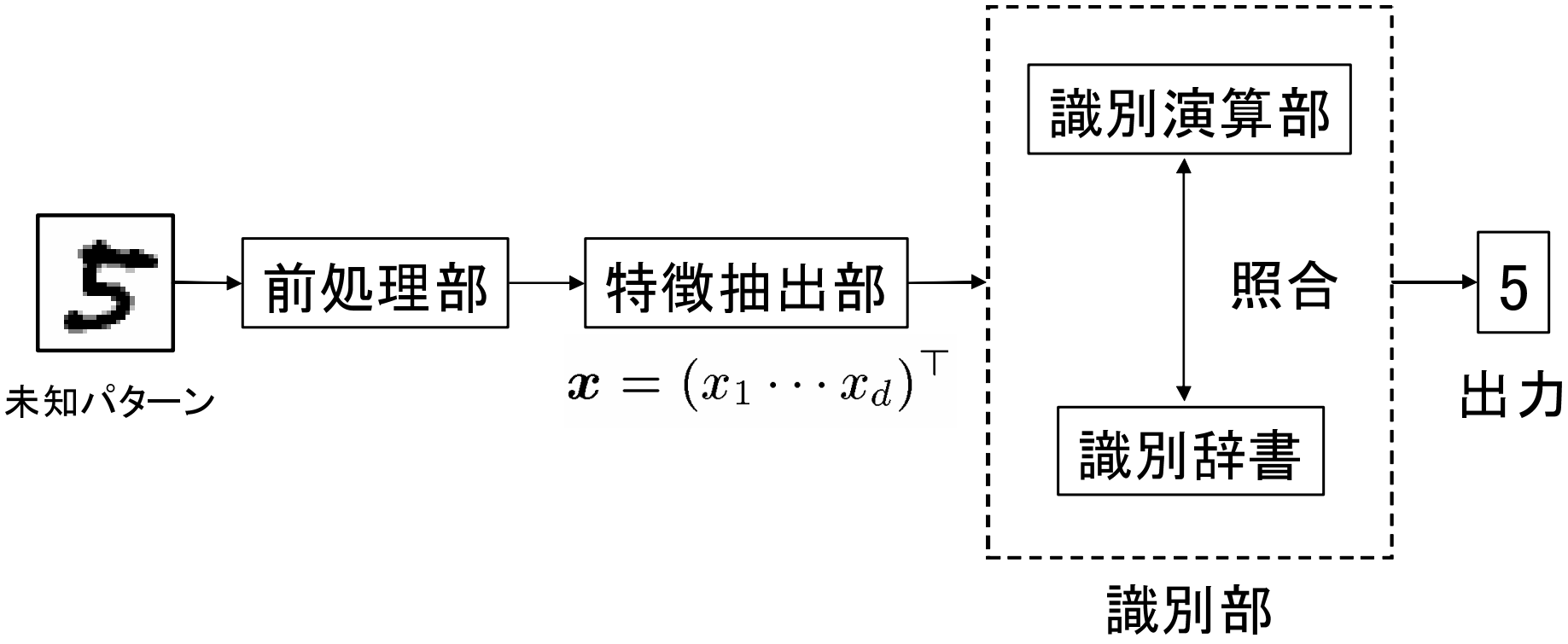
画像・映像認識は基本的にはパターン認識で解く

観測されたパターンを予め定められた複数の概念のうちの一つに対応させる (識別する) 処理 [9]



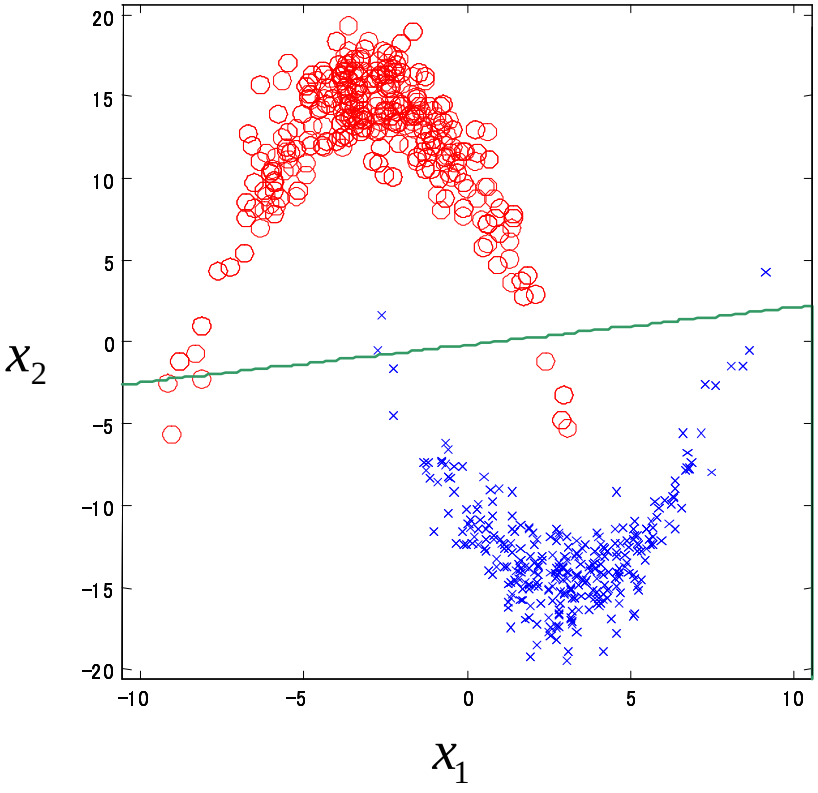
訓練パターン
(クラスが既知)

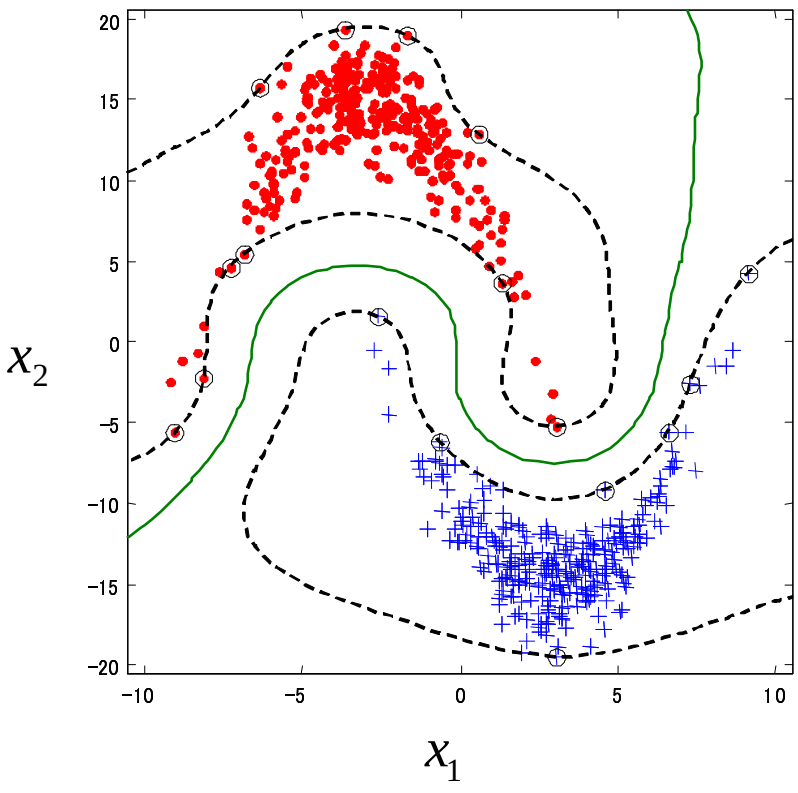
パターン認識の工学モデル [9]



- **本当にこのモデルで良いか，という疑問はある**

識別部の目的





- 未知パターンを精度良く識別できる (汎化能力の高い) 識別境界を引くこと

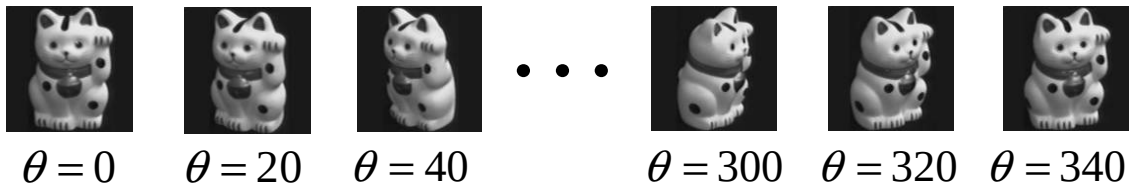
理想的な識別方法

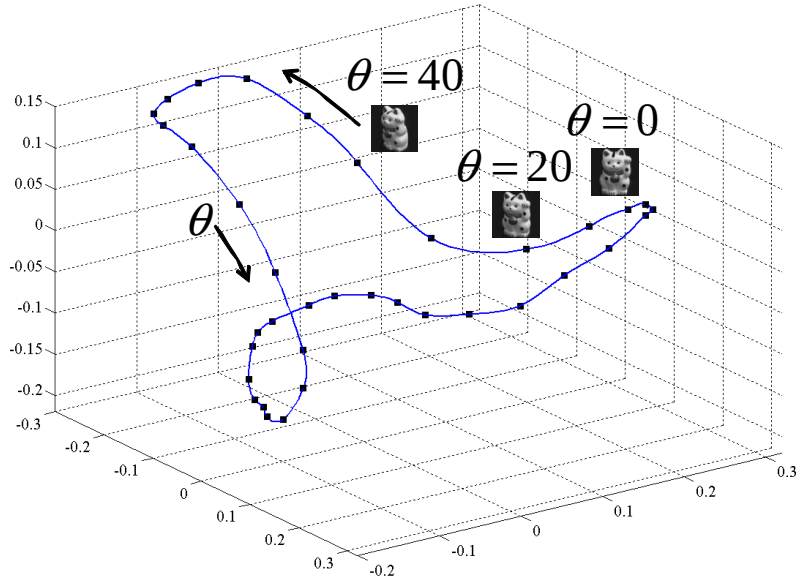
ベイズ決定則 (Bayes decision rule)

$$\max_j P(\omega_j|\boldsymbol{x}) = P(\omega_k|\boldsymbol{x}) \Rightarrow \boldsymbol{x} \in \omega_k \text{ ここで } P(\omega_j|\boldsymbol{x}) = \frac{P(\omega_j)p(\boldsymbol{x}|\omega_j)}{\sum_j P(\omega_j)p(\boldsymbol{x}|\omega_j)}$$

- $P(\omega_j)$: パターンを観測する前のクラス ω_j の生起確率 (事前確率)
- $P(\omega_j|\boldsymbol{x})$: パターン $\boldsymbol{x}$ を観測した後の ω_j の生起確率 (事後確率)
- $p(\boldsymbol{x}|\omega_j)$: ω_j における $\boldsymbol{x}$ の分布 (確率密度関数)
- $p(\boldsymbol{x}|\omega_j)$ を精度良く推定することが難しい (大量データが得られれば別)

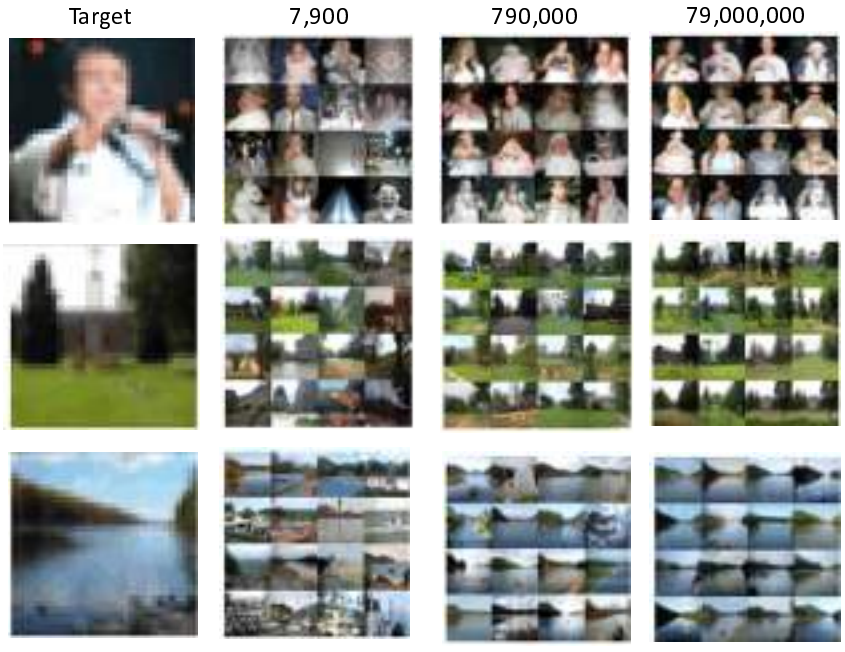
画像・映像認識で特に問題となる工学的課題





- 幾何学的変動は画素空間では非線形な多様体をなす [10]
- このような変動を吸収したり，多様体を近似したりするための特徴量や識別法の開発が必要

大量画像を用いる



A. Torralba, et al. ``80 million tiny images: A large dataset for non-parametric object,'' PAMI, 2008.

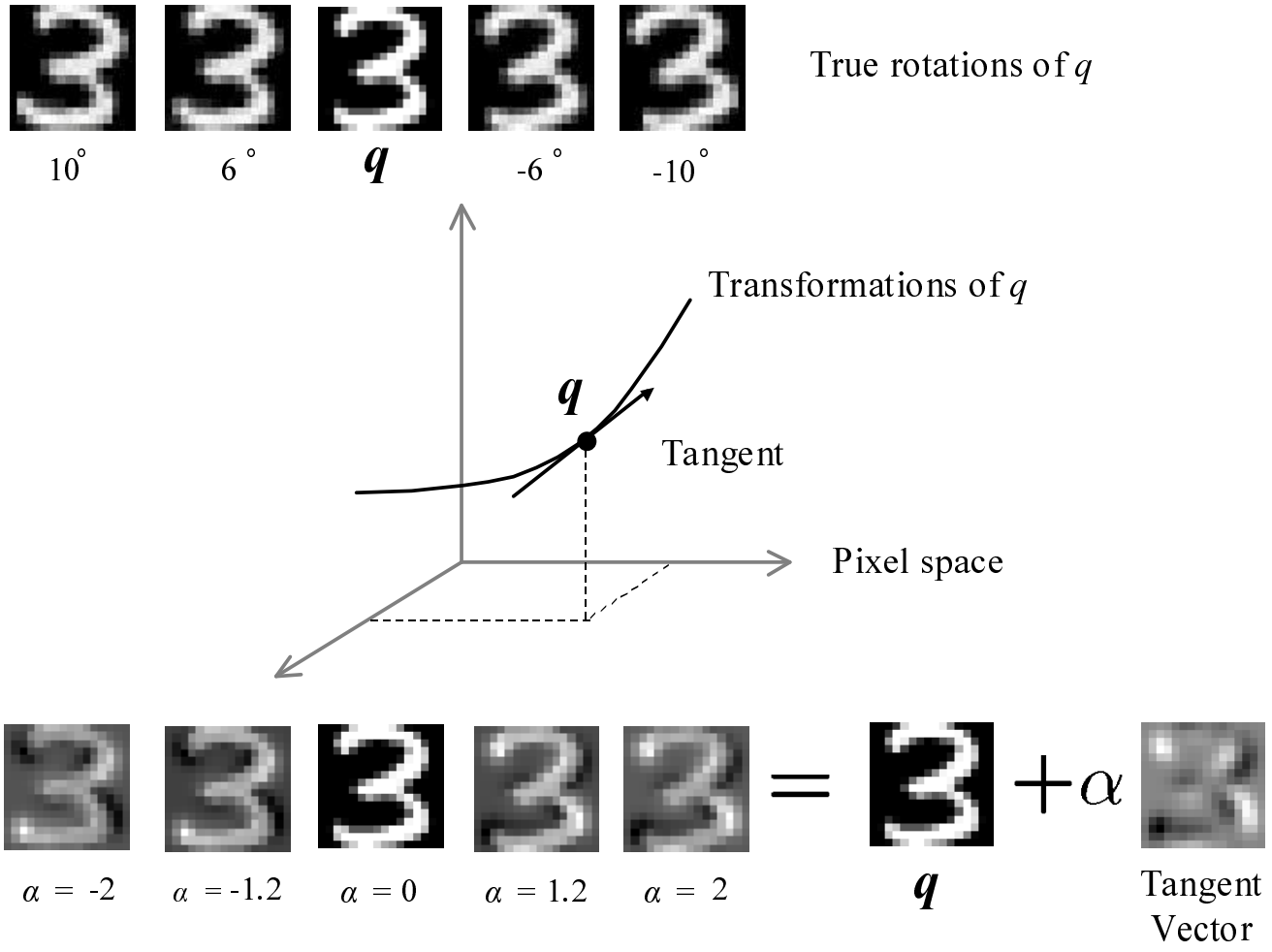
- 32×32 の 8000 万枚のカラー画像を使用 [11]
- k-nearest neighbor で認識．物体検出も可能
- More data beats better algorithms

32×32 ピクセルの画像例

- 32×32 ピクセルでも gist を認識できる
- 一方で物体を独立させると認識できない

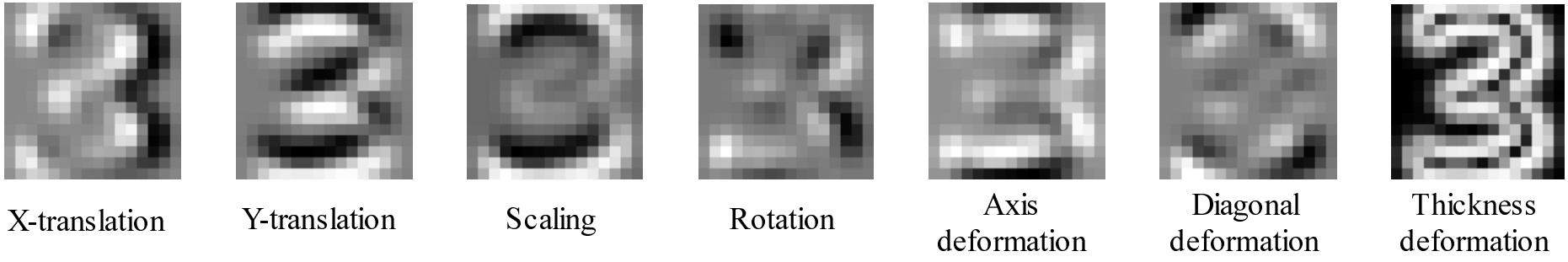
変形パターンを用いる

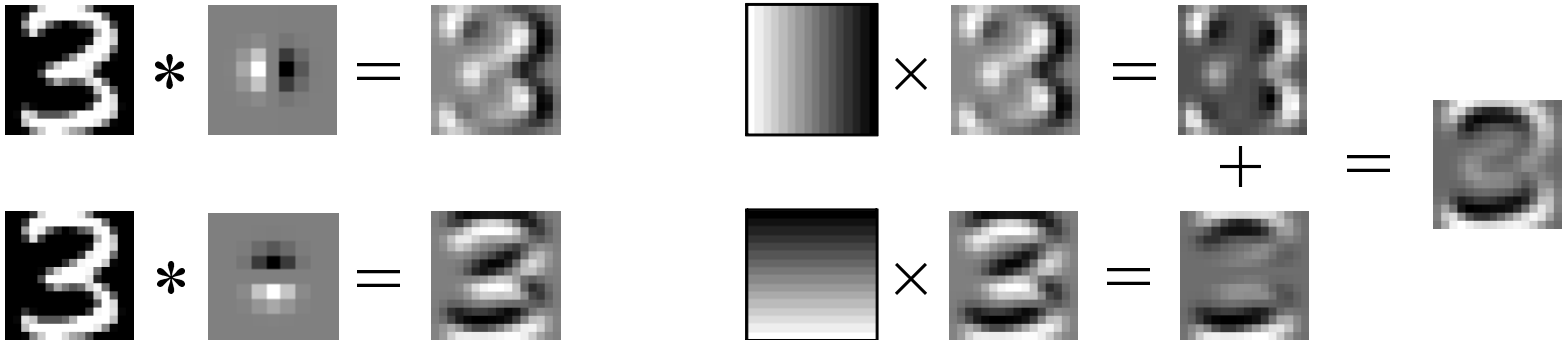
Tangent Vector による変形の近似 [12]



- **非線形多様体を線型近似するパターンを** Tangent Vector **と呼ぶ**

Tangent Vector の種類 (接ベクトル)

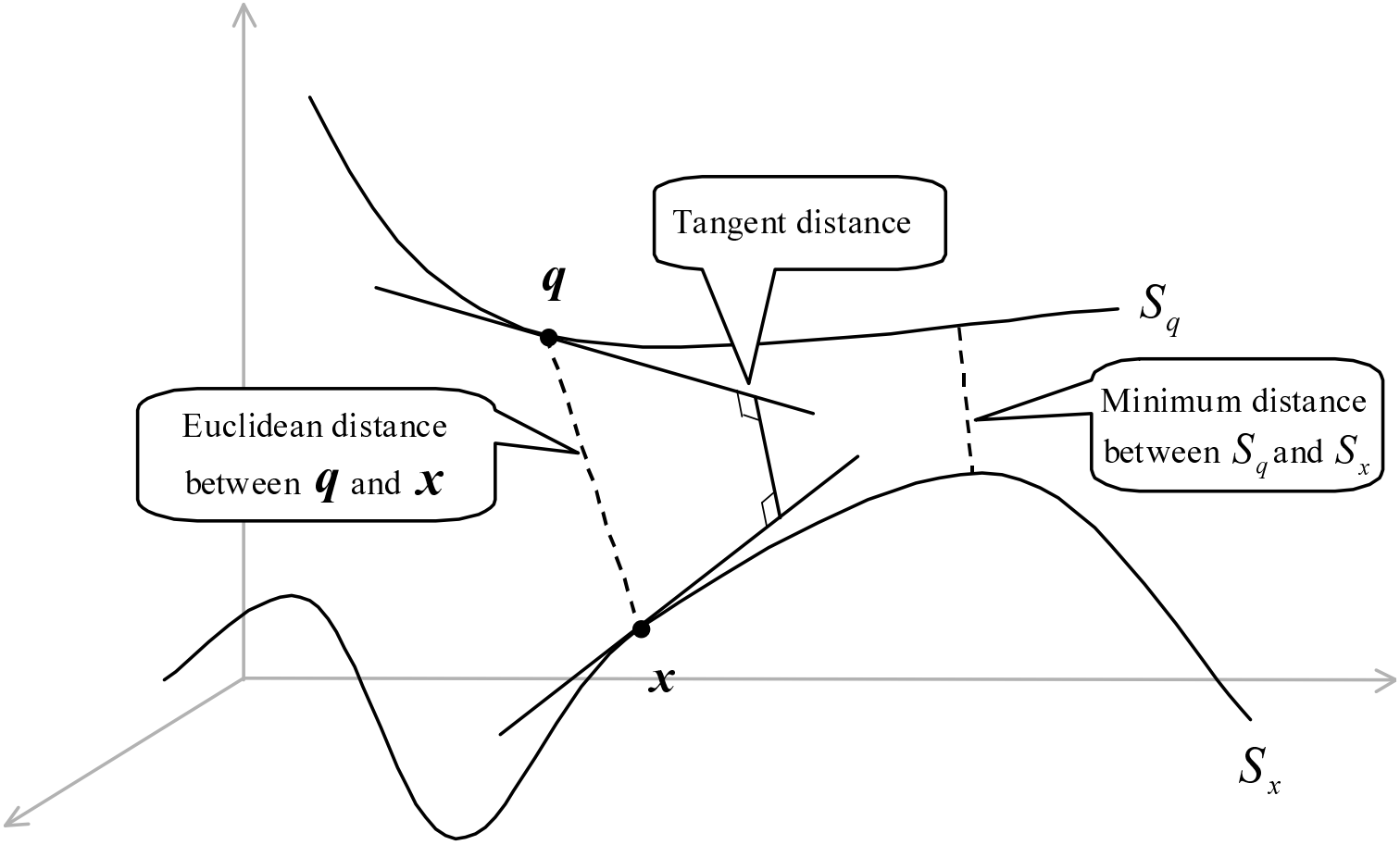




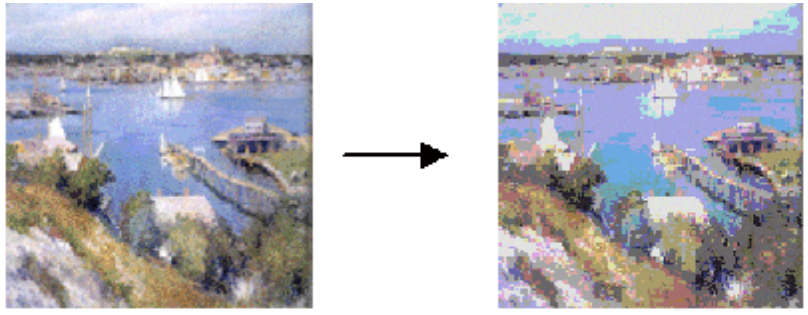
- **回転，移動，スケール，ひずみ等の変形を再現できる**

Tangent Distance

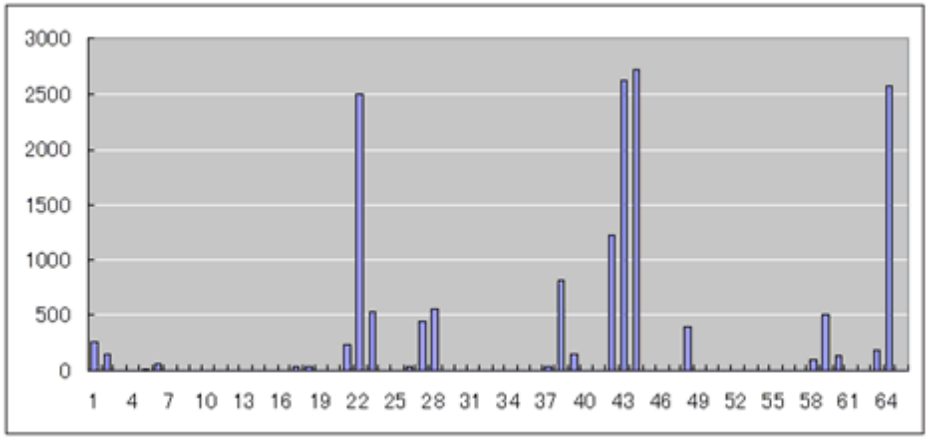
Tangent Vector **による変形したパターン同士の最短距離は解析的に求めることができる** [12]



カラーヒストグラム

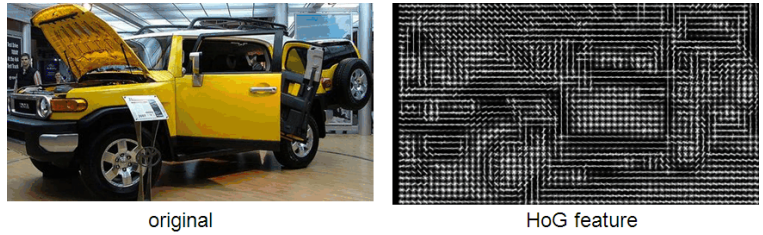


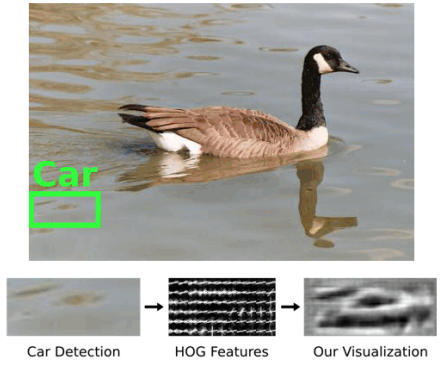
元画像　　減色画像



- もっとも典型的なもの．単純な追跡・検索に良く使われる
- 上下左右の回転に不変．規格化すれば大きさにも不変
- 位置・形状情報は失われる

エッジ特徴量

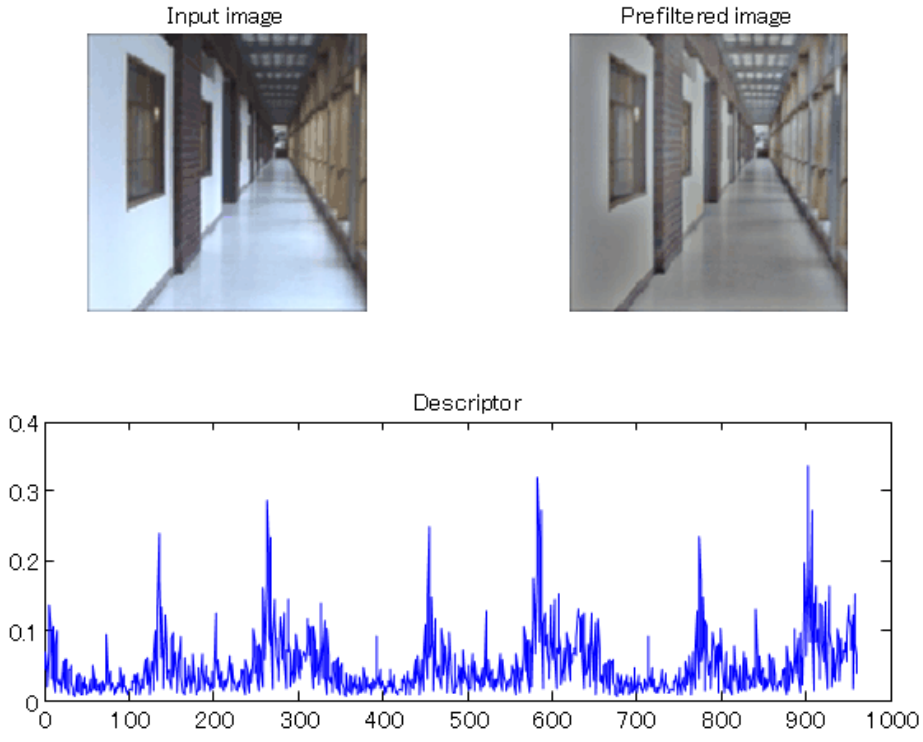




HOGgles [13] **より転載**

- **方向寄与度特徴** [14]，HoG [15]，Pyramid HoG (PHoG) [16] **などが有名**
- **位置にある程度不変．規格化すれば大きさにも不変**
- **人間にとって明らかに異なる物体でも特徴的には類似する場合がある**

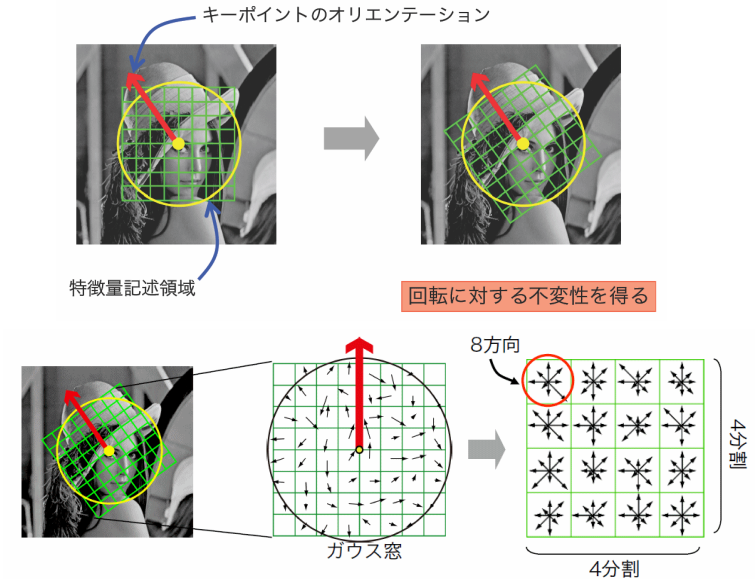
テクスチャ特徴量



- HLAC [17]，Local Binary Pattern (LBP) [18]，GIST [19] **などが有名**
- **大域・局所特徴量として用いられる**
- HLAC **は位置に対して不変で線型性を持つ．**LBP **は輝度変化に頑健．**GIST **は強力**

局所特徴量を用いる

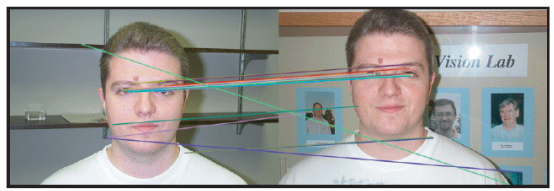
- **局所特徴量の代表格と言えば** SIFT (Scale-Invariant Feature Transform) [20]
- SIFT **では特徴点の検出と特徴量の記述を行う**
- **検出した特徴点に対して，画像の回転，スケール変化，照明変化等に頑健な特徴量を求める (アフィン変換に完全に不変ではない)**
- **開発者は** David Lowe (`http://people.cs.ubc.ca/~lowe/`)



藤吉弘亘, "**一般物体認識のための局所特徴量** (SIFT **と** HOG)," PCSJ/IMPS2008 **ナイトセッション** (2008) **より引用**

low-level 特徴量のまとめ

- **ヒストグラム特徴量はべき変換** [21] **により精度が向上することが多い．インタセクション，**χ^2 **距離との相性が良い**
- HoG **特徴量は線型** SVM **との親和性が高い** [22]
- Deep Learning **に組み込める** (sum-product network) [23]
- SIFT **特徴量は** Identification **に向いているが** Object Categorization **には向いていない** $\Rightarrow$ Bog of Features **アプローチの利用**



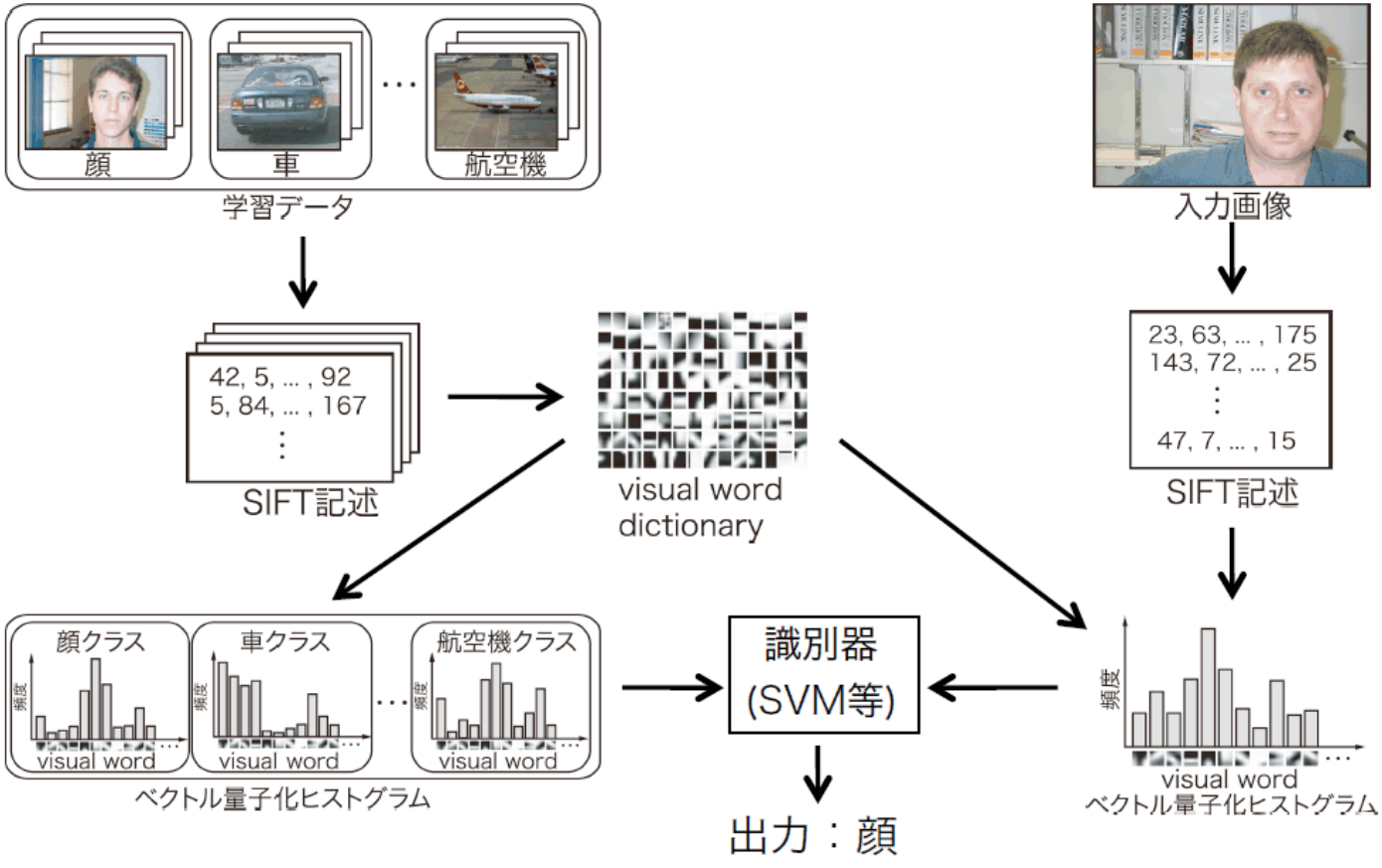
(a)同一人物による対応点探索


(b)異なる人物での対応点探索

藤吉弘亘, "**一般物体認識のための局所特徴量** (SIFT **と** HOG)," PCSJ/IMPS2008 **ナイトセッション** (2008) **より引用**

Bag of Features (BoF) [24]

各局所特徴量に最も類似した代表的な特徴量 (visual word) **の出現頻度で画像を特徴付ける**



山下，藤吉，”特定物体認識に有効な特徴量``，CVIM 2008

BoF の流れ

1. **局所特徴量の算出**

   全ての学習画像から SIFT 特徴量等の局所特徴量を算出

2. **クラスタリング（ベクトル量子化）**

   得られた全ての局所特徴量を k-means 法により k 個のクラスタに分割する (各クラスタの重心を visual word，visual alphabet と呼ぶ)

3. **画像のヒストグラム表現**

   ある画像に対して，その画像の個々の局所特徴量と k 個のクラスタの重心との距離を測り，最近傍のクラスタへ投票を行うことで，k 次元のヒストグラムを計算する

4. **識別器の構築**

   全ての学習画像をヒストグラムで表現した後，識別器を構築する

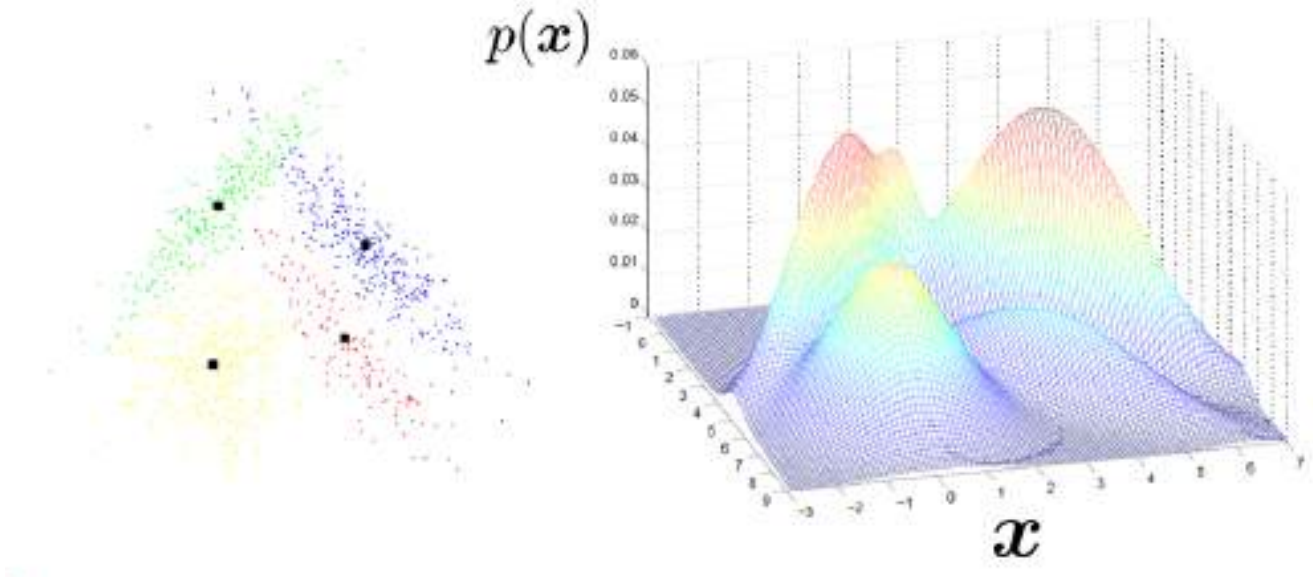
5. **テスト画像のカテゴリー分類**

   テスト画像についても visual word を使って k 次元ヒストグラムを作成し，そのヒストグラムを特徴としてクラス分類する

Gaussian Mixture Model (GMM)

k-means **の替わりに** GMM **も良く用いられる**

$$p(\boldsymbol{x}) = \sum_{i=1}^{k} \pi_i \mathcal{N}(\boldsymbol{x}|\boldsymbol{m}_i, \boldsymbol{\Sigma}_i)$$

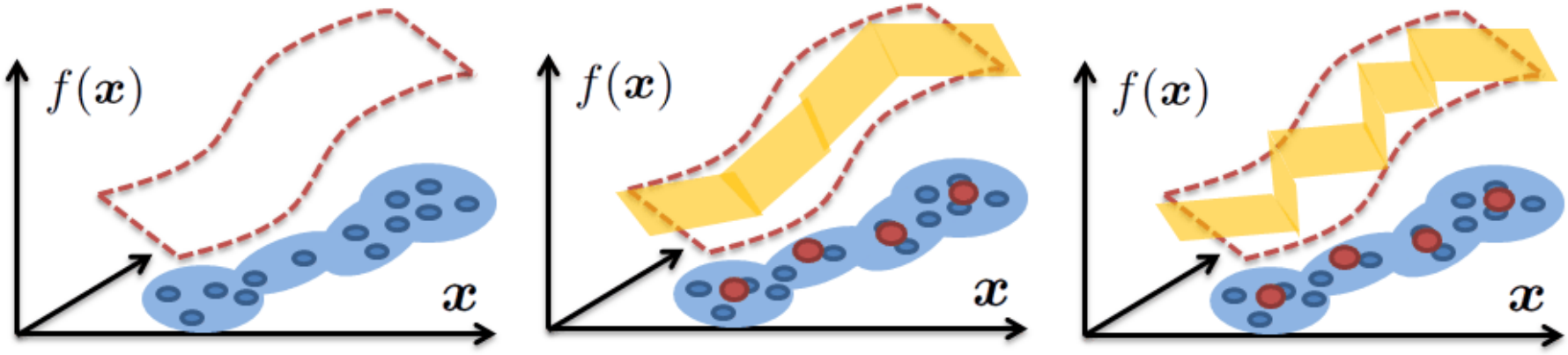


- **モデルへの尤度を投票値とする** soft-assignment **を実現可能**
- super vector **の作成に利用**

SIFT や BoF の拡張

- SURF: `http://www.vision.ee.ethz.ch/~surf/`
- Affine Invariant Keypoint [25]
- Sparse coding [26]: Sparse Coding **により一つの局所特徴を複数** word **へ割り当てる**
- Max pooling [27]: **各** visual word **について割り当てられた局所特徴のスコアの最大値を特徴ベクトルの成分とするもの**
- VLAD [28]: **各** visual word **に帰属する局所特徴量 (原点は重心) の平均ベクトルを連結したもの**
- Super vector coding [29]: BoF **と** VLAD **を組み合わせた特徴量**
- GMM Supervectors [30]: TRECVID2011 **の優勝チームはこれを用いている** [31]
- Fisher vector [32]: **局所特徴の平均に加え分散も用いる**

BoF や Super-Vector は何をしているのか？

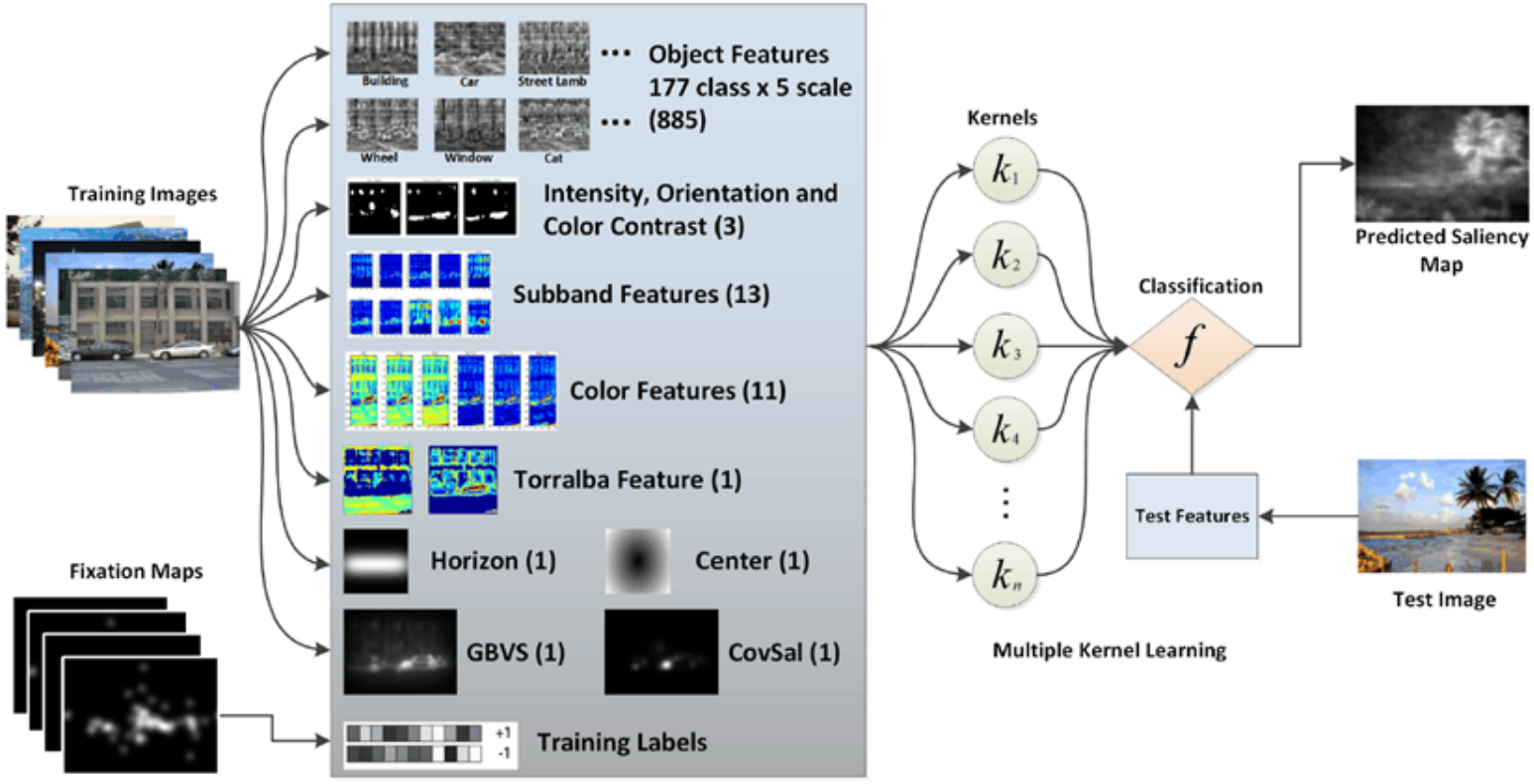


文献 [33] **より転載**

- **局所特徴量が成す非線形多様体 (左図) を近似しようとしている**
- BoF **は階段状 (右図)，**super vector coding **は超平面の組合せ (中央図) で近似しようとしている**

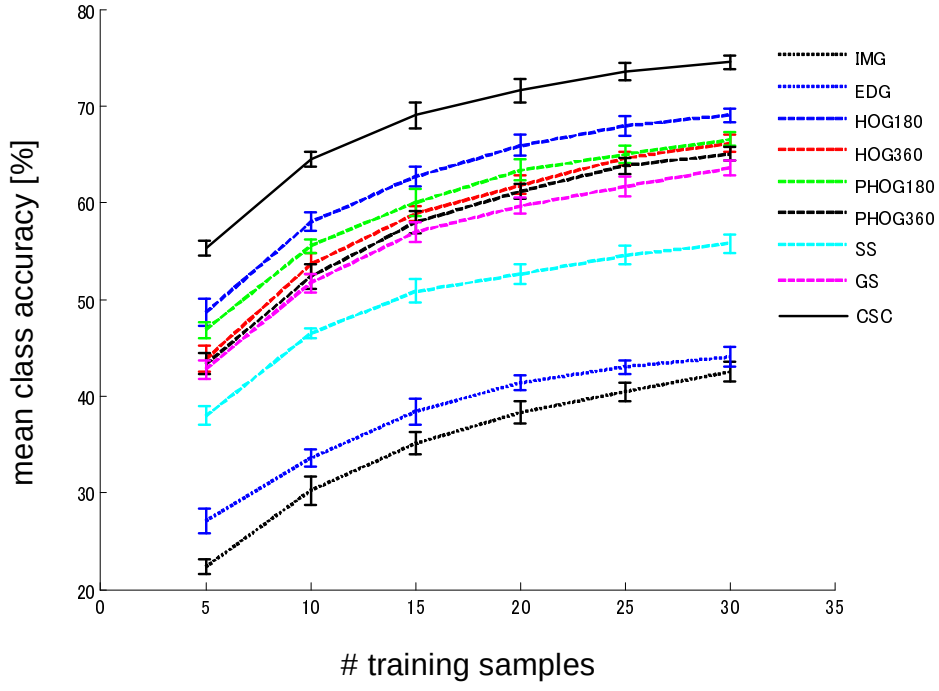
特徴毎に fusion する

- どの特徴量が識別に有効かは事前にはわからない
- さまざまな特徴量を組み合わせて識別を行う [34]
- multiple kernel learning [35, 36] など



文献 [37] より転載

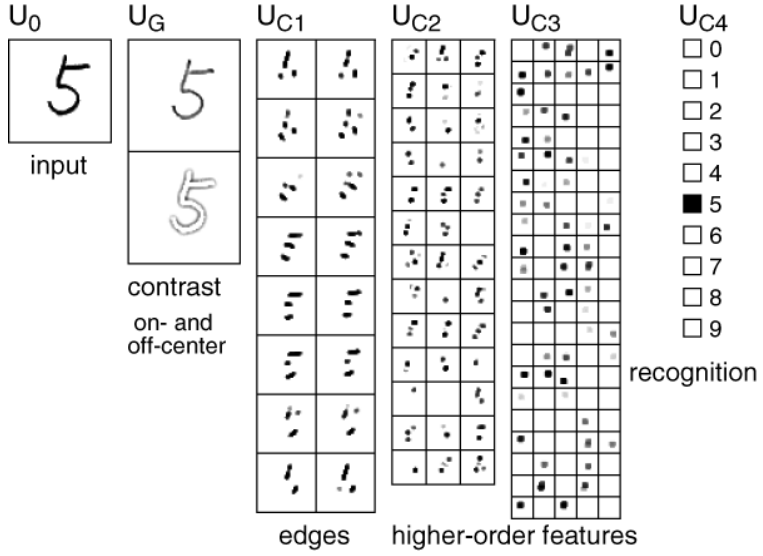
- **未知画像は** 15 **枚，訓練画像は** $5 \sim 30$ **まで** 5 **刻みで変化，**ROI**，**SELFIC **有**
- `IMG`: 32×32 (pixel **モノクロ**) **画像の輝度値**
- `EDG`: 32×32 **画像のエッジ強度**
- `HOG180`: 64×64 **画像の** 180 **度** HOG
- `HOG360`: 64×64 **画像の** 360 **度** HOG
- `PHOG180`: 128×128 **画像の** 180 **度** PHOG (**平方根変換**)
- `PHOG360`: 128×128 **画像の** 360 **度** PHOG (**平方根変換**)
- `SS`: 32×32 **画像の** self-similarity **特徴量**
- `GIST`: 128×128 **画像の** gist **特徴量**



訓練画像数	MKL	LP-beta	部分空間法
5	46.5 ± 1.1	59.5 ± 0.2	55.3 ± 0.8
10	59.2 ± 0.5	69.2 ± 0.4	64.4 ± 0.8
15	66.0 ± 0.9	74.6 ± 1.0	69.1 ± 1.4
20	70.8 ± 0.9	77.6 ± 0.3	71.6 ± 1.2
25	74.3 ± 0.8	79.6 ± 0.4	73.6 ± 0.9
30	77.7 ± 0.5	82.1 ± 0.3	74.5 ± 0.7

Deep learning

- **特徴抽出と識別部を多層ニューラルネットで構築**
- **元祖は福島先生のネオコグニトロン** [38]
- G.E. Hinton [39] **により一挙に有名に**

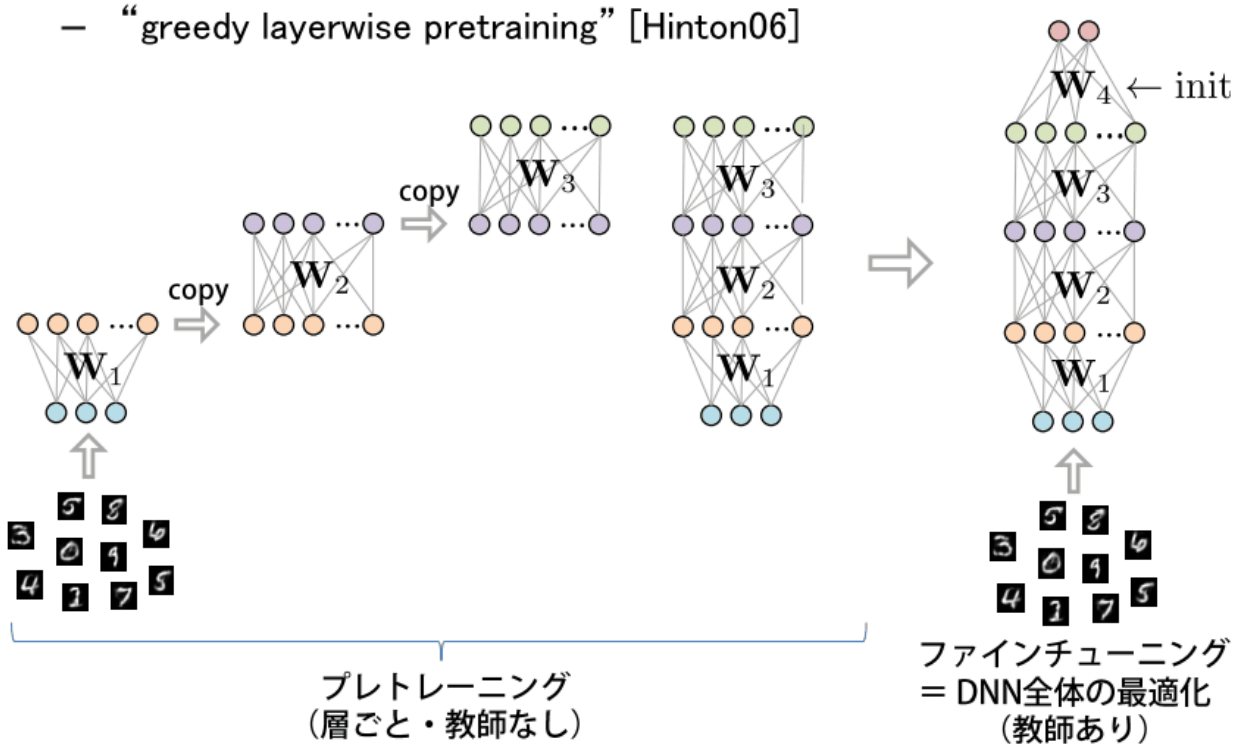


`http://www4.ocn.ne.jp/~fuku_k/` **より転載**

科学映像館「脳をつくる」（動画）
`http://www.kagakueizo.org/movie/industrial/349/`

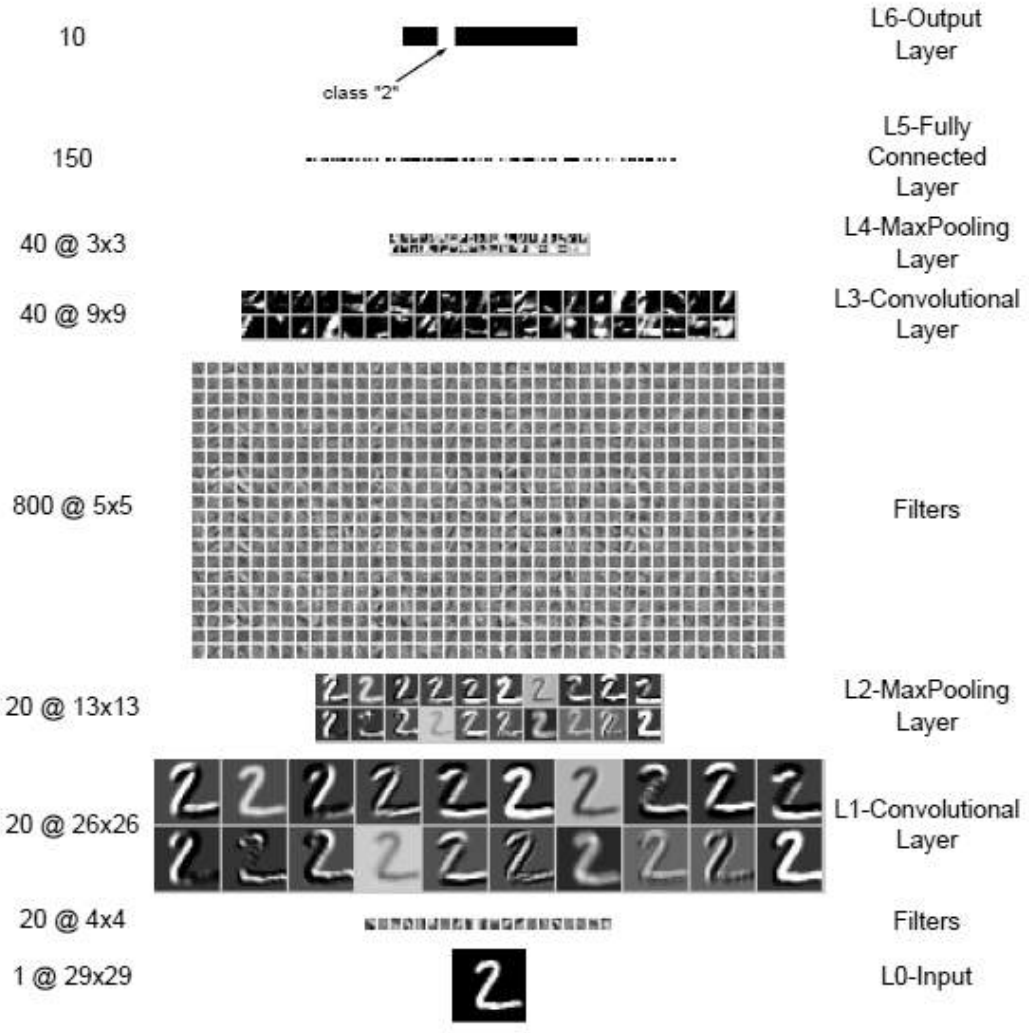
Deep learning の構造

- 4 層以上からなるネットワーク (ニューラルネットとは限らない)
- 最終層以外は独立に学習
- 過学習を避けるための Dropout 処理等の様々な工夫



岡谷，“ディープラーニングと画像認識への応用”， SSII2013 より引用

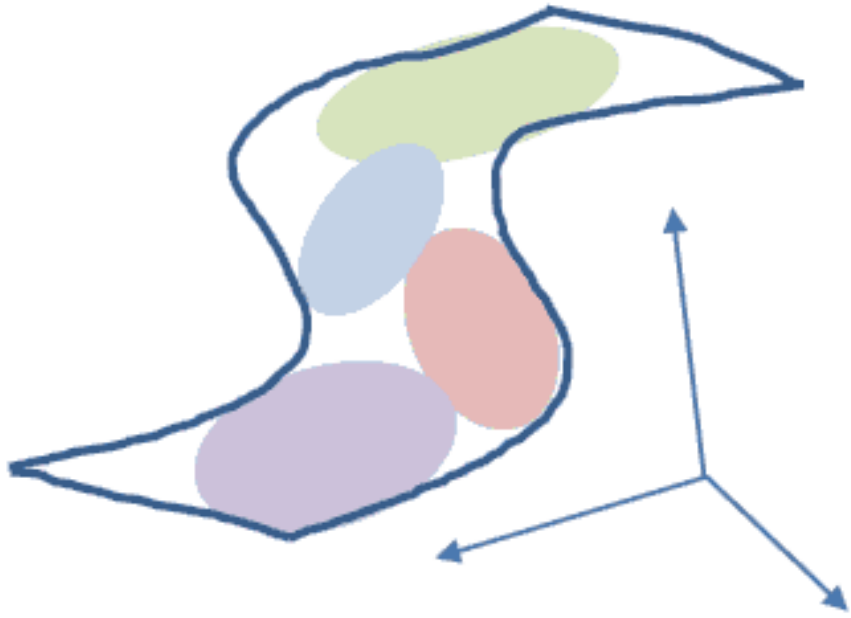
Deep Learning の一例



文献 [41] **より転載．接ベクトルに似ている**

Deep Learning は何をしているのか？

高い汎化能力．例えば MNIST [42] **のエラー率**:
kNN:3.1%, Tangent Distance: 1.1%，Deep Learning 0.23%



文献 [40] **より転載**

- **パターンの成す非線形多様体を上手く近似する超曲面を推定**
- **そこへパターンを射影したのちに識別境界を引く**

本講演の内容

- 幾何学的不変性を獲得するための基本的なアプローチや最近の動向について概説

- 非常に進歩の速い分野 (数週間単位ぐらい？)
- 最新の手法で何でも解けるわけではない (TRECVID の Semantic Indexing タスクで DL よりも従来法の方が良い場合もあった)
- テストデータにフィッティングさせないよう，評価は慎重に

参考文献 I

[1] http://pascallin.ecs.soton.ac.uk/challenges/VOC/

[2] **柳井，**“**一般物体認識の現状と今後**,” **情報処理**, vol. 48, no. SIG16(CVIM19), pp. 1–24, 2007.

[3] K. Grauman and B. Leibe, “Visual object recognition,” Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan & Claypool Publishers, 2011.

[4] http://opencv.jp/

[5] J. Shotton, M. Johnson, and R. Cipolla, “Semantic Texton Forests for Image Categorization and Segmentation”, CVPR 2008.

[6] **中山英樹**, “**タカとハヤブサはどこが違う？ ～新たな認識領域「詳細画像識別」の展開と応用～**”, SSII, 2014. http://www.nlab.ci.i.u-tokyo.ac.jp/pdf/ssii2014fgvc.pdf

[7] P. Yan, S.M. Khan, and M. Shah,“3D Model based Object Class Detection in An Arbitrary View,” ICCV, 2007.

[8] http://www.vision.caltech.edu/Image_Datasets/Caltech101/

[9] **石井健一郎**, **上田修功**, **前田英作**, **村瀬 洋**, “**わかりやすいパターン認識**,” **オーム社**, Aug. 1998.

[10] **村瀬，**S.K.Nayar, “2 **次元照合による** 3 **次元物体認識-パラメトリック固有空間法-**,” **信学論**, no. J77-D-II, vol. 11, pp. 2179–2187, 1994

参考文献 II

[11] A. Torralba, et al., "80 million tiny images: A large dataset for non-parametric object," IEEE Tran. on PAMI, vol. 30, no. 11, pp. 1958–1970, 2008.
http://groups.csail.mit.edu/vision/TinyImages/

[12] P.Y. Simard, Y. LeCun, J.S. Denker, and B. Victorri, "Transformation invariance in pattern recognition – Tangent distance and tangent propagation," Int'l J. of Imaging Systems and Technology, vol. 11, no. 3, 2001.

[13] C. Vondrick, et al. "HOGgles: Visualizing Object Detection Features", ICCV, 2013.
http://web.mit.edu/vondrick/ihog/

[14] **木村ら，"拡張外郭方向寄与度特徴と輪郭特徴とを用いた手書き漢字/非漢字のハイブリッド認識，" 信学誌，**J82-D-II(12), 1999**.**

[15] N. Dalal & B. Triggs, "Histograms of oriented gradients for human detection," Proc. of CVPR, pp. 886–893, 2005.

[16] A. Bosch, A. Zisserman, and X. Muñoz, "Representing shape with a spatial pyramid kernel," Int'l Conf. on Image and Video Retrieval, pp. 401–408, 2007.

[17] N. Otsu & T. Kurita, "A New Scheme for Practical Flexible and Intelligent Vision Systems," MVA, 1988.

[18] T. Ojala, et al., "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions", ICPR, 1994.

参考文献 III

[19] A. Oliva & A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," Int'l J. of Computer Vision, vol. 42, no. 3, pp. 145–175, 2001.

[20] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," Int'l J. of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

[21] **若林哲史**, **鶴岡信治**, **木村文隆**, **三宅康二**, "**手書き文字認識における特徴量の次元数と変数変換に関する考察**," **信学論**, vol. J76-D2, no. 12, pp. 2495–2503, 1993.

[22] H. Bristow & S. Lucey, "Why do linear SVMs trained on HOG features perform so well?", arXiv : 1406.2419v1, 2014.

[23] H. Nakayama, "Stacked local autocorrelation features,", MIRU, 2014.

[24] G. Csurka, C.R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," ECCV, 2004.

[25] K. Mikolajczyk and C. Schmid,"Scale & Affine Invariant Interest Point Detectors," IJCV, vol. 60, no. 1, pp. 63–86, 2004.

[26] J Yang, et al., "Linear spatial pyramid matching using sparse coding for image classification," CVPR 2009.

[27] Y-L. Boureau, et al., "Learning mid-Level features for recognition," CVPR 2010.

参考文献 IV

[28] H. Jegou, et al., "Aggregating local descriptors into a compact image representation," CVPR 2010.

[29] Xi Zhou, et al., "Image classification using super-vector coding of local image descriptors," ECCV 2010

[30] W.M. Campbell, et al., "Support vector machines using GMM supervectors for speaker verification," IEEE Signal Processing Lett., vol. 13, pp. 308–311, 2006.

[31] N. Inoue and K. Shinoda, "A fast MAP adaptation technique for GMMsupervector-based video semantic indexing," ACM Multimedia, 2011.

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

[33] **原田達也**, "**大規模画像データを用いた一般画像認識**," SSII 2012. `http://www.isi.imi.i.u-tokyo.ac.jp/~harada/`

[34] J. Kittler, M. Hatef, R. P.W. Duin, and J. Matas, "On combining classifiers," IEEE Trans. on PAMI, vol. 20, no. 3, pp. 226-239, 1998.

[35] F.R. Bach, G.R.G. Lanckriet, and M.I. Jordan, "Multiple kernel learning, conic duality, and the SMO algorithm," Proc. of ICML, 2004.

[36] P. Gehler & S. Nowozin, "On feature combination for multiclass object detection," Proc. of ICCV, 2009.

参考文献 V

[37] Y. Kavak, E. Erdem, and A. Erdem, "Visual saliency estimation by integrating features using multiple kernel learning,", ISACS, 2013.

[38] **福島邦彦，"位置ずれに影響されないパターン認識機構の神経回路のモデル - ネオコグニトロン -,"** **信学論** A, vol. J62-A, no. 10, pp. 658–665, 1979.

[39] G.E. Hinton and R.R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science, vol. 313. no. 5786, pp. 504–507, 2006.

[40] **岡谷貴之，"ディープラーニングと画像認識への応用**," SSII, pp. OS03-1-11, 2013. `http://www.vision.is.tohoku.ac.jp/index.php/download_file/view/41/136/`

[41] D. Ciresan, et al., "Multi-column deep neural networks for image classification," Proc. of CVPR, pp. 3642–3649, 2012.

[42] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Intell. Signal Process., pp. 306–351, 2001. `http://yann.lecun.com/exdb/mnist/`