【チュートリアル】一般物体認識

PRMU/CVIM チュートリアル講演

2009年 11月 26日

電気通信大学　情報工学科

柳井 啓司

「一般物体認識」大ブーム到来！

- 2000年以降に急速に研究が発展！
  - (1) 局所特徴量による新しい画像表現の提案
    SIFT と Bag-of-features
  - (2) 機械学習の進歩
    SVM, boosting, graphical model, MCMC, …
  - (3) 大規模データセットの入手の容易化
    Web, Web Image Search, CGM, Flickr, Youtube, Mturk
  - (4) 計算機(PC)の高速大容量化
    メモリ: GB, HDD: TB, マルチコア, クラスタ
    クラウドコンピューティング

アウトライン

- ［導入］一般物体認識とは？
  - 特定物体　と　一般物体
  - 現在どこまで出来るか？
- 基本的な手法：Bag-of-Features (BoF)
- 画像単位での分類
- 一般物体の位置検出
- データセット
- ［まとめ］今後の展望

1．一般物体認識とは？

【参考文献】
柳井啓司. 一般物体認識の現状と今後. 情報処理学会論文誌: コンピュータビジョン・イメージメディア, Vol.48, No. SIG16 (CVIM19), pp. 1-24, 2007.

「物体認識」とは？

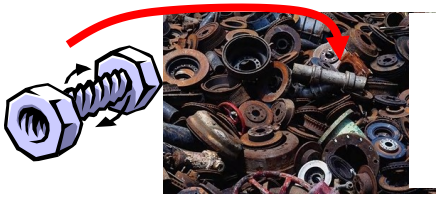
- 画像中の「物体」を認識する技術，研究分野

文字認識

顔画像検出

カテゴリー認識　シーン認識

カテゴリー認識



3Dモデル物体認識

顔画像認識

同一物体の認識

特定物体検出

カテゴリー・同一物体認識

- カテゴリー認識
  - 文字認識
  - 顔検出
  - 一般的な名称の物体・シーンを認識
    （e.g. ライオン，ラーメン，山，空）

一般物体認識

- 同一物体認識
  - 部品認識（3D モデルベースト認識）
  - 人物名認識
  - 登録物体の検索

特定物体認識

The University of Electro-Communications
Tokyo, JAPAN (UEC)

特定物体認識（同一物体認識）

- 特定の登録物体が画像中にあるかどうか認識

見た目（appearance）がまったく同じ物体を検出

現時点では剛体なら
95%以上認識可能．ただし，顔や動物などの変形物体は難しい．

一般物体認識（カテゴリー認識）

- “一般的な”実世界画像の認識
  - デジカメやWebの画像を自動認識.
  - 画像内容 を 言語（記号）で記述. 意味理解.

クマ

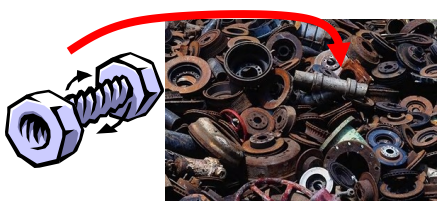
（草の上の）トラ

（草を食べる）ゾウ

静止画像に対して，その中に含まれる
物体もしくはシーンの一般名称（カテゴリー）を認識

究極的には人間以上に幅広く詳細な認識

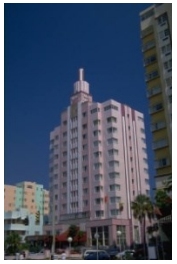
一般物体認識研究の背景



- 従来の画像認識
  - 認識対象を限定していた （例. 部品，顔，文字，自然風景）

↓

- 近年のデジタルカメラ等の普及
  - 一般画像データの入手の容易化
    一方，計算機は意味は分からずにただ蓄積.



対象を限定しない一般的画像の認識技術の必要性

画像の意味的処理．画像の取り扱いに関する
セマンティックギャップ解消のための技術.

一般物体認識の分類（1）

- **画像全体**のカテゴリー分類


→ クマ

→ トラ

→ ゾウ

- **画像アノテーション**：複数ラベルの付与

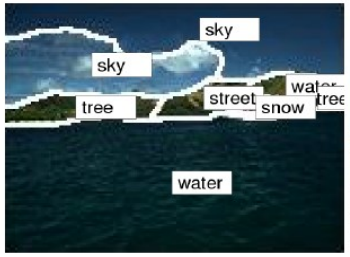

→ クマ
草
水

→ トラ
草
草原

→ ゾウ
キバ
空
草
草原

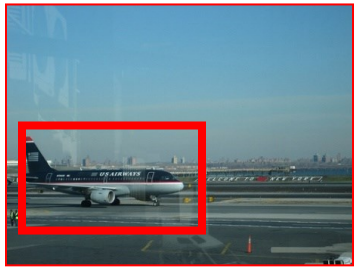
一般物体認識の分類（2）

- **画像ラベリング**：領域分割→分類

- **カテゴリー物体検出**：ウィンドウ探索



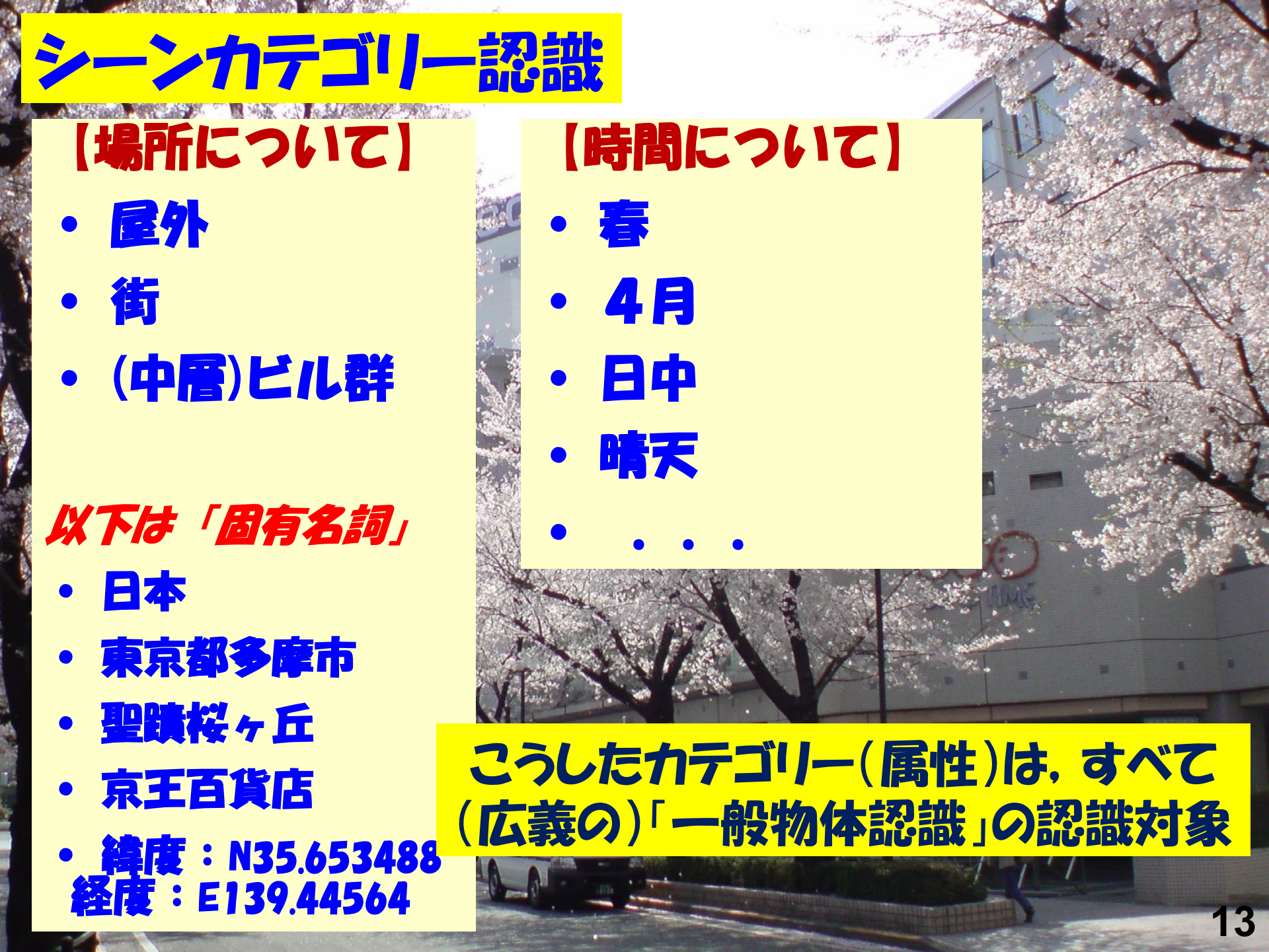


- **オブジェクト領域抽出**：認識＋領域分割




認識カテゴリーの例

シーンカテゴリー認識



【場所について】

- 屋外
- 街
- (中層)ビル群

以下は「固有名詞」

- 日本
- 東京都多摩市
- 聖蹟桜ヶ丘
- 京王百貨店
- 緯度：N35.653488
  経度：E139.44564

【時間について】

- 春
- ４月
- 日中
- 晴天
- ・・・

こうしたカテゴリー(属性)は，すべて(広義の)「一般物体認識」の認識対象

一般物体認識の困難性

- **認識対象が多様（カテゴリー内変化が大）**
  - 同一種類（カテゴリー）の物体でも形は様々．変形も．
  - 撮影時の条件が多様（視点位置，向き，変形，スケール，照明（天候），背景，オクルージョン）
- **認識対象が多い．（カテゴリー数が多い．）**
  - 辞書に出ている名詞の数だけある！　数万？
  - 何を認識するべきか？　レベルは？　動物orライオン?

様々な「ライオン」

The University of Electro-Communications
Tokyo, JAPAN (UEC)

カテゴリー内変化(1)：いろいろな「椅子」

The University of Electro-Communications
Tokyo, JAPAN (UEC)

カテゴリー内変化(1)：いろいろな「椅子」



どんな「椅子」が認識できればいい？

(1) 世の中の「椅子」すべて？

(2) 典型的なもののみ？

(3) 「座る」機能を提供する物体すべて？



The University of Electro-Communications
Tokyo, JAPAN (UEC)

カテゴリー内変化(2)：いろいろな視点からの見え方

[P.Yan, S. M. Khan and M. Shah:
3D Model based Object Class Detection in An Arbitrary View, CVPR 2007]より

3Dモデルを持たずに見た目(appearance)のみで認識するのが現在の一般的方法.

様々な視点からの見た目の学習が必要

The University of Electro-Communications
Tokyo, JAPAN (UEC)

カテゴリー内変化(2)：いろいろな視点からの見え方

どこからみた「バイク」が認識できればいい？

(1)すべての方向？　360度．下からも上からも？

(2) 典型的な見え方のみ？　真横．斜め前方．
canonical view

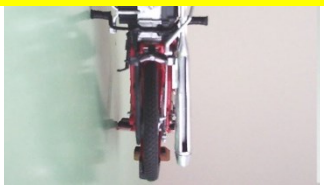
(3) 状況によって異なる．
地上からみた場合．　⟷　高層ビルや飛行機から．









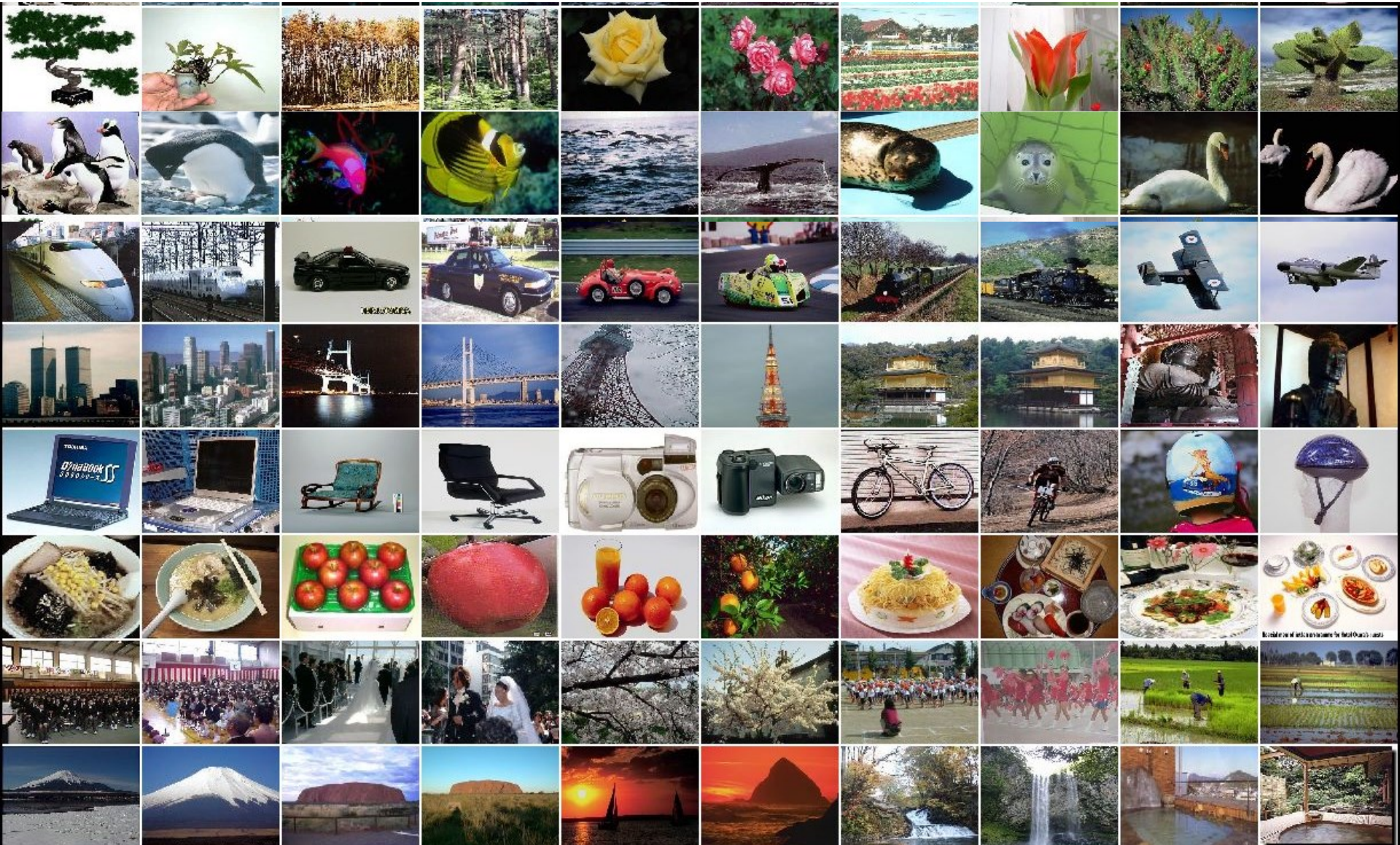



[P.Yan, S. M. Khan and M. Shah: [redacted]
3D Model based Object Class Detection in An Arbitrary View, CVPR 2007]より

カテゴリー数が多い：

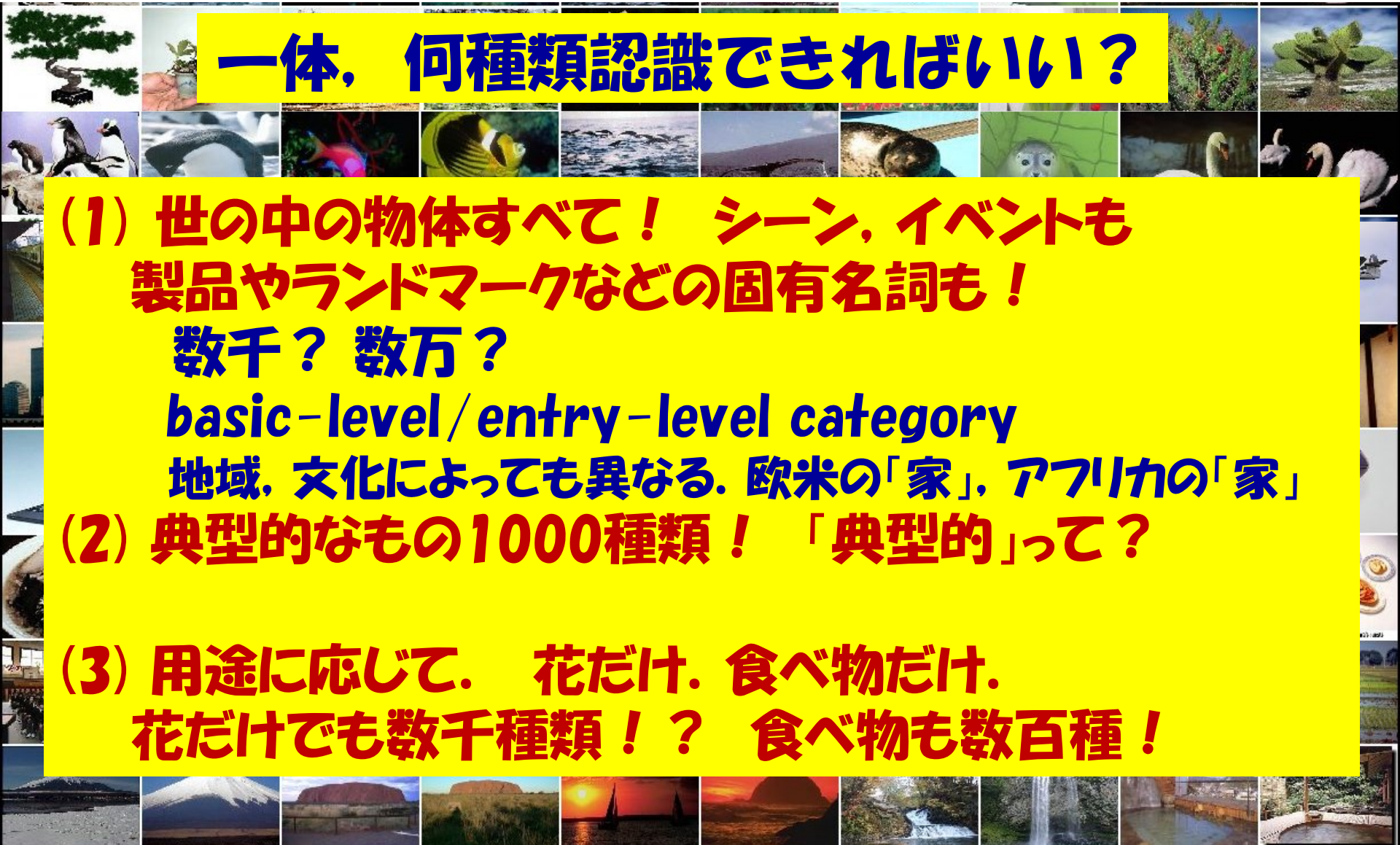
多様なカテゴリー

The University of Electro-Communications
Tokyo, JAPAN (UEC)



カテゴリー数が多い：多様なカテゴリー

The University of Electro-Communications
Tokyo, JAPAN (UEC)



一体，何種類認識できればいい？

(1) 世の中の物体すべて！　シーン，イベントも
　製品やランドマークなどの固有名詞も！
　　数千？ 数万？
　　basic-level/entry-level category
　　地域，文化によっても異なる．欧米の「家」，アフリカの「家」

(2) 典型的なもの1000種類！　「典型的」って？

(3) 用途に応じて．　花だけ．食べ物だけ．
　花だけでも数千種類！？　食べ物も数百種！

「カテゴリー問題」

[1] どの程度のカテゴリー内変化に対応すべき？

- どちらが「バイク」？

- 一般的な視点とは？

[2] 何種類のカテゴリーに対応すべき？

- 人間は数千種類に対応.
  - 256？1024？ Wordnet全部(5万)？ 応用次第？

「カテゴリー問題」は「フレーム問題」

- 人工知能での「フレーム問題」

- DBに含まれる「椅子」のみを認識できれば
  とりあえずいいことにして，研究する！

現状での「カテゴリー問題」への対処法

- 標準データセットを利用して研究を進める.
  - カテゴリー問題を棚上げして，研究を進めることが可能.
    アルゴリズム開発に集中出来る！　パターン認識の問題.

「一般物体認識」－「カテゴリー問題」＝「パターン認識問題」

リー.

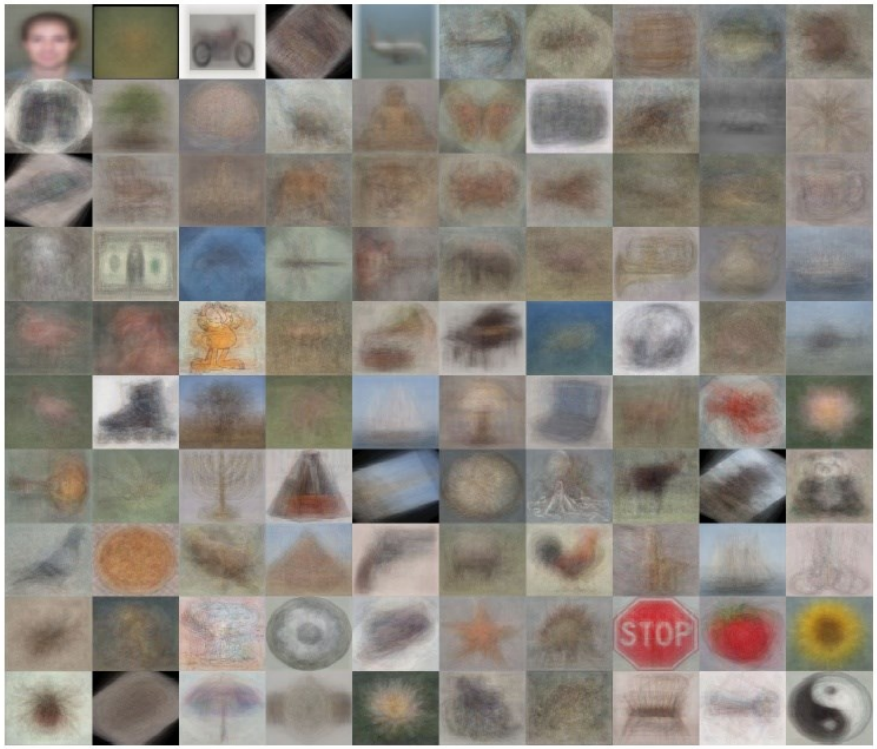
- 問題点：標準データセット作成者の主観でカテゴリ内変化やカテゴリ数が決められている.
- [それに対する新たな動き] Webの大量画像
  - Wordnetに出ている名詞(7万語)すべてについてデータセットを作成　[Tinyimage 2009][imagenet.org 2009]

「Webに存在する知識　≒　人間の知識」とみなすことにより対処.

[Nakayama et al. 2009]

実際の研究用データセット(1)

- Caltech-101：2004 年に登場の101種類カテゴリーの画像認識データセット（全体分類用）



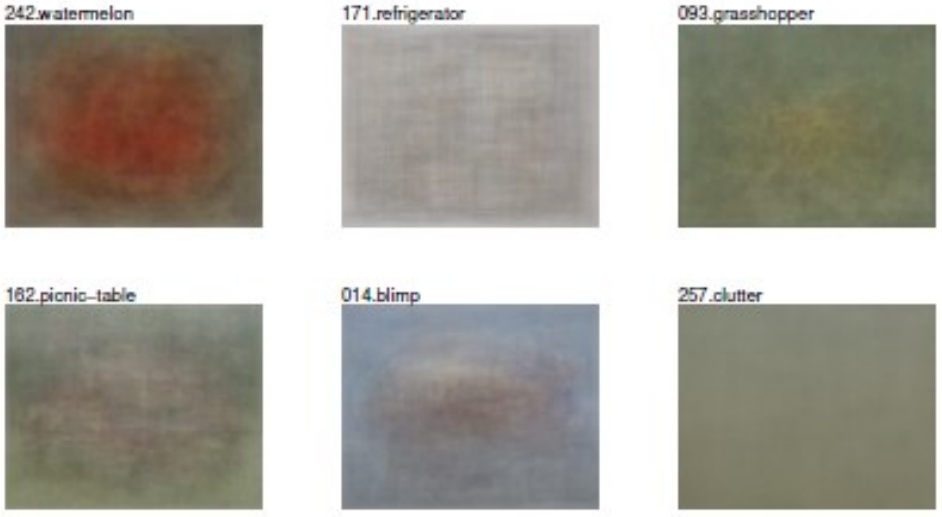
各カテゴリの平均画像

平均画像からカテゴリが分かる！
カテゴリ内変化が小

技術レベルに合わせて，意図的にやさしくしてある．

実際の研究用データセット(2)

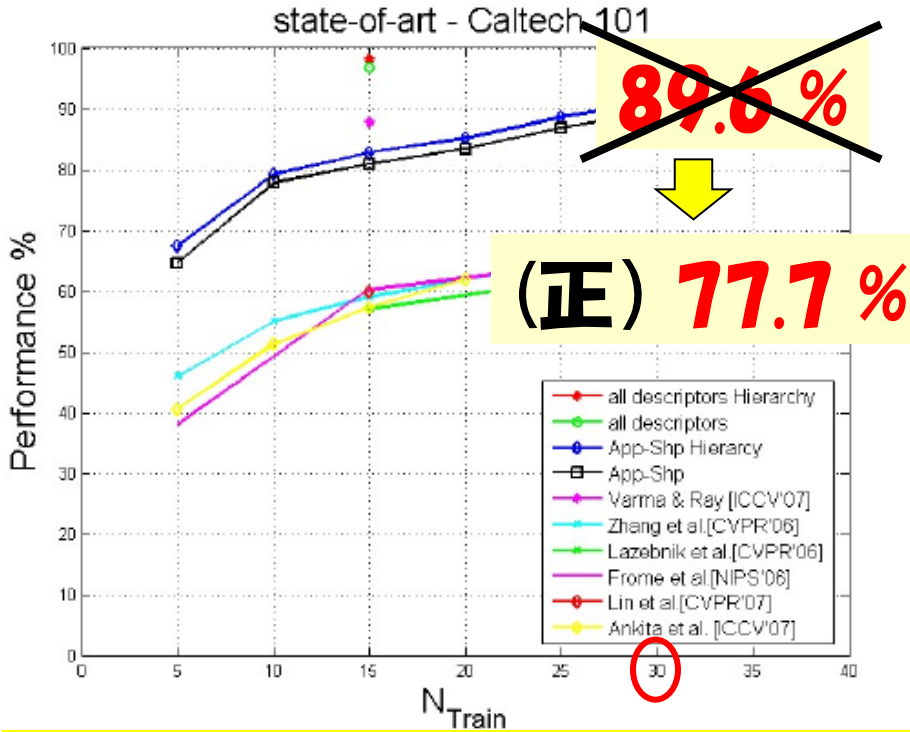
- Caltech-256 (2007年登場)
  - クラス数256，クラス内変動も大きくなっている．平均画像からカテゴリ認識不能．
    - （少し改善された．）

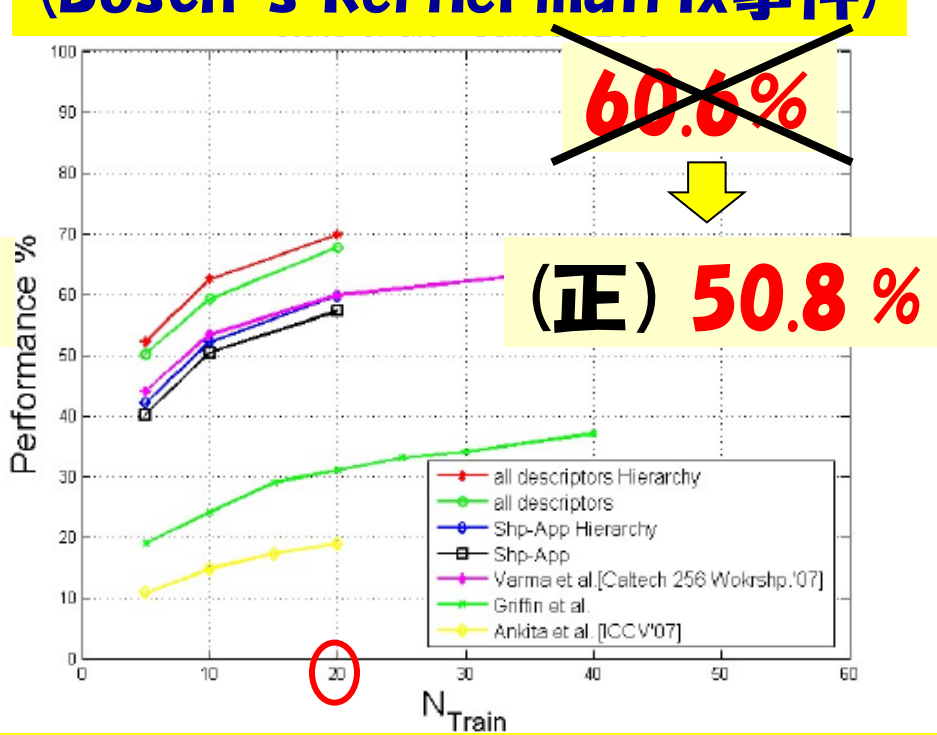


最新の一般物体認識の分類精度

- Caltech-101/256

[Varma et al. 2007]の
Caltechの結果は間違い！
(Bosch's Kernel matrix事件)





Caltech-101/256 の認識精度は，2007年から向上していない．

BoF+MKL が現在の最高精度 → 新しいブレークスルーが必要！

実際の研究用データセット(3)

- Caltech-256 の問題点
  - カテゴリー数が多すぎて，重複するカテゴリーがデータセットに含ま

「チャーハン/ピラフ」問題

認識カテゴリーを決めるのは実は難しい問題．重複なしで選ぶことは困難．

画像認識研究者は認識方法のみを研究している!?

The University of Electro-Communications
Tokyo, JAPAN (UEC)

PASCAL Visual Object Challenge

2005年～

Caltechと違って
多様であるが
種類は20のみ.
毎年データが代わる.

1)分類
2)検出
3)物体領域抽出
4)人間のパーツ検出
の
4つのタスクがある.

Aeroplanes – *all images contain at least one aeroplane.*

Bicycles – *all images contain at least one bicycle.*

Caltechよりも多様な画像. Flickrから収集した画像.

厳密な定義がない認識カテゴリー:
-どのような「認識」をするべきか？-

- 認識にふさわしい認識カテゴリは？
  - 人間の認識との一致.
    - Basic/entry-level （認知科学の用語. 厳密な定義はない. ）
    - Web上のアノテーションの利用(CGM).
      Wisdom of crowd
  - 認識・分類が容易なカテゴリ階層.
    - カテゴリのエントロピー評価 （visual/non-visual con.)
  - 実用性.
    - アプリケーション依存
      - 食べもの. 食事管理.
      - 図鑑. 動植物.

一般物体認識は「人工知能問題」

- 究極的には，人間が「○○」と思う物は，計算機も同様に「○○」と認識するべき．
  - 人間の言語と対応したカテゴリーの認識．
  - 特定物体認識の場合は，カテゴリー（クラス）がない．

理想の一般物体認識システム

- 一般物体認識：人間が基準．人工知能問題．
  - クラス定義があいまい．フレーム問題．
- 特定物体認識：全く同一の物を認識．
  - 問題が明確．アルゴリズムの問題．

2. 基本手法:
Bag-of-Features (BoF)
（ここまでは理想の一般物体認識，ここからは現実の一般物体認識）

【参考文献】
[Low99] Lowe, D.G.: Object recognition from local scale invariant features, *Proc. of IEEE International Conference on Computer Vision, pp. 1150–1157* (1999).
[Siv03] Sivic, J. and Zisserman, A.: Video Google: A Text Retrieval Approach to Object Matching in Videos, *Proc. of IEEE International Conference on Computer Vision*, pp.1470–1477 (2003).
[Csu04] Csurka, G., Bray, C., Dance, C. and Fan, L. “Visual categorization with bags of keypoints,” in *Proc. of ECCV Workshop on Statistical Learning in Computer Vision*, pp. 59–74 (2004).

一般物体認識の歴史

研究者の知識がすべて！

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.

- 人手によるルール記述に一般性がない. 知識爆発.

80年代後半　3次元の復元. モデルベースト.

- 既知形状の物体のみ. 実世界でうまくいかない.

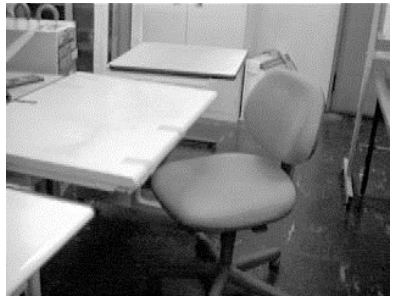
90年代 学習による認識. 顔画像や既知物体の検出　.

- 顔画像認識(Eigenface)の成功. 固有空間法.
- 画像DBにおける画像の意味的分類.

00年代　局所特徴＋機械学習 により 大きく進歩

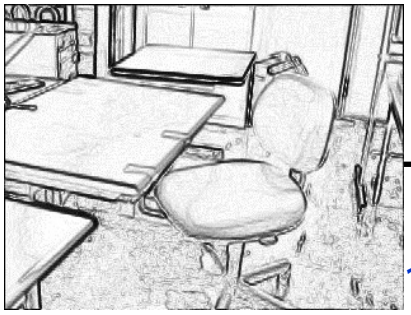
90年代までは，画像認識においてはマイナーな研究分野.

昔の研究（1995年くらい）



入力画像

エッジ抽出

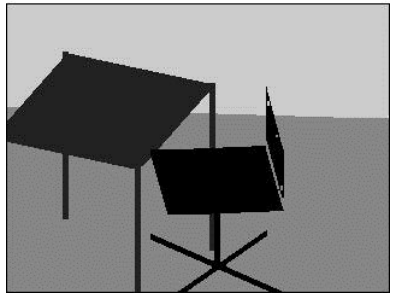


エッジ画像

特徴選択

有意特徴

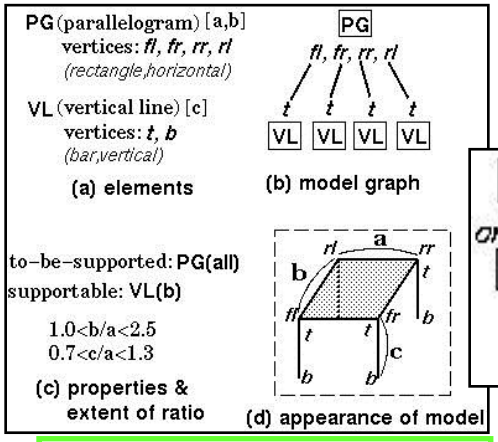
モデル当てはめ



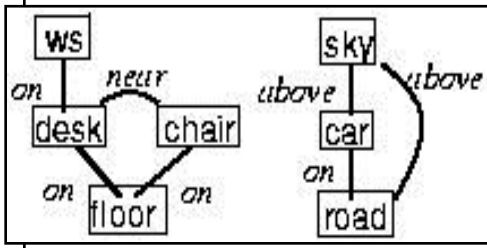
認識結果

領域分割

分割領域



モデル知識ベース



関係知識ベース

2000年以降の発展 突然ブレイク！

Bag-of-Features, SVM + PCの進化,Webの発展

2000年 Constellation model (確率モデル)

2001年 確率手法による単語と画像の対応付け [RWCP]

2002年 Word-image translation model

2003年 Video Google (image sea 特定物体認識の基本手法

2004年 Bag-of-Features(BoF 一般物体認識の基本手法

2004年 Caltech101 2005年 Pascal VOC

2005年～ BoF + probabilistic graphical model
(PLSA, LDA, HDP, their modifications)

BoF + SVM with modified kernel

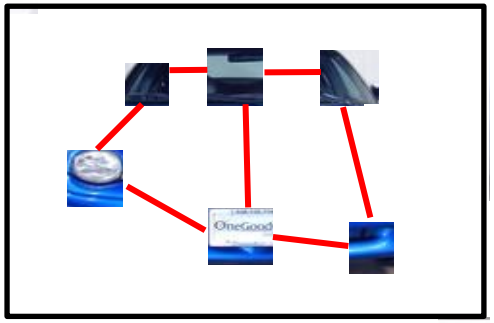
BoF + MRF for semantic region segmentation

2007年 Caltech256 (256カテゴリーのデータセット)登場

全体特徴から局所特徴へ

原画像

- 2000年以前の認識
  - 認識対象の全体を利用
  - 問題点：物体の隠れや変形に弱い．



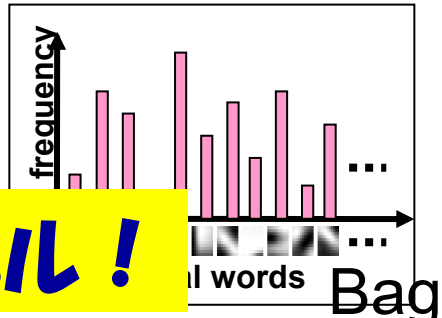
星座モデル

- 部分の組み合わせによる認識:
  複数の部分と位置関係で認識

問題点：位置関係の扱いは面倒！視点変化にも弱い

Bag-of-featuresモデル

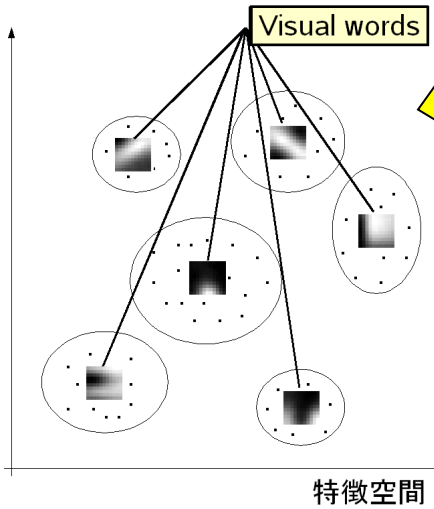
- 局所パターンの分布に基づく認識(bag-of-features)
  - 位置関係はすべて無



シンプル！

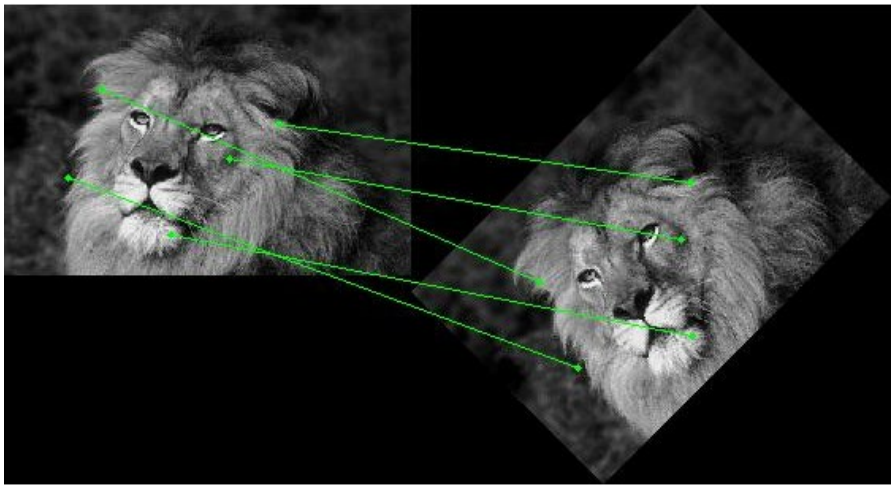
局所特徴量 による 特定物体認識

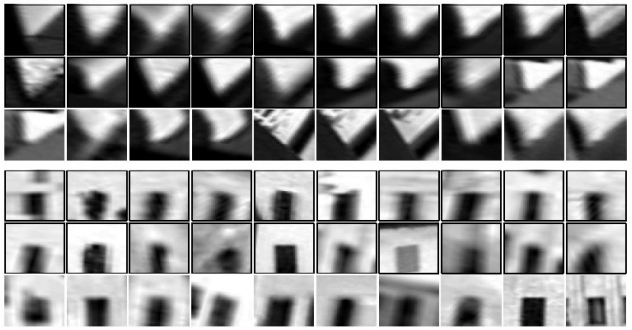
- SIFT [Lowe 99]
  - 回転およびスケール変化に不変な局所特徴量
    ただし，D.Loweが想定したのは，**特定物体検出**
- Video Google [Sivic et al. 03]
  - SIFT特徴ベクトルをベクトル量子化し，画像を
    visual wordsの集合とみなす．
  - テキスト検索の手法(Google)を
    応用し(inverted file)，高速画像
    検索を実現．（同一部分の検索）



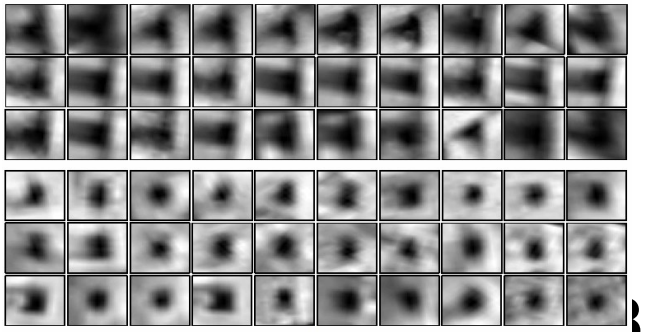
局所特徴量

- 局所パターン(10x10～30x30程度)をベクトル化(64～128次元)
  - 類似パターンは，類似したベクトルになる
  - SIFT法 と SURF法が有名
    - フリーソフト．SURFはOPENCV．
    - 商用利用は有償(特許出願済み)





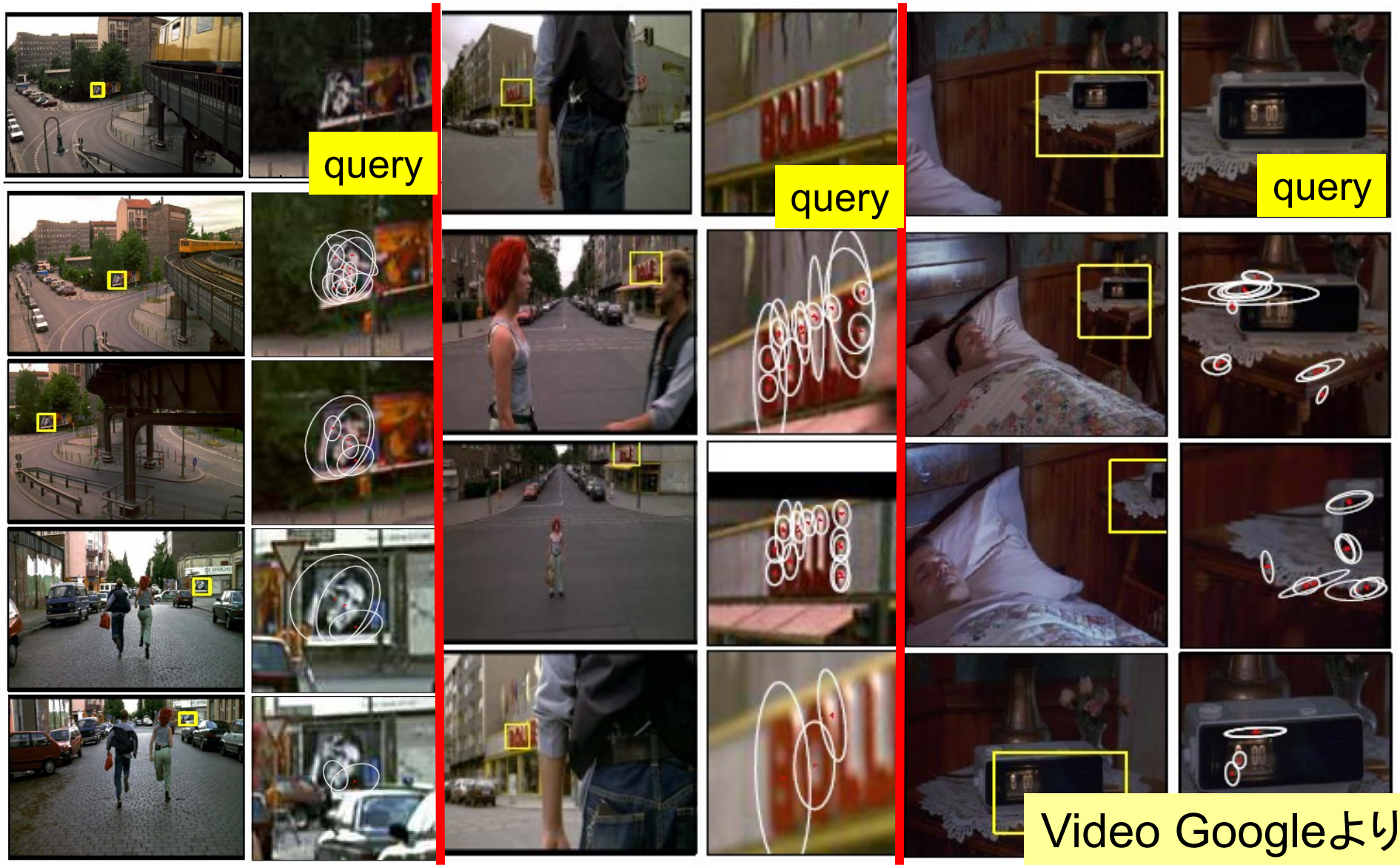
(a)



(b)

The University of Electro-Communications
Tokyo, JAPAN (UEC)

局所特徴による特定物体検索



The University of Electro-Communications
Tokyo, JAPAN (UEC)

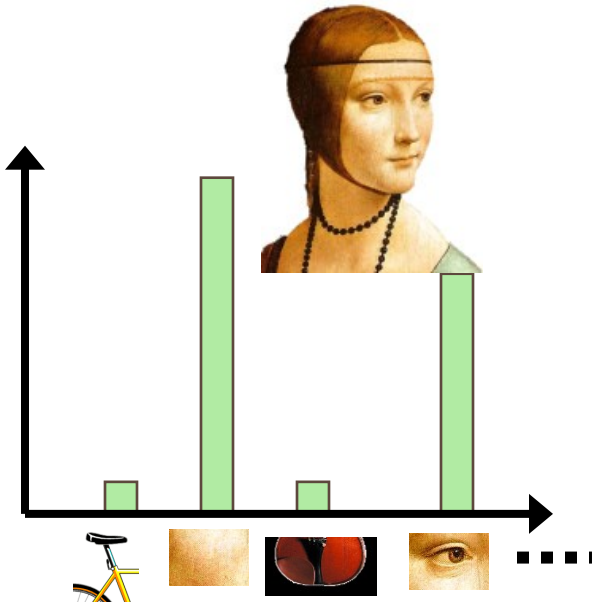
Bag-of-features 表現：代表パターンの集合による表現

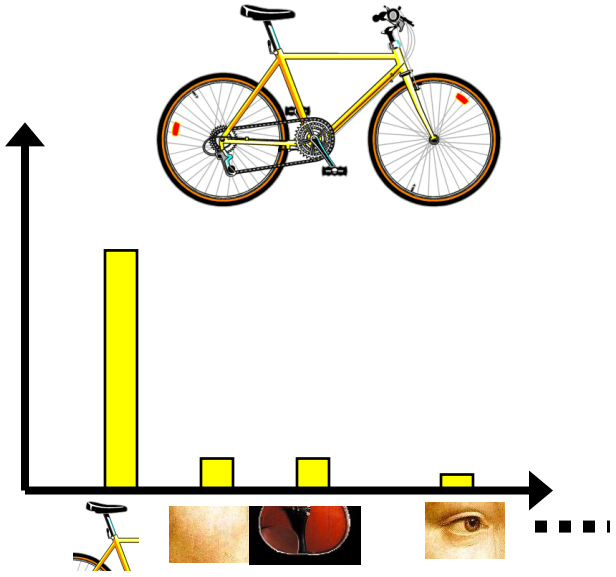
Object → **Bag of 'words'**

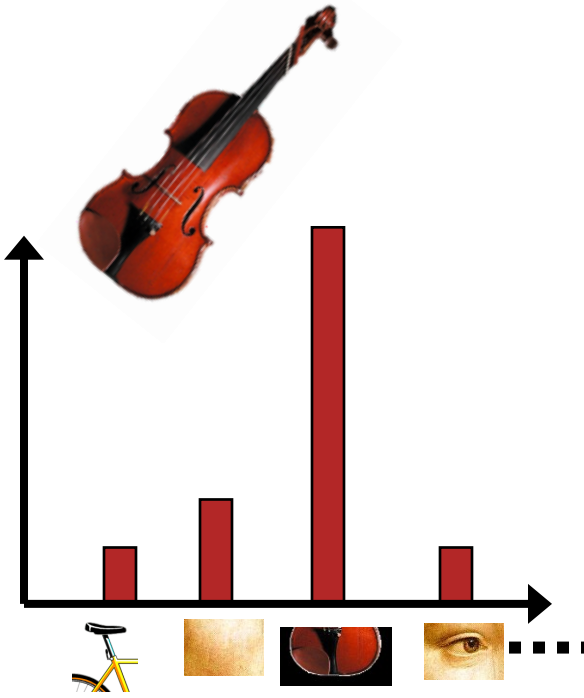
[ICCV 2009 tutorial スライドより]

同一カテゴリー：同一代表パターンを含む異なるカテゴリ：代表パターンは異なる.

- 代表パターンのヒストグラム表現







[ICCV 2009 tutorial スライドより]

Bag-of-features [Csu04]: visual wordの一般物体認識への適用

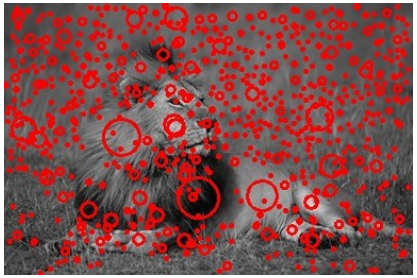
- Visual words の集合として画像を表現
  - Visual words のヒストグラムを画像特徴とする
  - 単語出現頻度によりテキストを表現する方法のbag-of-wordsの考え方を画像に応用．語順を無視するのと同様に，位置を無視．
- Bag-of-features によって表現された特徴ベクトルをNaive Bayes, SVMなどの機械学習手法で分類．テキスト分類と同じ！
- Bag-of-visual-words (BoVW), Bag-of-keypoints (BoK) とも言うことがある．

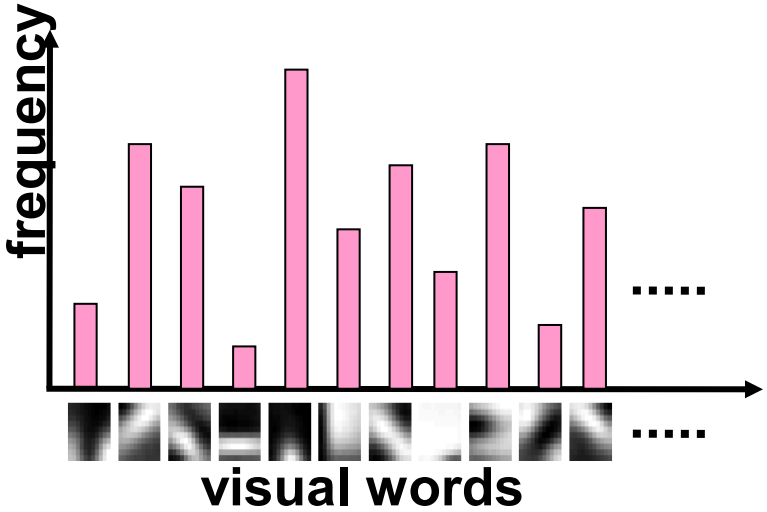
Bag-of-featuresのアルゴリズム: bag-of-features表現への変換

特定と同じ

- 画像を visual word の出現頻度ヒストグラムで表現
  1. 各画像について，数千個の特徴点を抽出.
  2. SIFT記述子により特徴点周辺パターンをSIFT特徴ベクトルとして抽出.

  SIFT法
  （特徴点抽出＋記述）
  3. 予め求められた visual words (codebook)に基づいてSIFT特徴ベクトルをベクトル量子化.
  4. 画像毎にヒストグラムを作成.





特徴点のサンプリングの方法

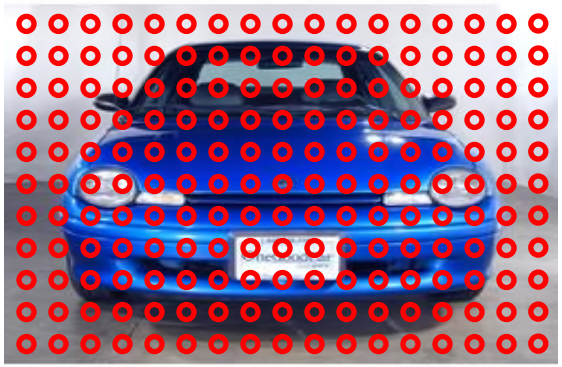
特定と違う！

- 主な3つの方法
  - Difference of Gaussian(DoG)　SIFT法の方法　sparse sampling と呼ぶ
  - Random sampling
  - Grid sampling

dense sampling と呼ぶ

DoG (sparse)

random(dense)
(スケールもランダム)



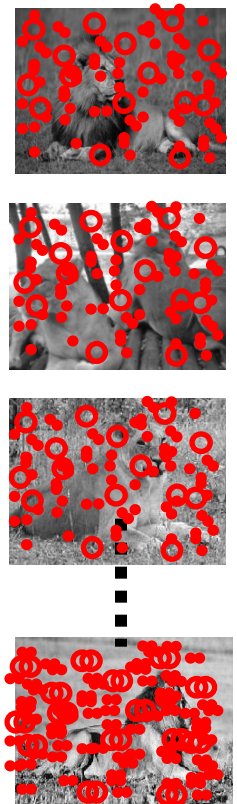
grid(dense)
(同一点に複数固定スケール)

カテゴリー分類においては，パターンのない部分も重要．
Random/gridは,点の数の制御可能．一般には多い方がいい．

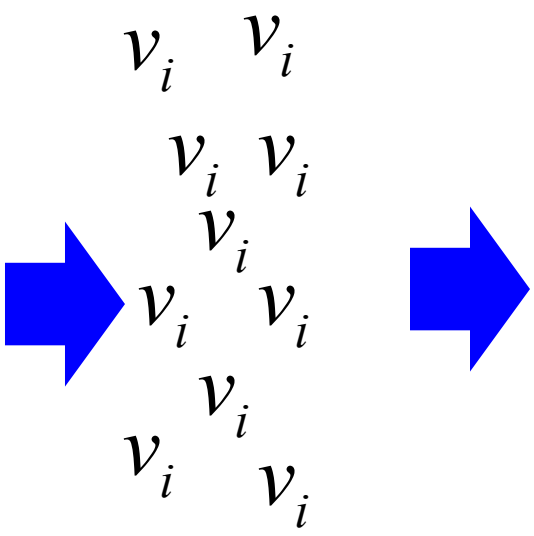
Visual words の求め方

特定と同じ

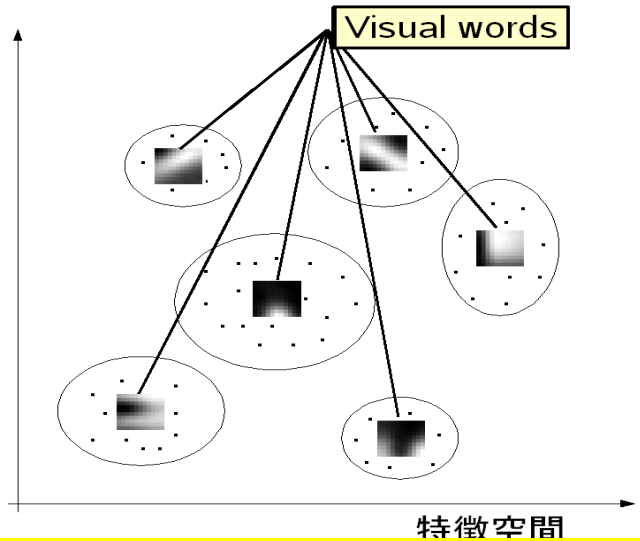
- 学習画像(正例，負例)を用意し，SIFT特徴ベクトルを全画像から抽出 （枚数が多い場合は，ランダムサンプリング）
- k-means クラスタリングを実行



各クラスタの中心が“visual words”



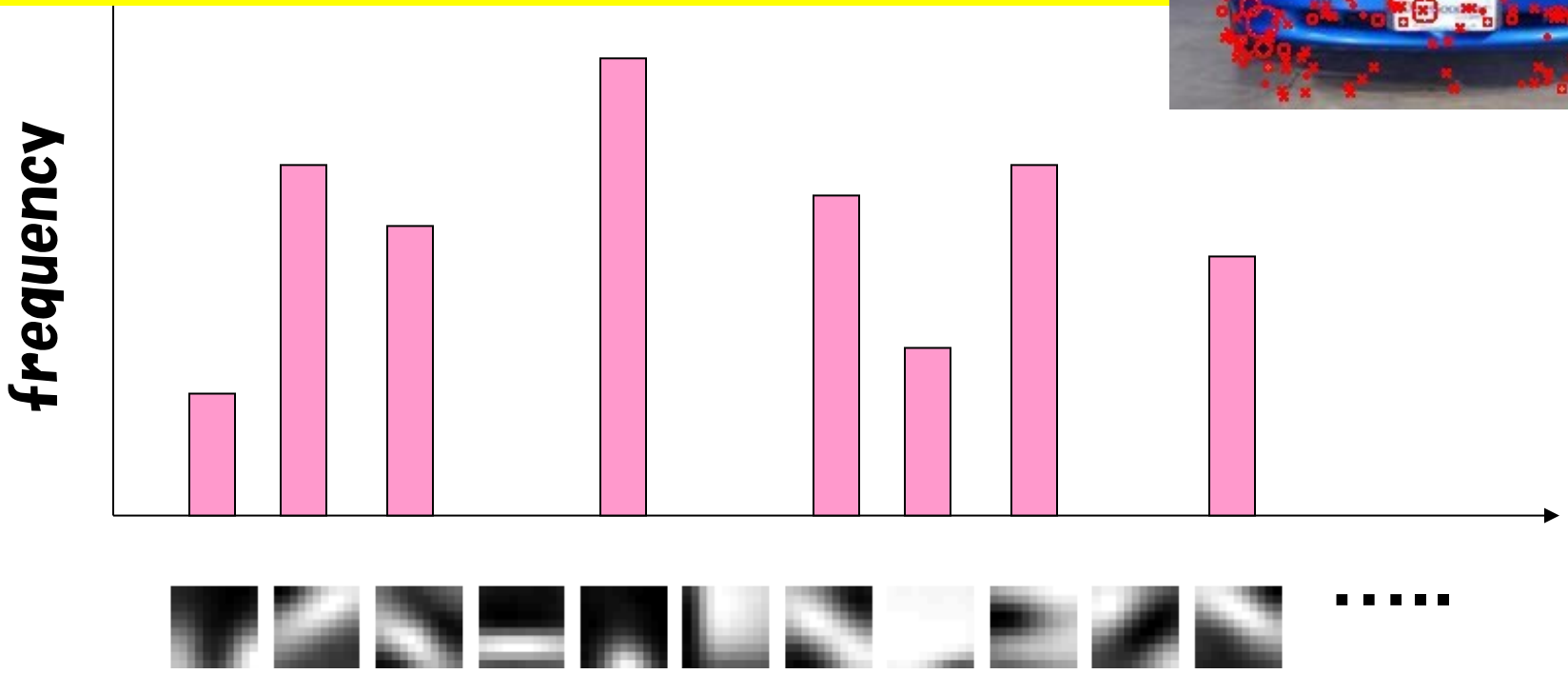
SIFT vectors



“Visual words”は，代表的な局所パターンに相当する．

Bag-of-features 表現 (BoF)

Visual wordsの出現頻度(ヒストグラム)によって画像を表現.



Visual words
(数百～数千個)

**次元は, 数百～数千次元.
スパースなベクトルになる.**

Bag-of-features表現を用いた画像認識

- あとは，多次元ベクトルの分類問題
  - 最初の論文[Csu04]では，以下の2つの手法で実験
    - SVM (support vector machine)
    - Naive Bayes

    ⇒ 従来手法の結果を大きく改善
- 例：Web画像の分類
  - 10種類のキーワードについて，平均適合率
    - 従来手法(領域分割＋GMM)： 73.5%
    - BoK + SVM : 82.4%

(ブラウザーにジャンプ)

[まとめ] Bag-of-featuresの特徴

- 認識方法ではなくて，**画像の表現方法**.
- 局所出現パターンのヒストグラム.
  - カラーヒストグラムに似ているが，色空間の代わりにSIFT記述子空間(128次元)をk個に分割．SIFT記述子の回転，拡大縮小に不変な特性を受け継ぐ.
  - 単語の順番を無視して，単語の出現頻度のみで文章を表現するbag-of-words表現と等価．Bag-of-visual-words.
- **Bag-of-words用のテキスト処理手法が適用可能.**
- SIFT/SURF と k-means で容易に実装可能.
  - SIFT/SURFは公開ソフトが利用可能．SIFT++, OPENCVなど.
  - k-meansは教科書レベル.
  - 分類は，SVMを使えば簡単高性能．(LibSVM, OPENCV ml lib)

BoFベクトルの分類

- 機械学習手法なら何でもいい．
  - Naive bayes法
    - 単純で実装容易だが，意外とうまくいく．多クラス対応．
  - Nearest neighbor法
    - データが大量にないとあまりうまくいかない．多クラス対応．
  - Support vector machine (SVM)
    - 次元が高いので，線形カーネルでもまずます．
    - カイ2乗RBFカーネルが，最近のスタンダード．
    - 1-vs-restはクラス数が多いと計算コスト大，並列化可能．
  - アンサンブル学習
    - Random forest, boosting など．

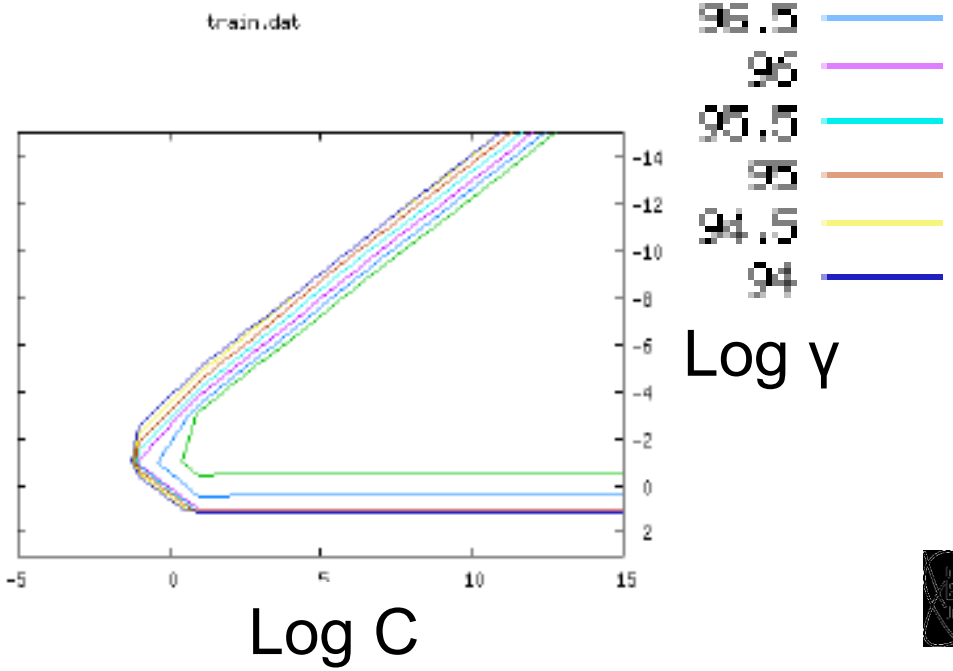
BoF + SVM が最も手軽な方法

1. 学習画像(2/マルチクラス)とテスト画像を準備.
2. グリッドかランダムで，特徴点を決定
3. SIFTかSURFで，特徴点記述
4. K-meansで，codebook 生成.
5. Codebookから各画像をBoFベクトルに変換.
6. LIBSVM か OPENCVのSVMで，学習・分類.
   - LIBSVMはマルチクラス対応なので簡単.
     内部で1-vs-restを自動的に実行.

SVMに関する tips（1）

- SVM は，必ずパラメータの最適化を行おう！
  - 学習データによるクロスバリデーションで行う．
    - C：ソフトマージン(コスト)パラメータ
    - γ：カーネルパラメータ．(RBFカーネル)

- LIBSVMの場合，grid.py で自動探索
- 劇的に結果がよくなることも多い！



SVMに関する tips (2)

[Zhang et al. IJCV2007]

- カーネルは，カイ2乗(χ^2)RBFカーネルを使う！
  - 最近のスタンダード．単に「カイ2乗カーネル」とも．

$$K(\mathbf{x},\mathbf{y}) = \exp\left(-\gamma\chi^2(\mathbf{x},\mathbf{x})\right)$$

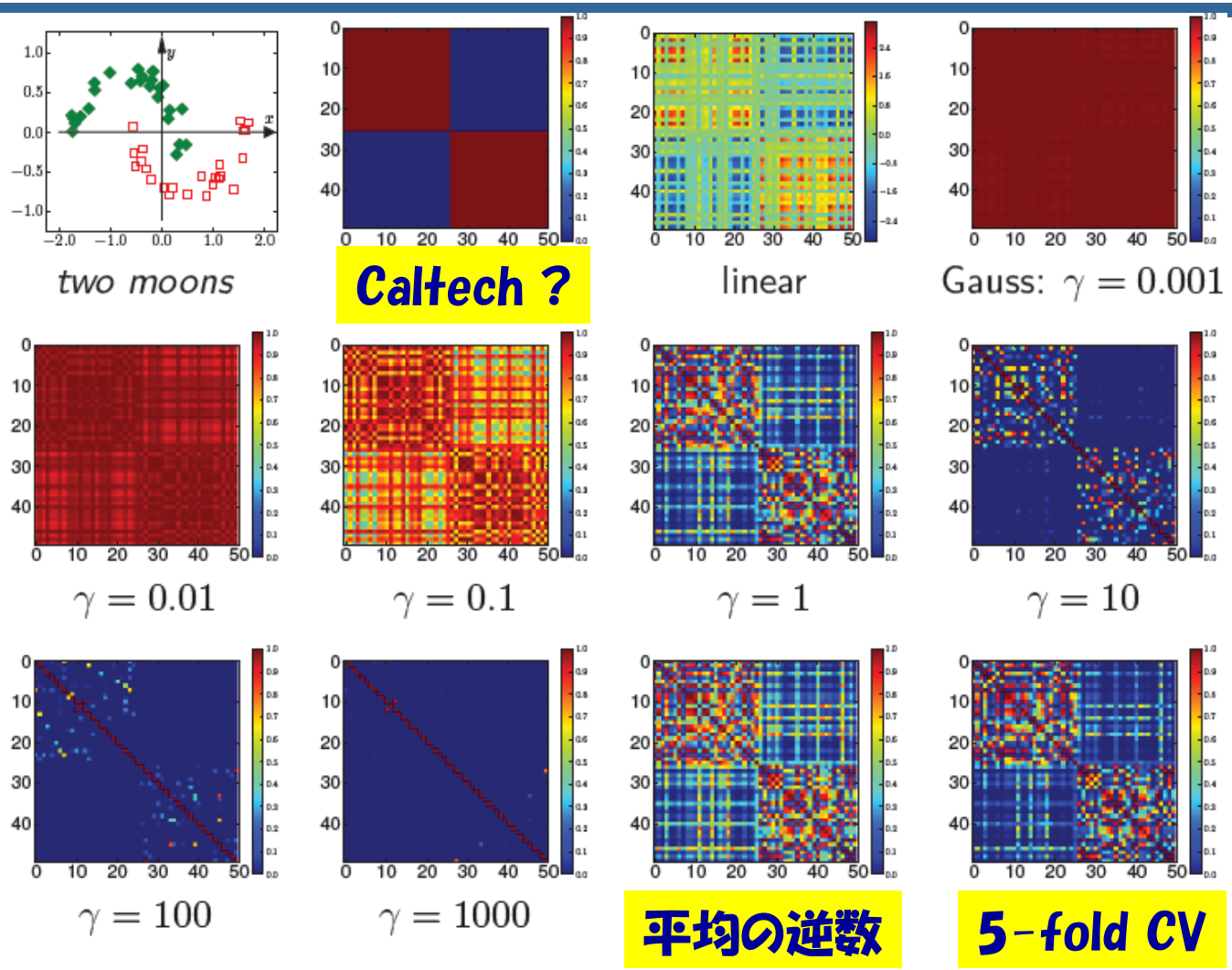
$$\text{where } \chi^2(\mathbf{x},\mathbf{y}) = \sum \frac{(x_i - y_i)^2}{x_i + y_i}$$

  - しかも，γ には，全学習データ間のカイ2乗距離の平均(or中央値)の逆数を設定するとよい．（経験則）

$$\gamma = (median_{i,j=1,\ldots,n}\chi^2(x_i,x_j))^{-1}$$

  - ただし，LIBSVMには自分で追加する必要がある．Cの値の探索は，依然として必要．

γ によるカーネルマトリクスの変化

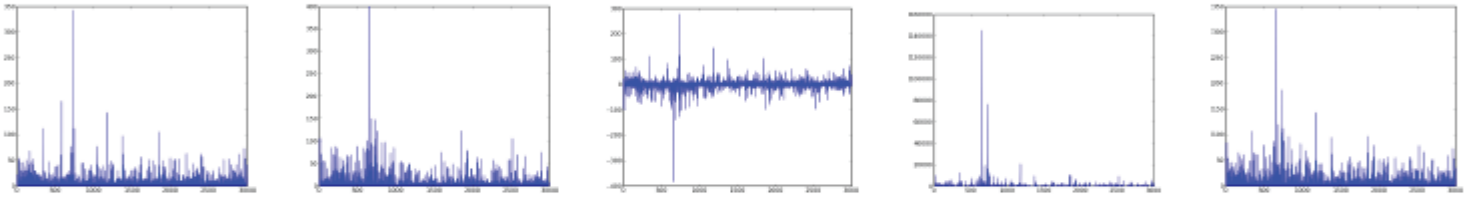


[C. Lampert CVPR2009 tutorial on Kernel Methods スライドより]

カイ2乗カーネルがいい理由

Especially k_{HI} and k_{χ^2} seem to work very well for computer vision:

$$k_{HI}(h,h') = \sum_j \min(h_j, h'_j) \qquad k_{\chi^2}(h,h') = \exp\left(-\frac{1}{\gamma}\sum_j \frac{(h_j-h'_j)^2}{h_j+h'_j}\right).$$



h h' $h-h'$ $(h-h')^2$ $\frac{(h-h')^2}{h+h'}$

- Feature-histograms have few large and many small entries.
- Quadratic measures (L^2 or Gaussian kernel) concentrate on the largest differences: 3 bins (out of 3000) contribute 25%
- $1st$-order (L^1 or χ^2-kernel) consider bins more equally: 3 largest terms contribute 3.5%

[C. Lampert CVPR2009 tutorial on Kernel Methods スライドより]

学部3年生でも「一般物体認識」！

情報工学実験第二2c（2009年度）

電気通信大学　情報工学科　第6学期　月水曜日3，4限　第2ラウンド　担当：柳井

Home
Overview
Lecture Notes
Report Upload
References
Jikken HP

課題資料

(練習問題の提出方法はこちら)

1) MATLABの基礎.
ベクトル・行列操作，ファイル読み込み

2) MATLABによる画像処理.
画像の入出力，色ヒストグラムによる類似画像検索.

3) 局所特徴量.
画像認識の基礎．局所特徴量SIFT．SIFTによる対応点探索．同一物体探索.

4) Bag-of-Visual-Wordsの作成.
K-means法によるコードブックの作成．ベクトル量子

5) Support Vector Machine による画像分類.
SVMによる特徴ベクトルの分類.

6) Web画像を用いた実験.
Webから収集した画像を用いた一般画像認識実験.

6日間で，一般物体認識とMATLABもマスター！

テキストも6日で完成.（6晩徹夜！）

実質3日なので，「特定」に対抗可能!?もしくは「特定」に＋1日.

3. BoFを用いた画像分類

【参考文献】
[Laz06] Lazebnik, S., Schmid, C. and Ponce, J.: Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories, *Proc. of IEEE Computer Vision and Pattern Recognition,* pp.2169–2178 (2006).
[Var07] M. Varma and D. Ray. Learning the discriminative power-invariance trade-off. In *Proc. of IEEE International Conference on Computer Vision, pp.*1150–1157 (2007).
[Rab07] A. Rabinovich, A.Vedaldi, C. Galleguillos, E. Wiewora and S. Belongie: Objects in context, In *Proc. of IEEE International Conference on Computer Vision, pp.*1150–1157 (2007).

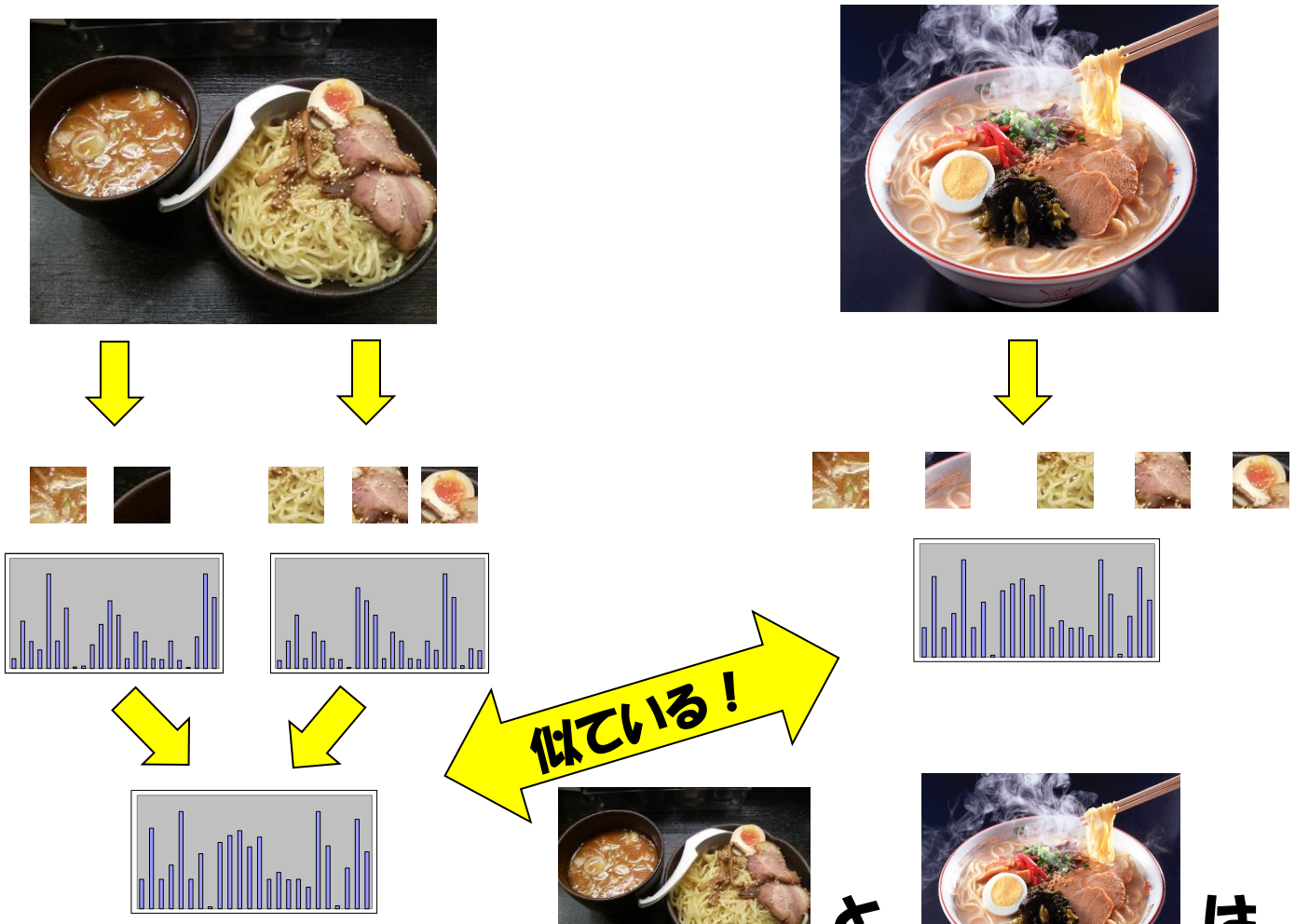
BoF の改良：codebook生成法

- Visual words：代表パターンをいかに見つけるか
  - Unsupervised
    - K-means の代わりに mean-shift
    - K-means の代わりに GMMでソフトクラスタリング
    - Hierarchical k-means で vocabulary tree の生成
  - Supervised：クラスラベルの利用
    - Supervised clustering
      - Information bottleneck (IB) 法など
    - Discriminative visual words のみの利用
    - 空間分割と分類を交互に行う
      - Unifying Discriminative Visual Codebook Generation with Classifier Training [Yang et al. CVPR 2008]
    - 他にもいろいろ多数．カバーし切れません！

BoFの改良 (2)

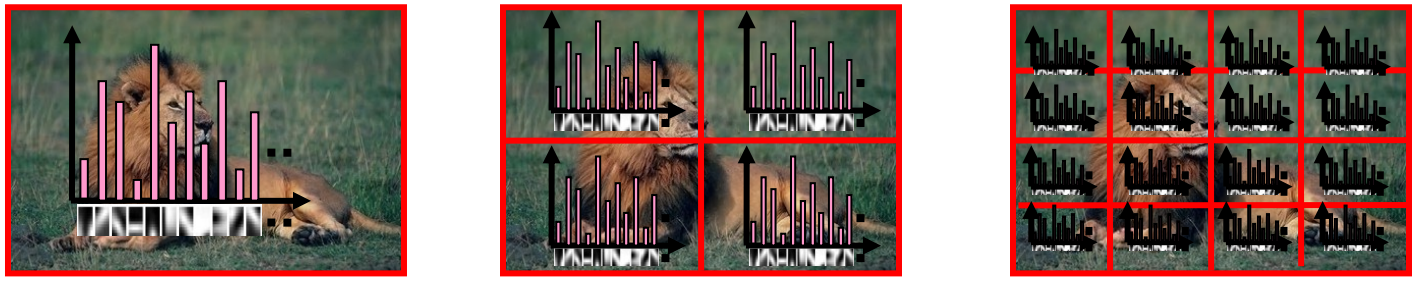
- ヒストグラムのため位置情報を利用しない
  - ブロック分割して，サブヒストグラムを作成
    - Spatial pyramid kernel
- SIFTを利用するので，色情報を利用しない
  - RGBやHSV, Labなど，3つの色成分ごとにSIFTで特徴抽出(128次元×3)
    - Color SIFT
  - 色情報や形状情報，テクスチャ情報を統合
    - 重み付き線形和カーネル による統合
- どこに物体があるか分からない．物体検出．

[一休み] BoFの欠点は．．．



位置情報の導入
Spatial pyramid kernel [Laz06]

- BoKをグリッド分割して階層的にlocal BoF作成

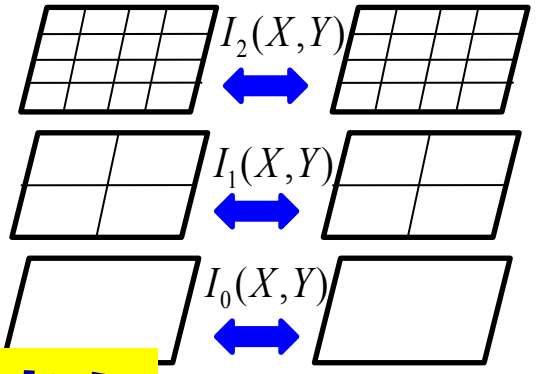


- 各レベルごとにヒストグラムインターセクションを求め，
  レベルごとに異なる重みで統合．SVMのカーネル関数とする．

$$k(X,Y)=\frac{1}{2^L}I_0(X,Y)+\sum_{l=1}^{L}\frac{1}{2^{L-l+1}}I_l(X,Y)$$

$$=\frac{1}{4}I_0+\frac{1}{4}I_1+\frac{1}{2}I_2 \quad (\text{in case of L}=2)$$

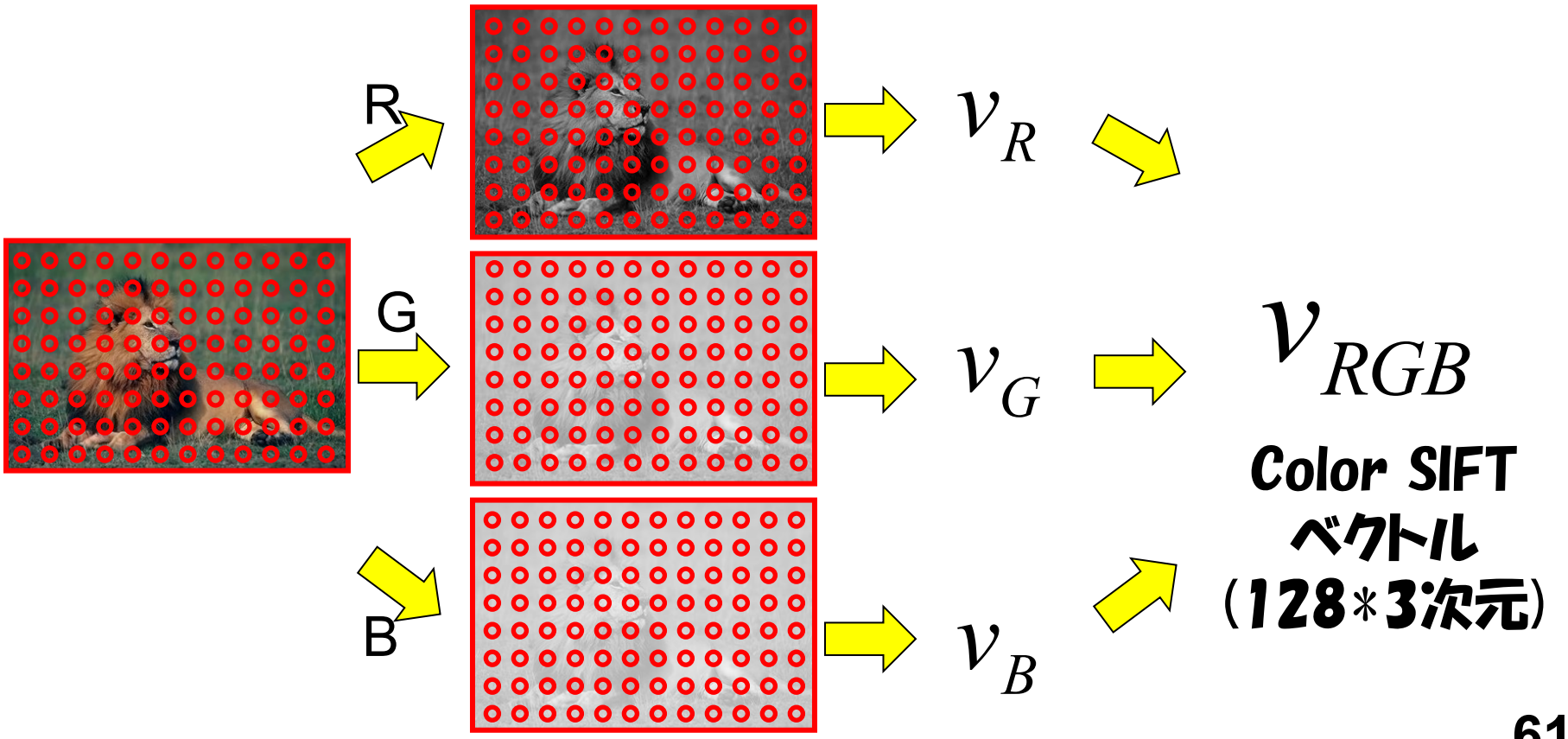
I_l : Histogram intersetion in level l



簡単！　でも，次元が21倍になるのが難点！

色情報の利用：Color SIFT

- 各特徴点についてRGB(HSV, Lab)の3つのSIFTベクトルを計算し，1つに結合．



他の種類の特徴との統合［Var07］
(Multiple Kernels Learning)

- 多種類特徴を統合するSVMのカーネル関数
  - 重み付き線形和カーネル関数による，
    Bag-of-features，色，形の統合．
    ＋ 重みの自動推定．
  - 各特徴のカーネルを $\mathrm{K}_1, \ldots, \mathrm{K}_{N_k}$ とすると，
    統合カーネルは，

$$\mathrm{K}_{\mathrm{opt}} = \sum_k d_k \mathrm{K}_k$$

ただし，$d_i (i = 1,.., N_k)$ は最適化問題を解いて求める．

- カーネルの重みを求めるのは，機械学習の研究では
  近年よく研究されている．

認識精度を上げるには，BoFのみでなく，様々な特徴量を
カテゴリーに応じて選択的に利用することが重要．

50種類食事カテゴリー分類

は人手で囲んだ食べ物の位置

各種類100枚ずつ　計5000枚

実験結果

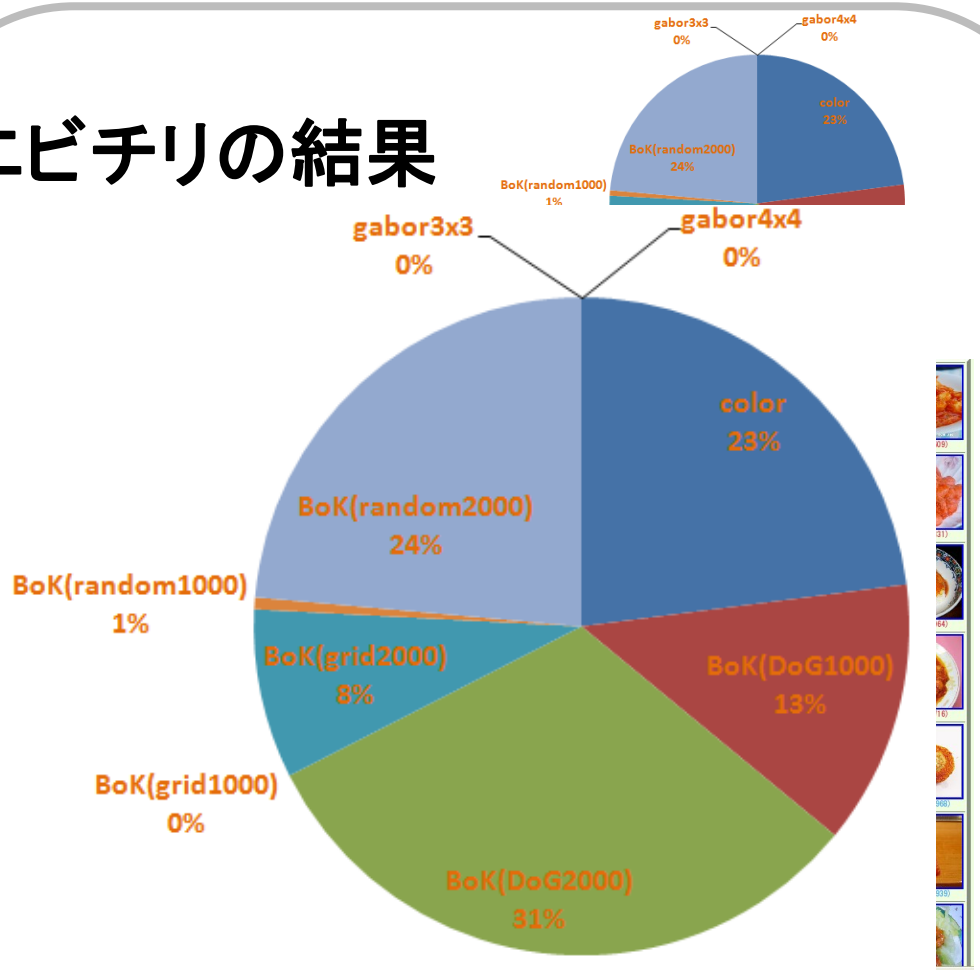
Multiple Kernel Learning法による特徴統合による結果と特徴単独による の50種類の平均分類率

特徴	平均分類率
Color	**38.18%**
BoF(dog2000)	**27.48%**
BoF(grid2000)	**27.68%**
BoF(random2000)	**29.70%**
MKL(mean-x2 distance)	**61.34%**

The University of Electro-Communications
Tokyo, JAPAN (UEC)

MKLで学習した特徴の重み

エビチリの結果



MKLが最善の特徴統合手法？

- Gehler et al. : On Feature Combination for Multiclass Object Classification [ICCV 2009]
  - 各種統合手法の比較
    - MKL
    - LPBoost: SVMを弱識別機にするboosting
    - CGBoost: LPBoostとは異なるboosting
    - Average: 複数カーネルの平均をカーネルとする
    - Product: 複数カーネルの積をカーネルとする

特徴統合の方法

Name	Test-time function	Coefficients	Training	Parameters
Averaging	$y(x) = \underset{c=1,\dots,\mathcal{C}}{\operatorname{argmax}} [\left(\frac{1}{F}\sum_{m=1}^{F} K_m(x)\right)^T \alpha_c + b_c]$	$\alpha \in \mathbb{R}^{\mathcal{C}\times N}$ $b \in \mathbb{R}^{\mathcal{C}}$	$(\alpha, b)_c$, ind.	C_c
Product	$y(x) = \underset{c=1,\dots,\mathcal{C}}{\operatorname{argmax}} [\left(\left(\prod_{m=1}^{F} K_m(x)\right)^{1/F}\right)^T \alpha_c + b_c]$	$\alpha \in \mathbb{R}^{\mathcal{C}\times N}$ $b \in \mathbb{R}^{\mathcal{C}}$	$(\alpha, b)_c$, ind.	C_c
MKL	$y(x) = \underset{c=1,\dots,\mathcal{C}}{\operatorname{argmax}} \sum_{m=1}^{F} \beta_m^c \left(K_m(x)^T \alpha_c + b_c\right)$	$\beta \in \mathbb{R}^{\mathcal{C}\times F}$ $\alpha \in \mathbb{R}^{\mathcal{C}\times N}$ $b \in \mathbb{R}^{\mathcal{C}}$	$(\alpha_c, b_c, \beta^c)_c$ ind.	C_c
CG-Boost	$y(x) = \underset{c=1,\dots,\mathcal{C}}{\operatorname{argmax}} [\sum_{m=1}^{F} K_m(x)^T \alpha_{c,m} + b_c]$	$\alpha \in \mathbb{R}^{\mathcal{C}\times F\times N}$ $b \in \mathbb{R}^{\mathcal{C}}$	$(\alpha, b)_c$, ind.	C_c
LP-β	$y(x) = \underset{c=1,\dots,\mathcal{C}}{\operatorname{argmax}} \sum_{m=1}^{F} \beta_m \left(K_m(x)^T \alpha_{c,m} + b_{c,m}\right)$	$\beta \in \mathbb{R}^{F}$ $\alpha \in \mathbb{R}^{\mathcal{C}\times F\times N}$ $b \in \mathbb{R}^{\mathcal{C}\times F}$	1. $(\alpha, b)_c$, ind 2. β, jointly	1. C_m 2. $\nu \in (0, 1)$
LP-B	$y(x) = \underset{c=1,\dots,\mathcal{C}}{\operatorname{argmax}} \sum_{m=1}^{F} B_m^c \left(K_m(x)^T \alpha_{c,m} + b_{c,m}\right)$	$B \in \mathbb{R}^{F\times\mathcal{C}}$ $\alpha \in \mathbb{R}^{\mathcal{C}\times F\times N}$ $b \in \mathbb{R}^{\mathcal{C}\times F}$	1. $(\alpha, b)_c$, ind 2. B, jointly	1. C_m, 2. $\nu \in (0, 1)$

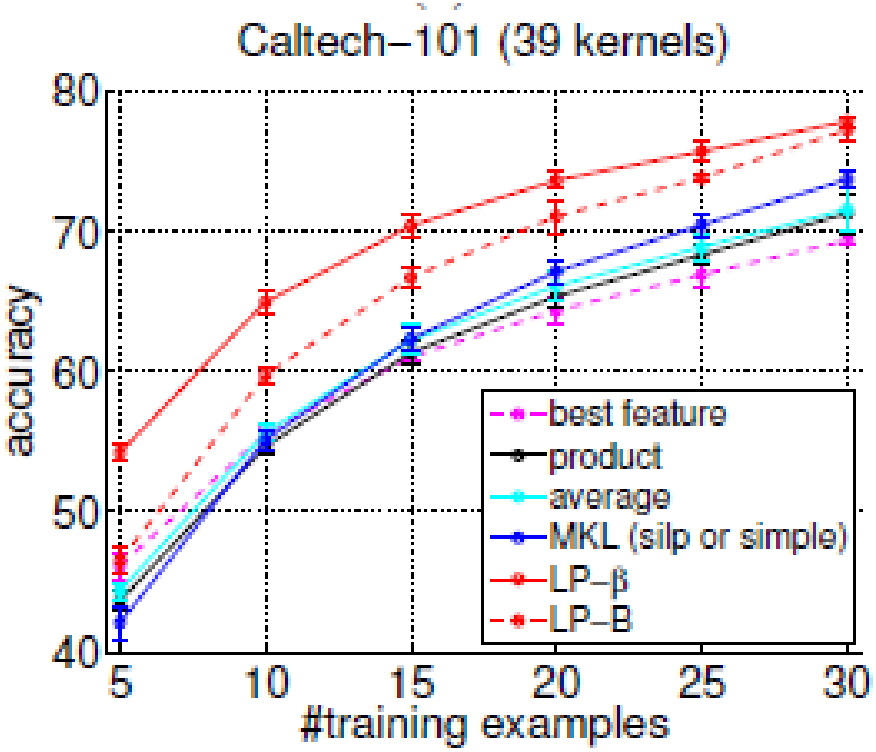
実験：Oxford Flowers

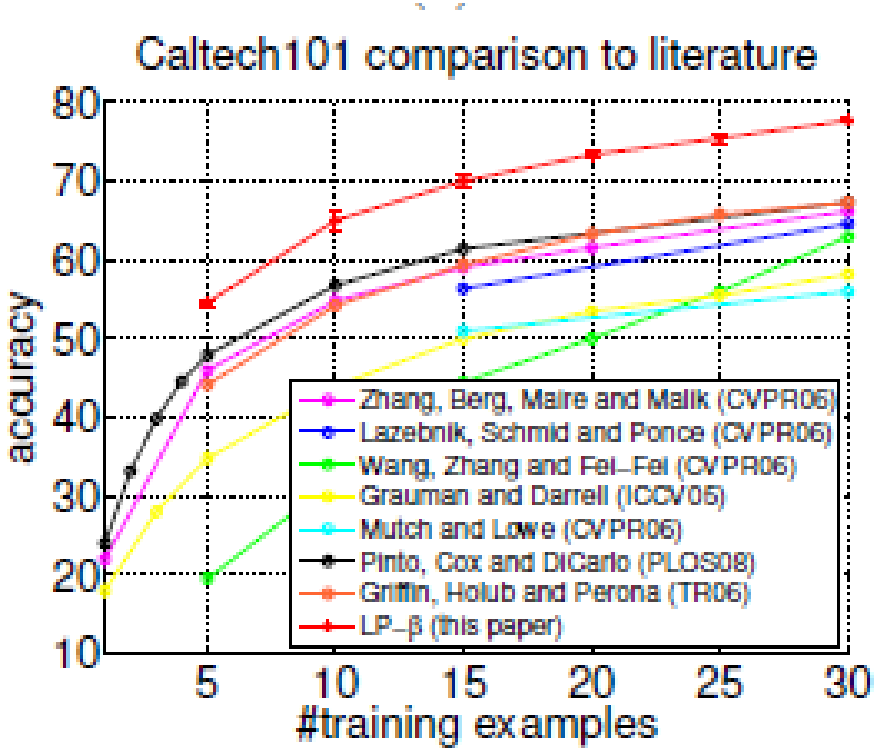
- オックスフォードの花データセット[13]
  - 17種類の花、各80枚

Single features			Combination methods		
Method	Accuracy	Time	Method	Accuracy	Time
Colour	60.9 ± 2.1	3	product	85.5 ± 1.2	2
Shape	70.2 ± 1.3	4	averaging	84.9 ± 1.9	10
Texture	63.7 ± 2.7	3	CG-Boost	84.8 ± 2.2	1225
HOG	58.5 ± 4.5	4	MKL (SILP)	85.2 ± 1.5	97
HSV	61.3 ± 0.7	3	MKL (Simple)	85.2 ± 1.5	152
siftint	70.6 ± 1.6	4	LP-β	85.5 ± 3.0	80
siftbdy	59.4 ± 3.3	5	LP-B	85.4 ± 2.4	98

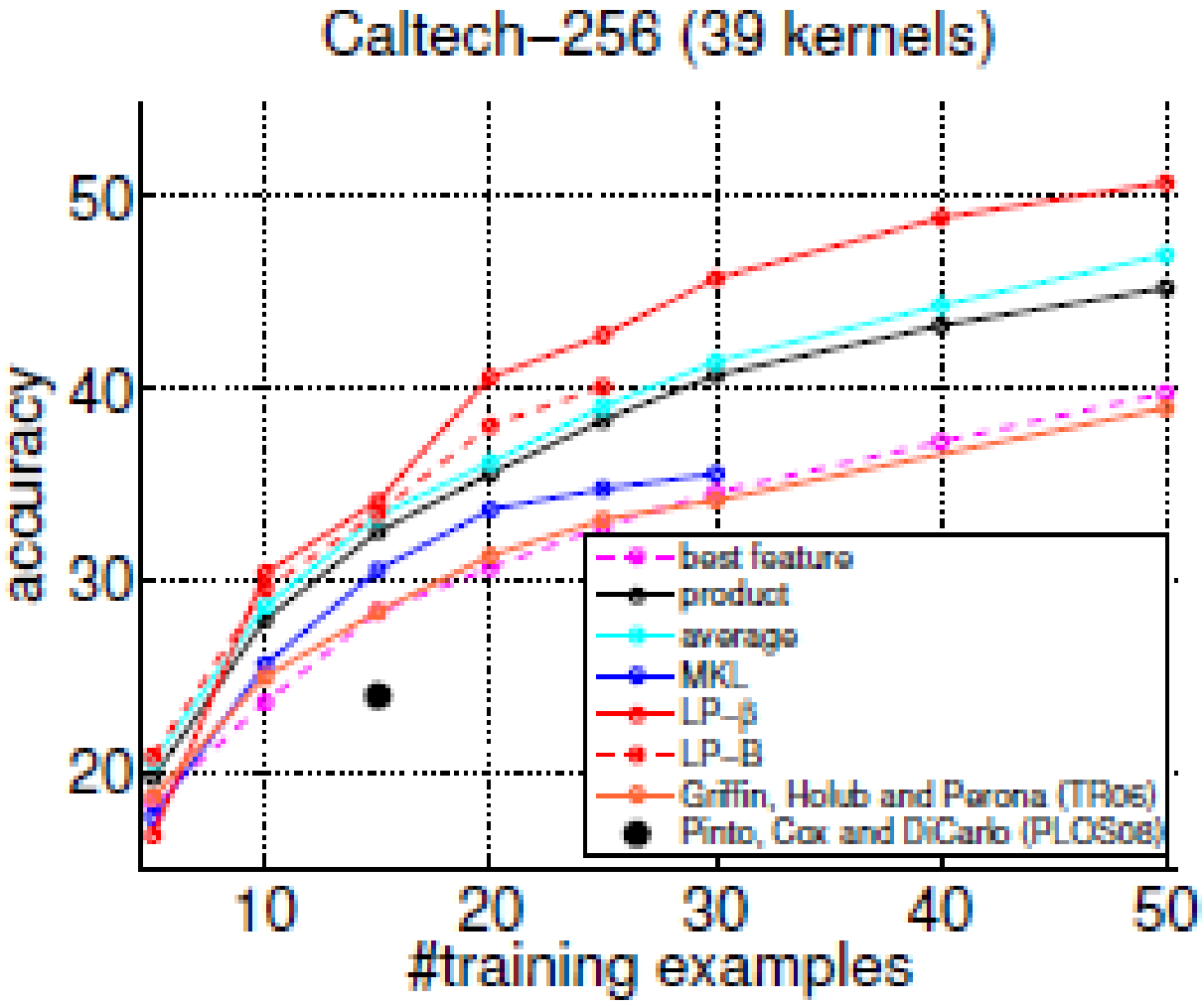
MKLがベストという訳ではなく，単純なproductも健闘．

Results for caltech-101





Results for Caltech-256



Bag-of-words用の手法の導入（テキスト解析手法の導入）

文書：単語の集合 → bag-of-words
画像：VWの集合　→ bag-of-visual-words

- Video Google [Siv03]
  キーワード検索手法（転置インデックス）の画像検索への応用
- 確率トピックモデルの画像への応用：
  元々はテキスト解析用bag-of-wordsを前提とする
  - PLSA (Probabilistic Latent Analysis)
  - LDA (Latent Dirichlet Allocation)
  - HDP (Hierarchical Dirichlet Process)

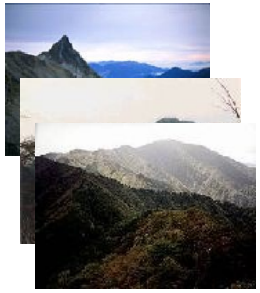
The University of Electro-Communications
Tokyo, JAPAN (UEC)

高次元でスパースなBoF向けの確率的クラスタリング: PLSA と LDA

- テキスト解析向けの確率トピックモデル
  - Bag-of-words表現された文書を確率的にトピック分類する
  - トピック数は，事前に指定する．K-meansと同じ．
  - トピックをz，文書（画像）をdとすると，各文書について P(z|d) が求まる
- Probabilistic Latent Semantic Analysis
  - (ヒストグラムは離散なので)混合多項分布によるモデル
  - $P(w,d) = P(d)\sum_z p(w|z)P(z|d)$ をEMでパラメータ推定
- Latent Dirichlet Allocation (判別分析ではありません！)
  - PLSAを改良．多項分布の代わりに混合ディリクレ分布．オーバーフィッティングを解消．データが多いとpLSAで十分．

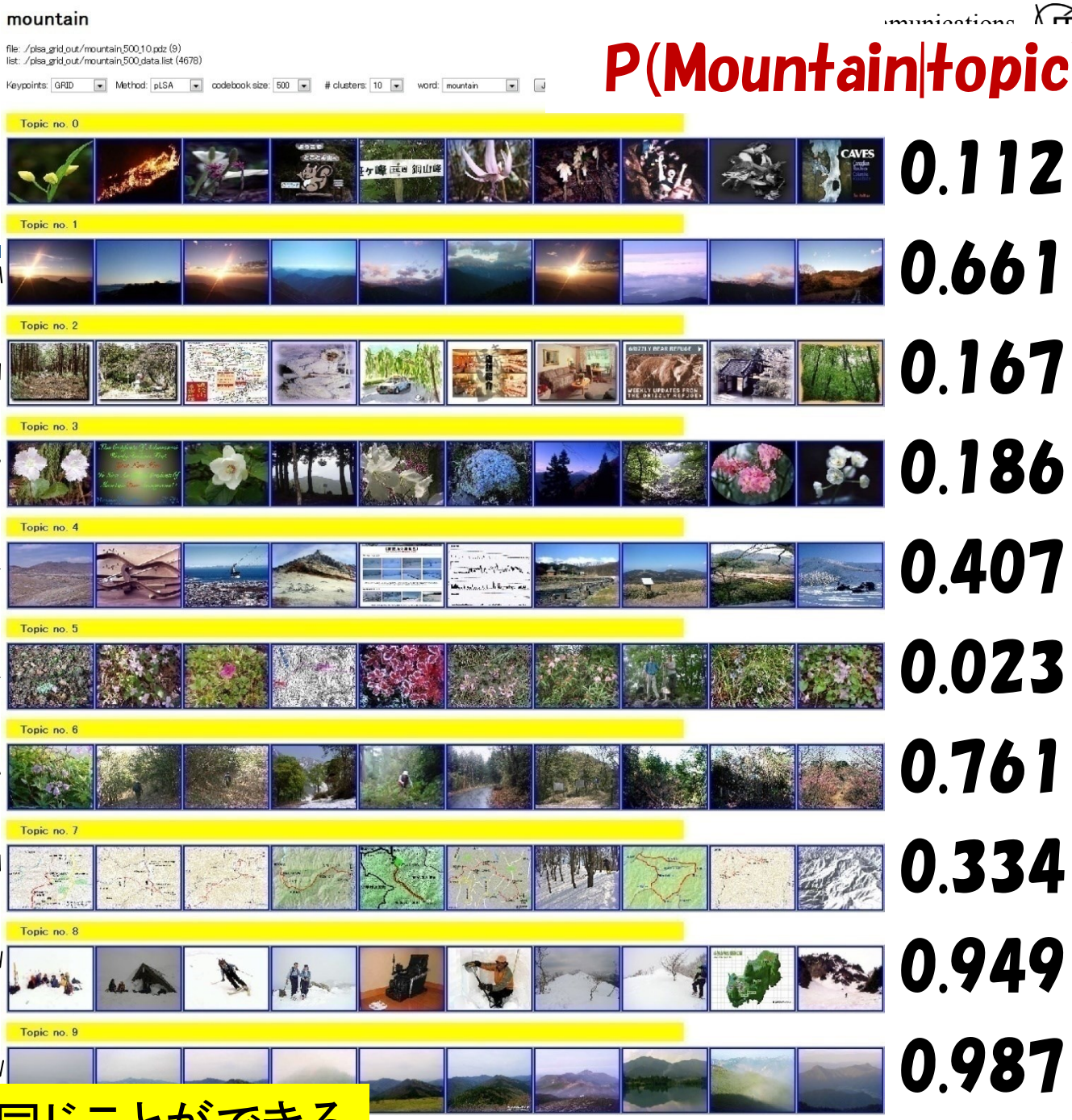
例：

- Mountain
  10 topics



正例

負例



P(Mountain|topic)

0.112

0.661

0.167

0.186

0.407

0.023

0.761

0.334

0.949

0.987

GMMによる認識と同じことができる

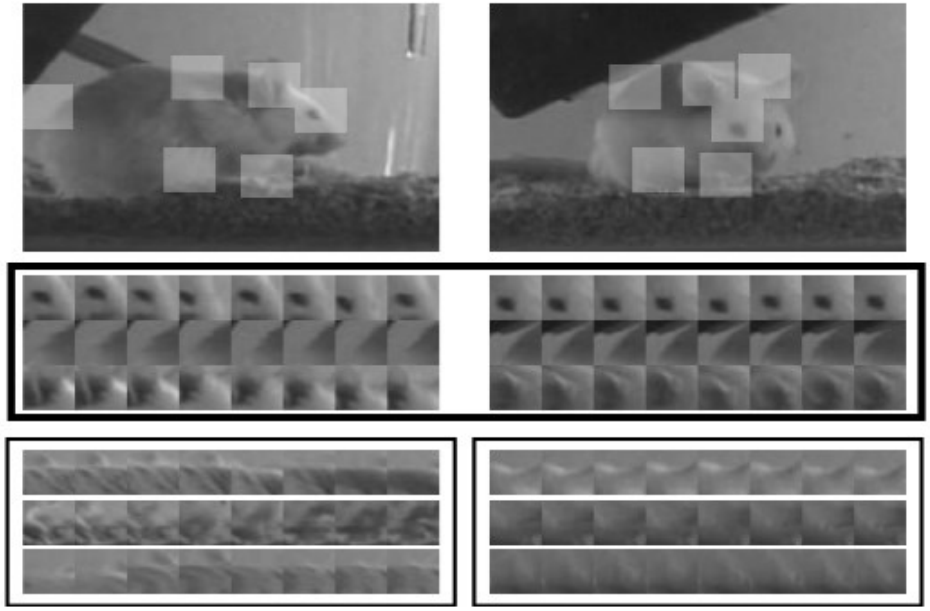
自動カテゴリー発見

- [Sivic ICCV05]
  - Bag-of-keypoints modelを採用.
  - テキスト解析においてtopic detectionに利用されるpLSA (probabilistic Latent Semantic Analysis) を用いて,topic detection を行う.
  - 4種類の物体が含まれる画像セットから4種類の物体を分離させることに成功.

もしスケールアップできれば，10億枚のWeb画像からクラス発見！？
実際は，スケールアップできない．クラス内外の変化を区別不可能．

時間方向(動画)への拡張：Bag-of-video-words

- 時空間特徴量をVQ化し，spatio-temporal visual wordsによって動画を表現
  - 動きを考慮した特徴点を抽出し，cuboidなど[画像＋動き]の特徴量を抽出.



KTHデータセット

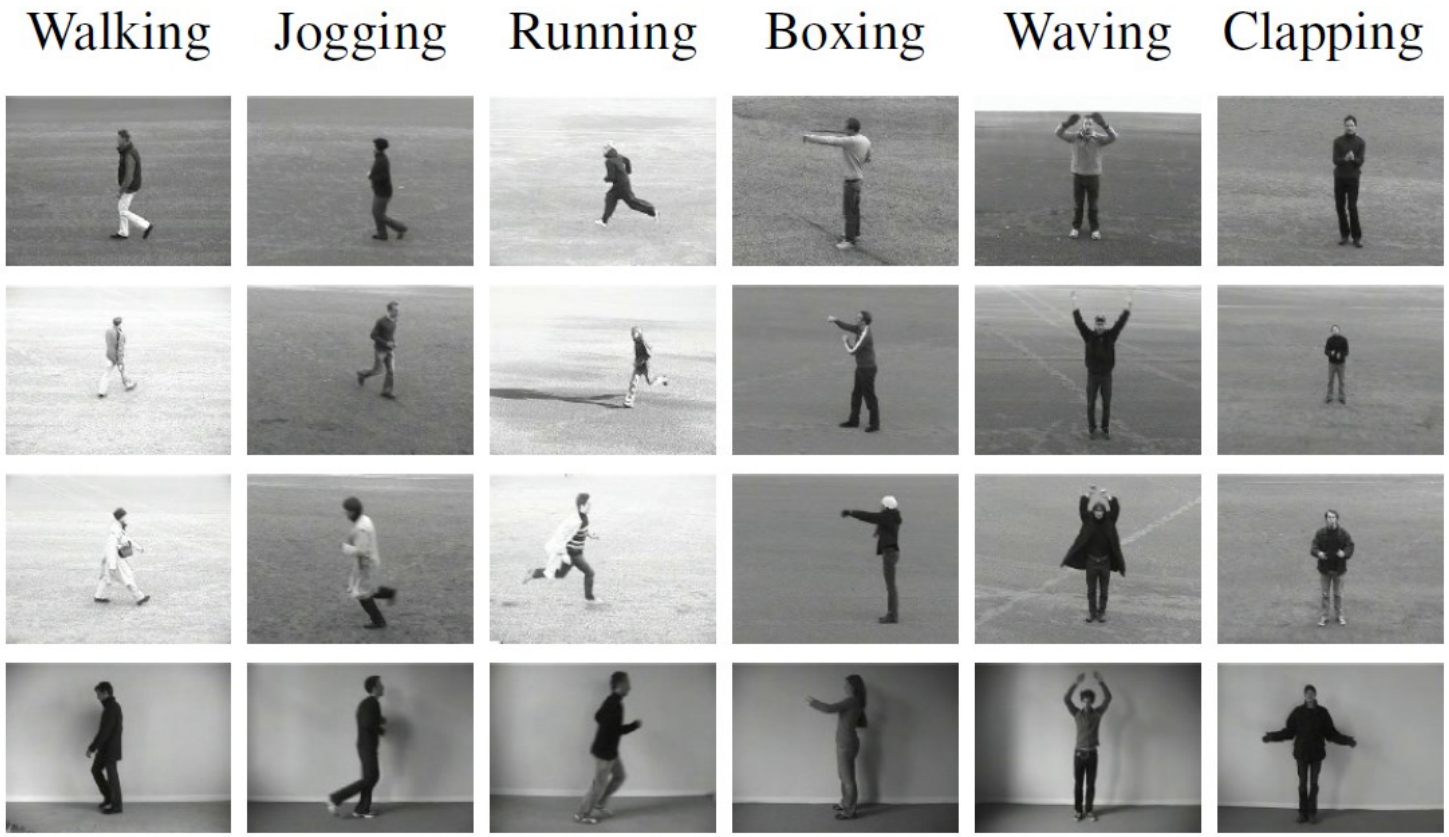


Figure 8. Sample frames from the KTH actions sequences. All six classes (columns) and scenarios (rows) are presented.

最新結果 [Lap08]

- 分類は SVM + χ2乗カーネル

Method	Schuldt et al. [15]	Niebles et al. [13]	Wong et al. [18]	ours
Accuracy	71.7%	81.5%	86.7%	**91.8%**

Table 3. Average class accuracy on the KTH actions dataset.

- 我々が実験中の結果は，約93％
- ICCV2009での最高は，97.1％

動作認識もBoFで可能．局所特徴以外は画像分類とほとんど同じ．

最新結果

	Walking	Jogging	Running	Boxing	Waving	Clapping
Walking	.99	.01	.00	.00	.00	.00
Jogging	.04	.89	.07	.00	.00	.00
Running	.01	.19	.80	.00	.00	.00
Boxing	.00	.00	.00	.97	.00	.03
Waving	.00	.00	.00	.00	.91	.09
Clapping	.00	.00	.00	.05	.00	.95

Table 4. Confusion matrix for the KTH actions.

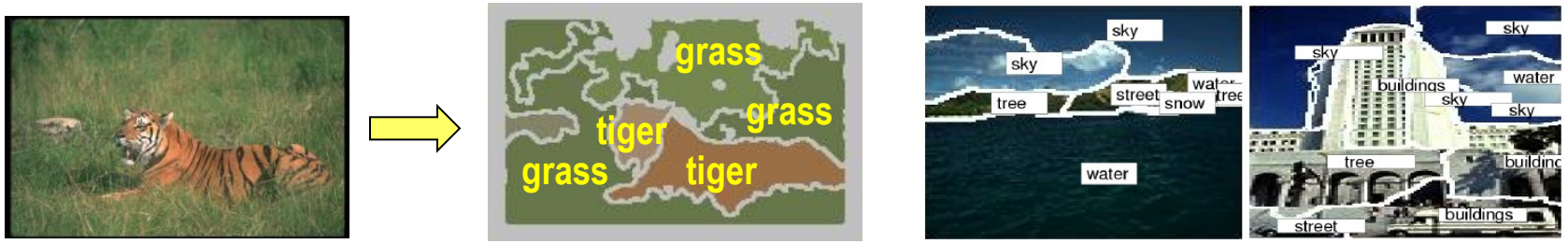
4. 一般物体検出

【参考文献】

Lampert, C. H., Blaschko, M. B. and Hofmann, T.: Beyond Sliding Windows: Object Localization by Efficient Subwindow Search, *CVPR 2008.*

一般物体の検出

- **画像ラベリング**：領域分割→分類



- カテゴリー**物体検出**：ウィンドウ探索

- オブジェクト**領域抽出**：認識＋領域分割

画像丸ごとのクラス分類は，かなりできるようになった．
(e.g. Caltech101 約80%) 次の課題は，物体位置検出．

The University of Electro-Communications
Tokyo, JAPAN (UEC)

元祖画像ラベリング：Word-image-translation model

[P.Duyguru, K.Barnard et al. ECCV 02]

- Statistical translation （統計的機械翻訳）
  - **大量の2ヶ国語テキスト(対訳コーパス)を学習データとし，確率モデルを学習し，翻訳を行う．文法知識が不要．**

I have a red pen in my pocket.
私は 赤い ペン を ポケット の中に 持っている.

大量の対訳テキスト

I have a blue eraser in my pencil case. ⇒ **未知の文が翻訳可能**

- Word-image-translation model
  - **大量のキーワード付画像を学習データとし，確率モデル構築**


lion, grass


panda, grass

大量の単語付画像 (bounding box不要)

⇒

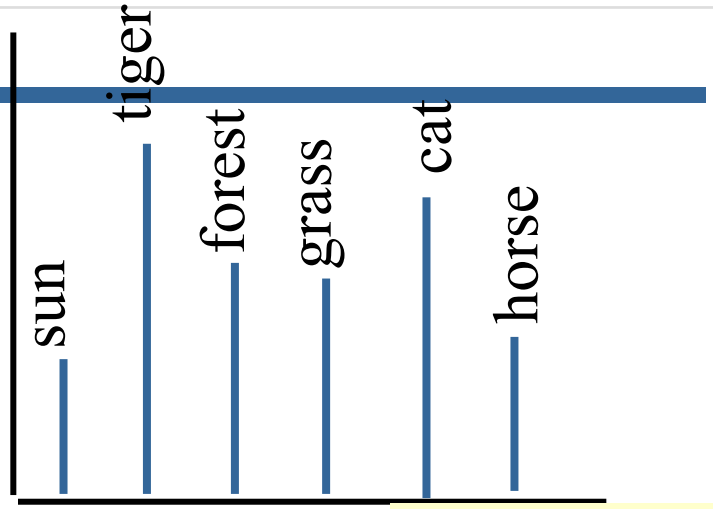
未知の画像に単語を付与することが可能

The University of Electro-Communications
Tokyo, JAPAN (UEC)

領域への確率モデルに基づくラベリング

領域分割

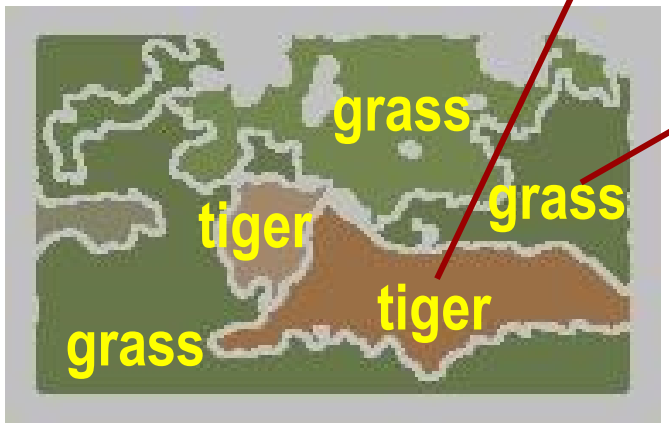
領域毎に特徴ベクトルを生成
P(word|region) を学習．GMMでモデル化

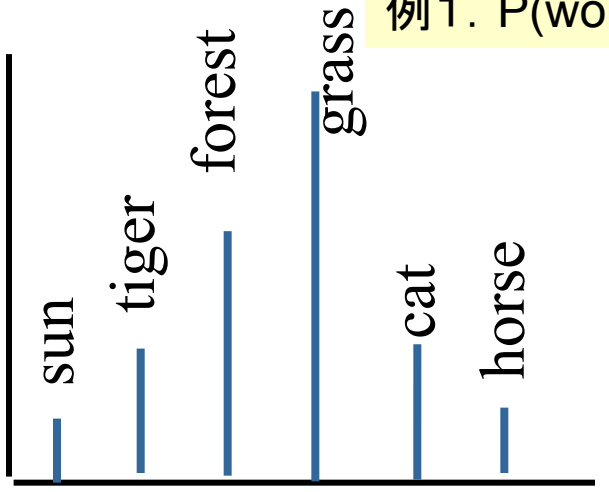


例1．P(word|region1)

未知画像の領域の特徴ベクトルから
P（word|region)の値を計算．

領域毎に，
テクスチャ，
色，形状
などの
特徴量
を抽出

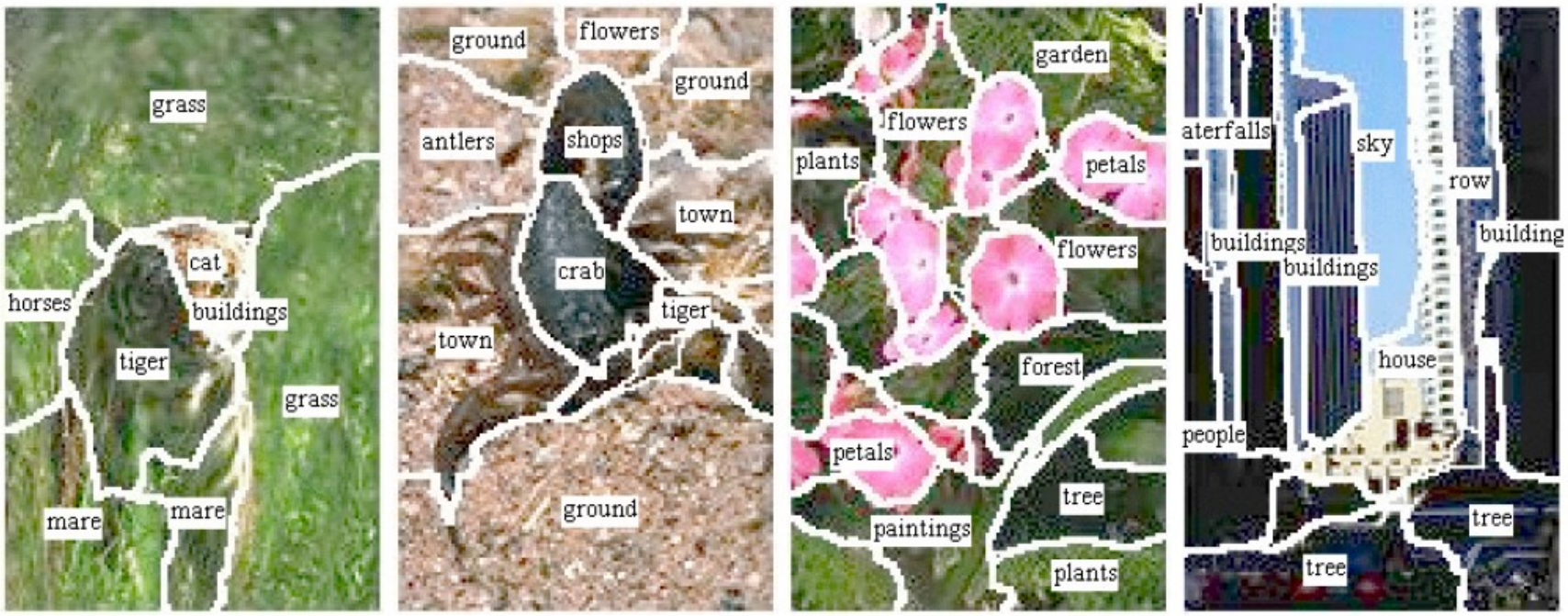




例2．P(word|region2)

領域分割は事前に行う．

画像ラベリングの例


grass
cat
horses
buildings
tiger
grass
mare
mare
ground
flowers
ground
antlers
shops
town
crab
tiger
town
ground
garden
flowers
plants
petals
flowers
forest
petals
tree
paintings
plants
aterfalls
sky
row
building
buildings
buildings
house
people
tree
tree

The University of Electro-Communications
Tokyo, JAPAN (UEC)

画像ラベリング: Multiple Instance Learning

- Multiple instance learning
  - positive bag : positive instance を含む
  - negative bag : positive instance を含まない

positive bag

ライオンの
positive instance

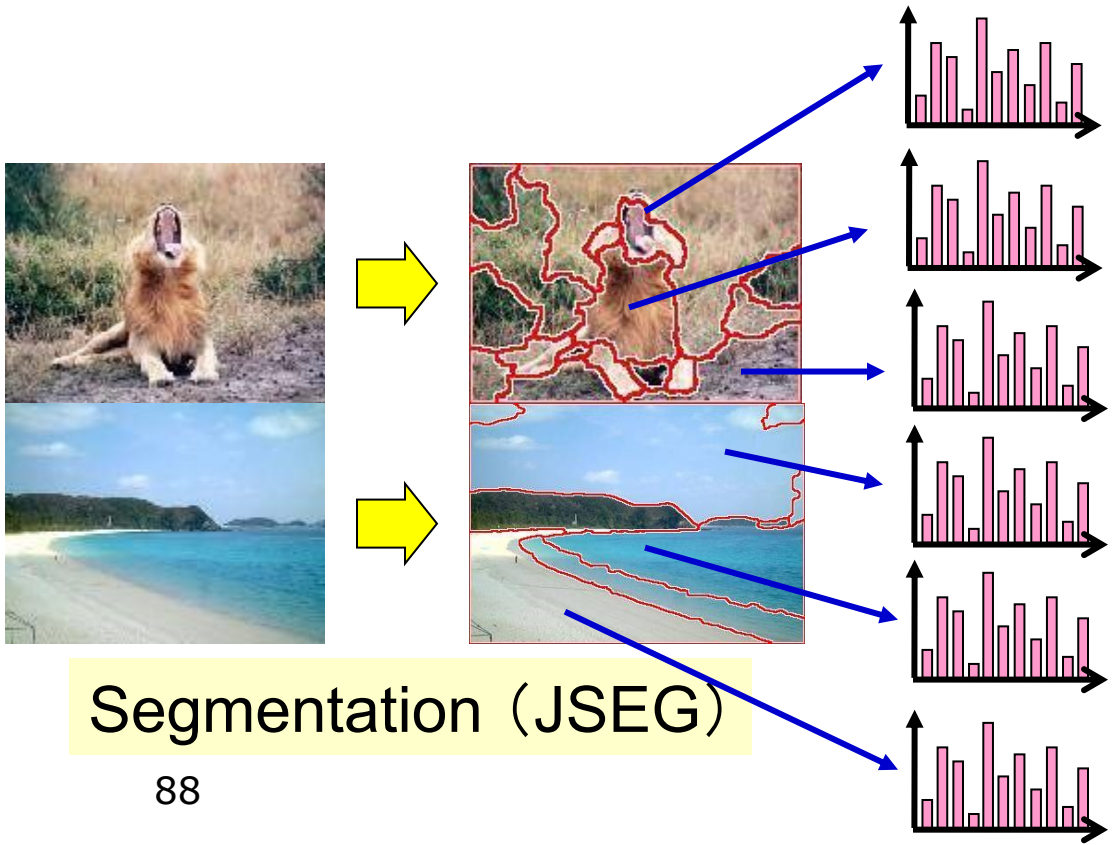
negative bag

ライオンの
positive instance
はない.

- Positive/negative bags より pos. inst. を求める
- 確率モデルよりも，判別手法の方が性能が上.
  - Mi-SVM , Sparse Multiple Instance SVM
    (Diverse Density (DD) はあまり使われない.)

Region-based BoF

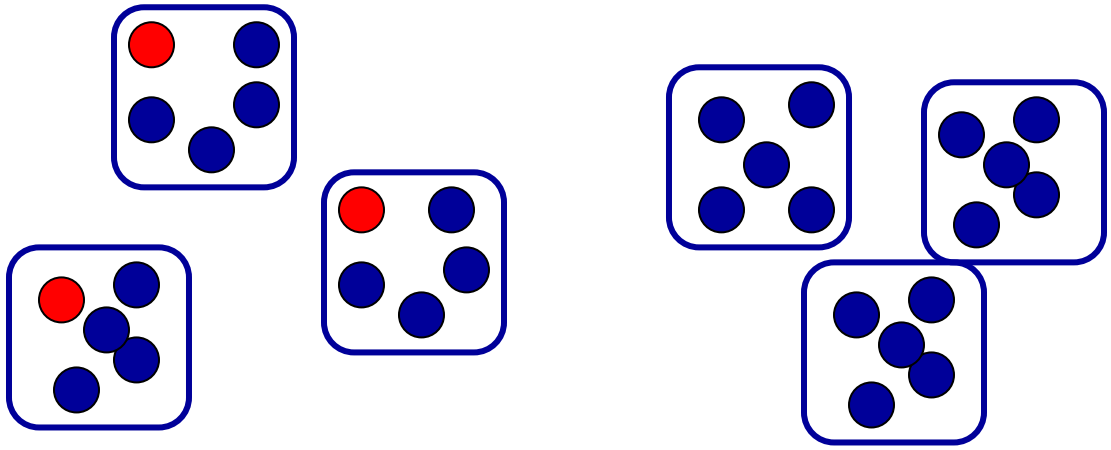
- Divide each image into regions by JSEG
  (8 regions on the average)



Bag-of-features (BoF) histograms (1000-dim)

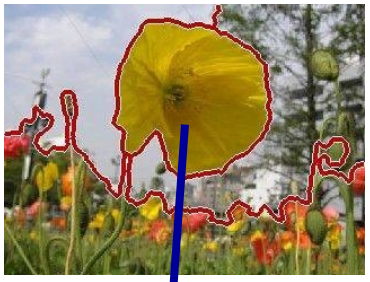
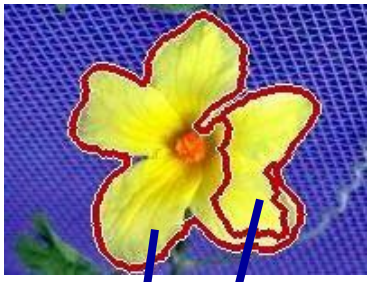
Multiple Instance Setting

- Positive bags / Negative bags

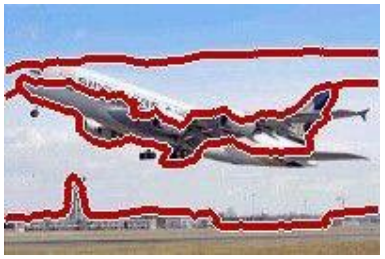


positive ins. (foreground)

negative ins. (background)

Positive instances of "flower"

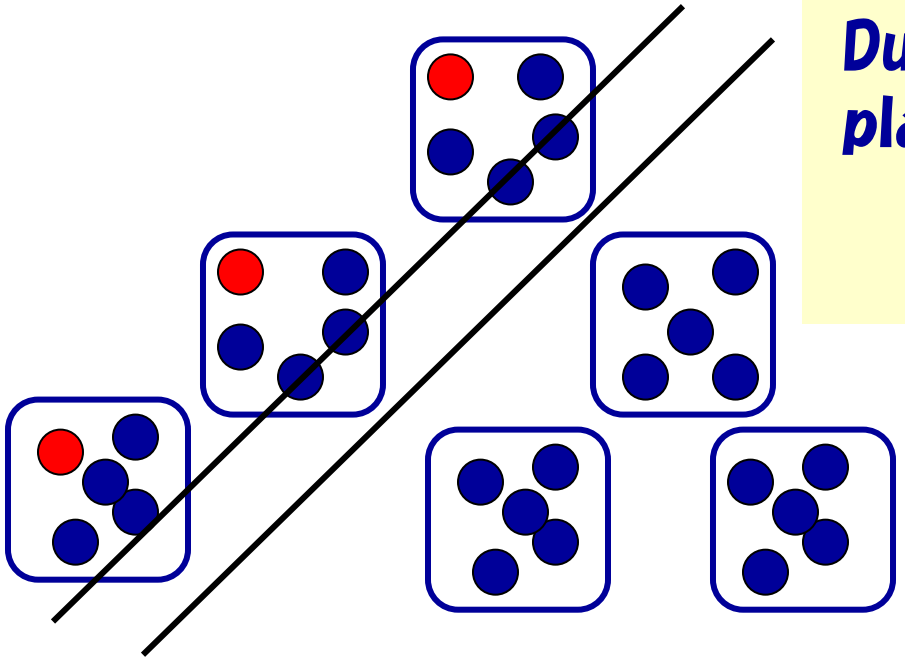
The rest of regions are negative regions.

pseudo-training images

random images

mi-SVM [Andrew et al. NIPS 03]

- Apply soft-margin SVM iteratively
  - Training → classifying → training → classifying → ……

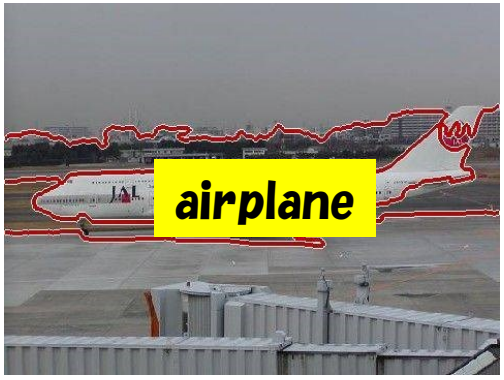


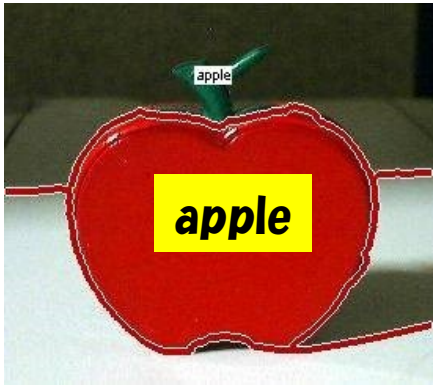
During the iteration, the hyper-plane is approaching the optimal plane to discriminate positive instances from negative ones.

- positive ins. (foreground)
- negative ins. (background)

Final Image Re-ranking

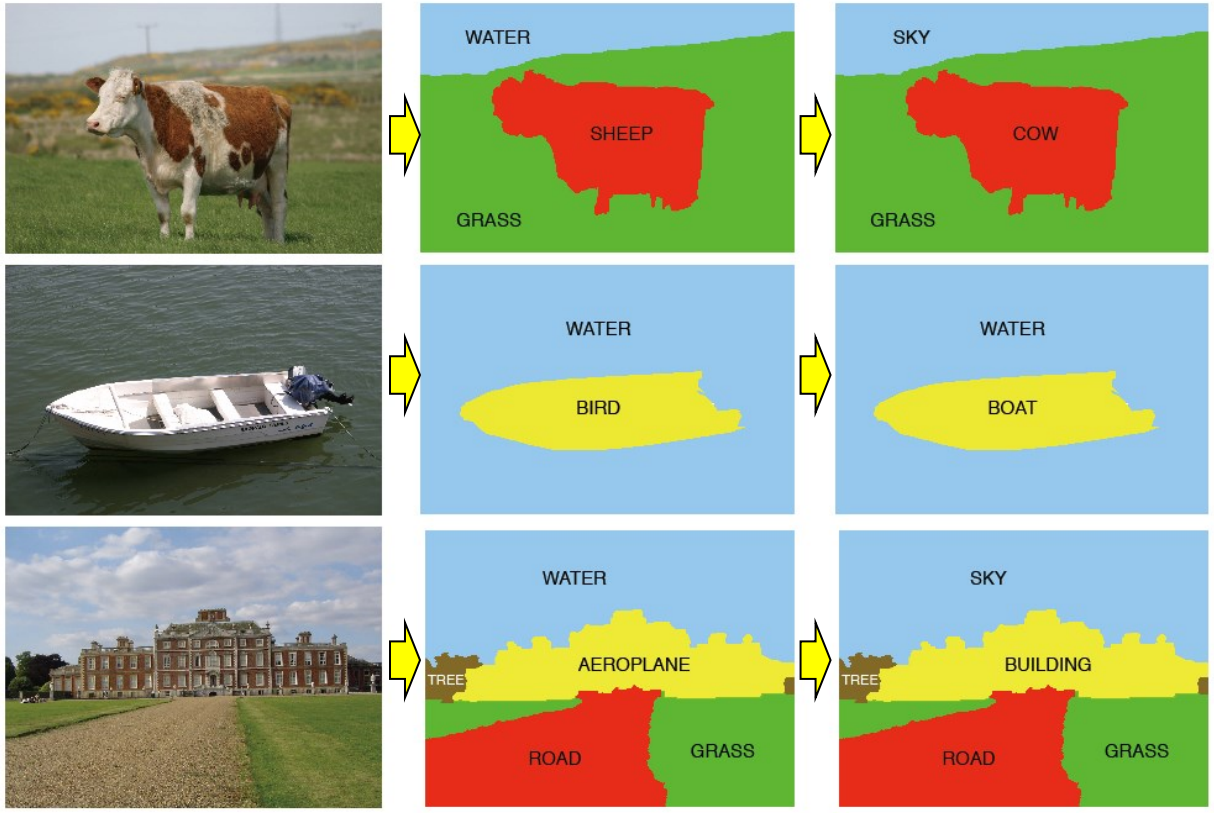
- Regard the best SVM output score of the regions within an image as the score of the image
  - An image having one positive region at least is a positive image !
- Rank images based on the scores





対象の検出へのBoFの利用：領域分割との組み合わせ［Rab07］

- 領域分割し，領域毎にBoFベクトルを作成，領域毎に分類．最後に共起関係より修正．



［Rab07］より
図を引用

領域分割は，
Normalized Cuts.
を利用．

共起関係は
確率モデル（MRF）
によって表現．

Sliding Window

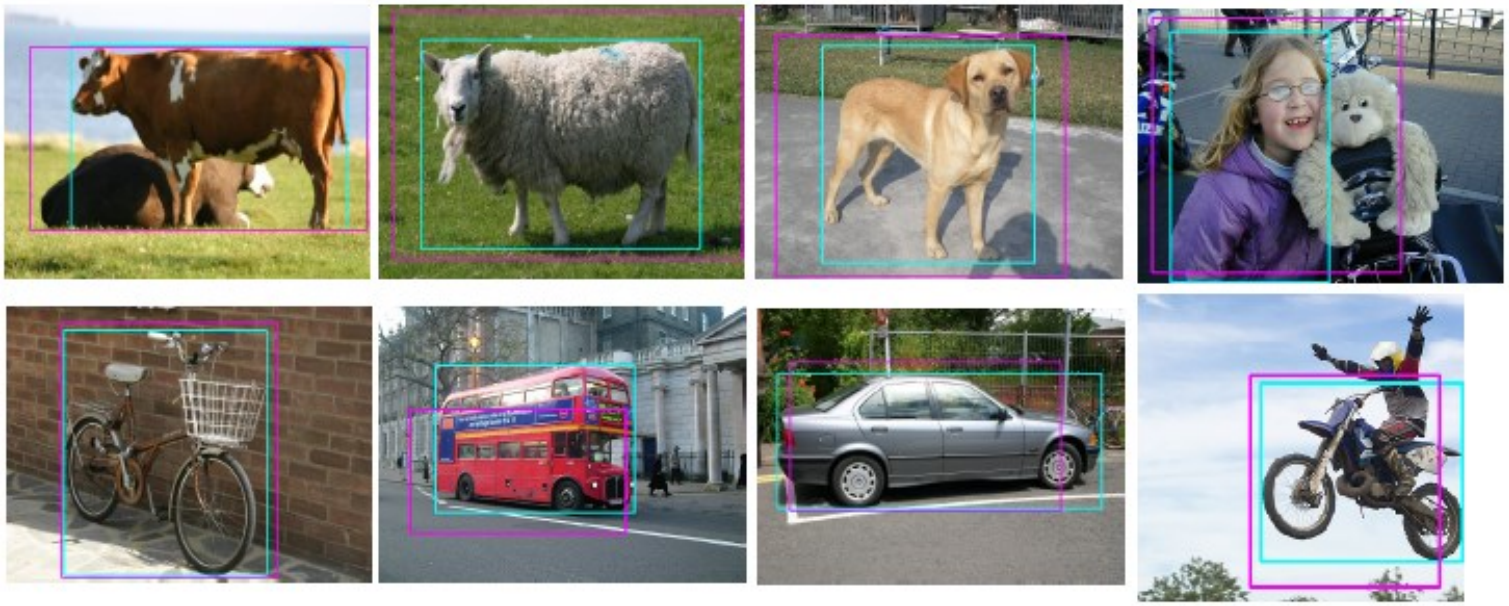
- 位置検出の基本は，sliding window

  例）顔検出(Viola-Jones detector)，HoGによる人物検出

各windowに対しては画像全体に対する手法(ただし学習画像はBB付き)
でよいが，計算コストは非常に大きい！！

Pascal VOC 2009

- Classification (画像全体分類) よりも Detection (物体検出)の方が中心タスク.



Sliding windowsの高速化手法

- Efficient Subwindow Search (ESS) [Lampert et al. 2008] 線形SVMのみ. 線形性を利用.
  - 分枝限定法によって, high score BBを探索する
- Jumping window [Chum et al. 2007]
  - 対象カテゴリーに関係によく表れるwordsの位置周辺を探索
  - Discriminability(word) = (#target obj. containig w)/ (#obj. containing w)
  - Generate pairs of high D(w) words and generate rectangles, aggregate them.

The University of Electro-Communications
Tokyo, JAPAN (UEC)

高速Windowサーチ：Efficient Subwindow Search [lam08]

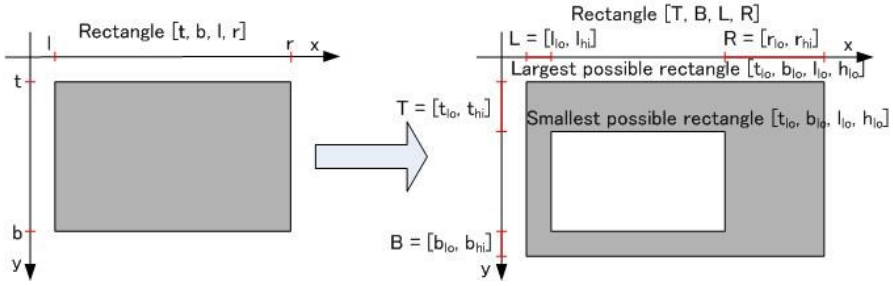
- **評価関数の計算に積分画像を導入**
  - 高速に，評価関数（線形SVM）が計算可能
- **バウンティングボックス(BB)の表現**
  - 座標でなく幅を持ったBBとして表現→[T、B、L、R]

  $T=[t_{min}, t_{max}]$、$B=[b_{min}, b_{max}]$、$L=[l_{min}, lmax]$、$R=[r_{min}, r_{max}]$
- **分枝限定法を使用する**
  - 走査領域を反復に分割して物体の位置を検出

ESSの挙動

ESSの特徴

- 走査領域の表現
  - 分割する際の計算を容易にするため、上下境界の幅で記述する



- 評価関数
  - 上限かつ収束部の線形カー

$$f^{upper}(R_x) \geq \max_{R \in R_x} f(R) \qquad (1)$$

$$f^{upper}(R_x) = f(R) \qquad s.t. R_x = R \qquad (2)$$

ESSアルゴリズムの特徴①
積分画像の導入

The University of Electro-Communications
Tokyo, JAPAN (UEC)

- **SVMの評価関数**

$$f(\vec{h}) = \sum_{i=1}^{n} \alpha_i y_i k(\vec{h}_i, \vec{h}) + \beta$$

α_i, β：学習パラメータ
$y_i, \vec{h}_i$：学習データ
$\vec{h}$：未知画像ベクトル

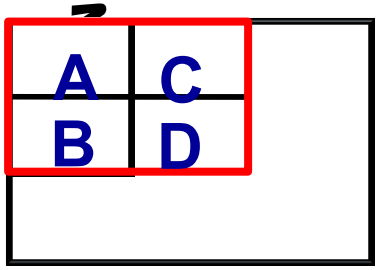
- **線形カーネルの場合、内積になる**

$$f(\vec{h}) = \sum_{i=1}^{n} \alpha_i y_i (\vec{h}_i \bullet \vec{h})$$

$$= \vec{w} \bullet \vec{h} = (\vec{w}^{+} + \vec{w}^{-}) \bullet \vec{h}$$

ただし、$\vec{w} = \sum_{i=0}^{n} \alpha_i y_i \vec{h}_i$

- **「窓」を分解すると，ヒストグラムが分解され，内積も分解され**

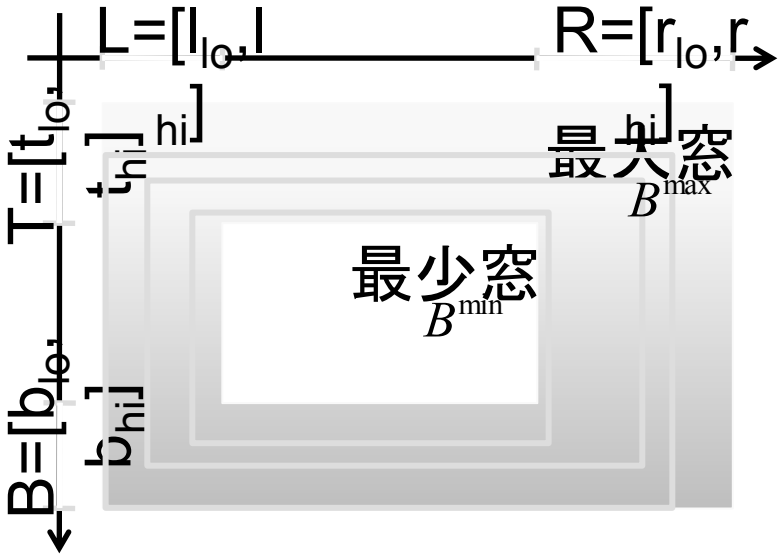


$$f(\vec{h}) = \vec{w} \bullet \vec{h} \quad = \vec{w} \bullet (\vec{h}_A + \vec{h}_B + \vec{h}_C + \vec{h}_D)$$

$$= \vec{w} \bullet \vec{h}_A + \vec{w} \bullet \vec{h}_B + \vec{w} \bullet \vec{h}_C + \vec{w} \bullet \vec{h}_D$$

ESSアルゴリズムの特徴②
バウンディングボックスの表現

- 通常の場合、座標をそのまま(上t、下b、左l、右r)
- ESSでは、幅で表現する→[T、B、L、R]
  - T=[t_{min}、t_{max}]、B=[b_{min}、b_{max}]、L=[l_{min}、lmax]、R=[r_{min}、r_{max}]



$$f^{+}(B) = \vec{w}^{+} \bullet \vec{h}$$

$$f^{-}(B) = \vec{w}^{-} \bullet \vec{h}$$

$B \in \beta : \beta$は窓の集合とすると、

$$f_{upper}(\beta) \geq \max f(B)$$

$$f_{upper}(\beta) = f^{+}(B^{\max}) + f^{-}(B^{\min})$$

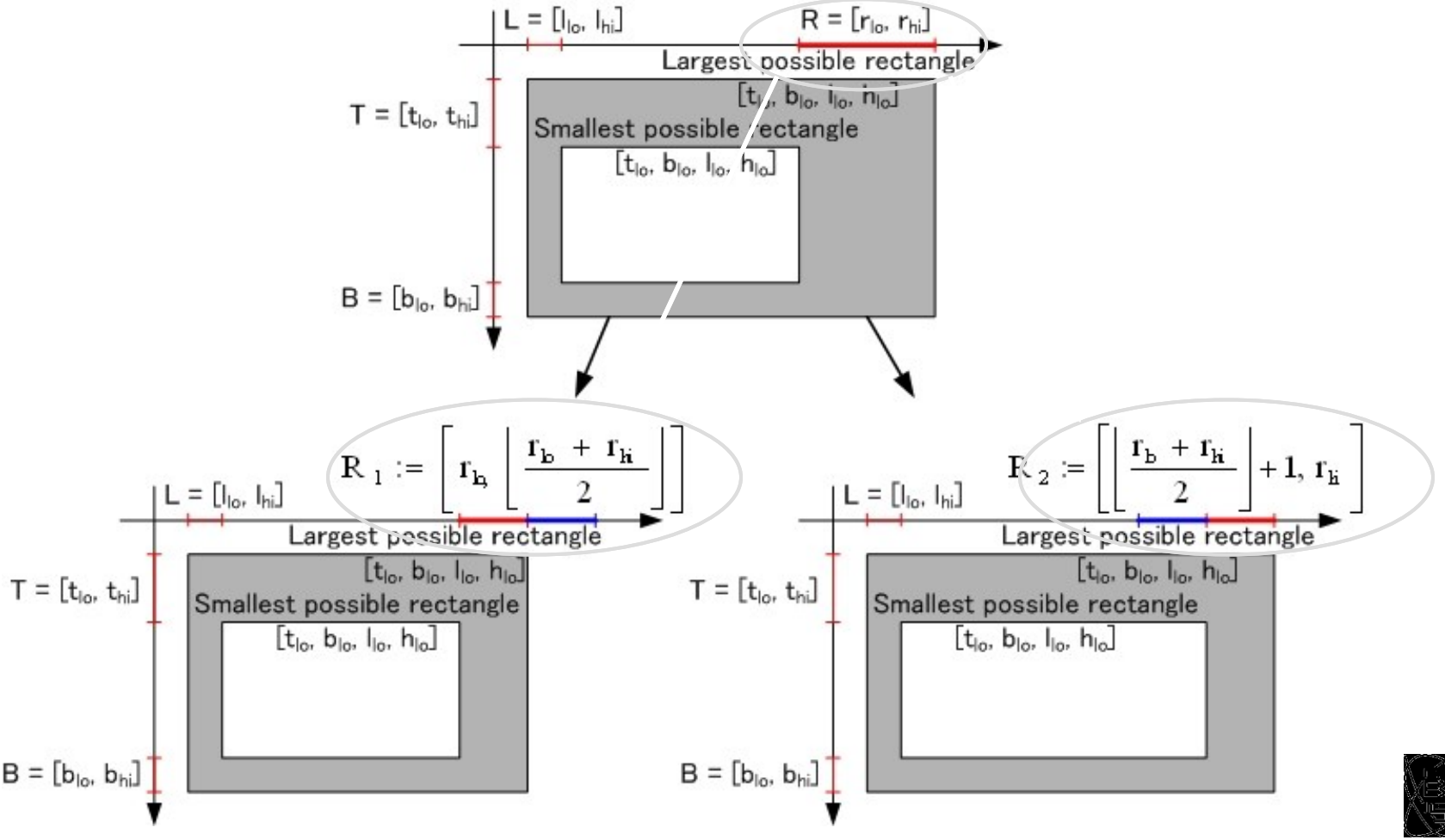
$B^{\max}$: 最大窓ヒストグラム

$B^{\min}$: 最少窓ヒストグラム

The University of Electro-Communications
Tokyo, JAPAN (UEC)

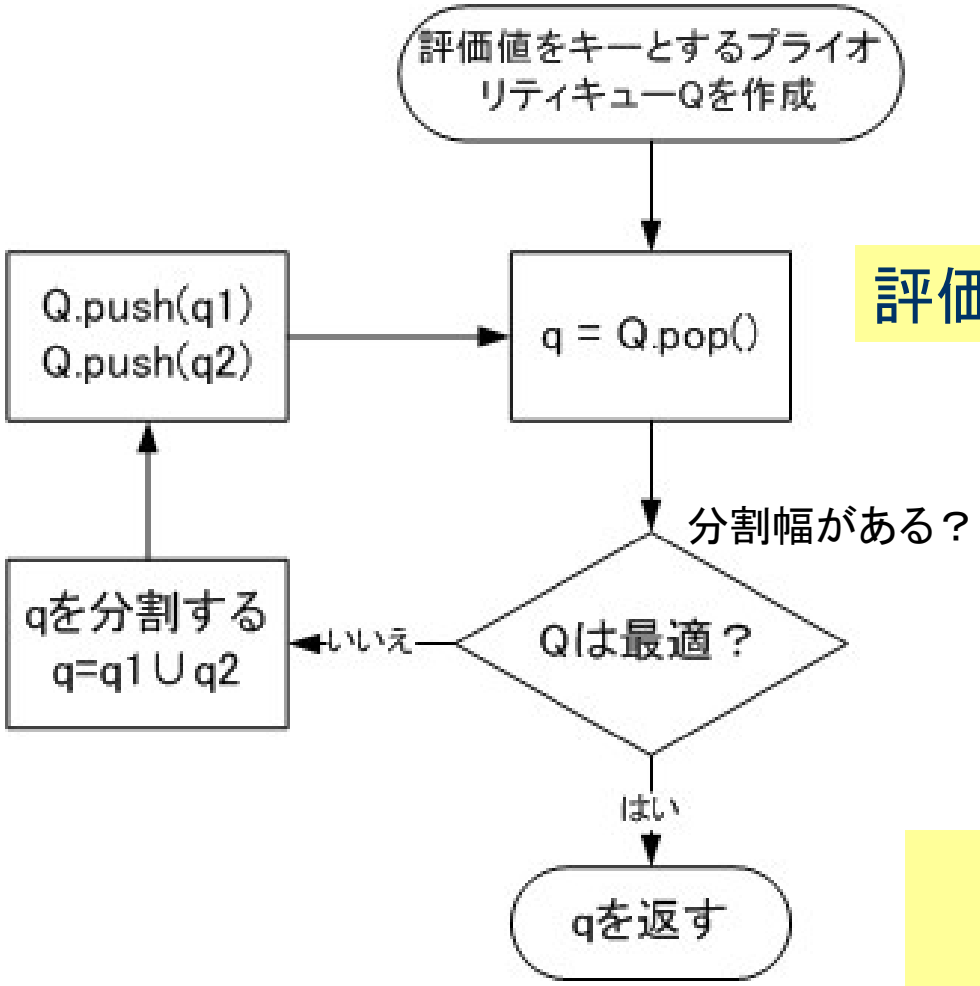
ESSアルゴリズムの特徴
バウンディングボックスの分割

- 領域分割：幅が最大の要素が対象



The University of Electro-Communications
Tokyo, JAPAN (UEC)

ESSアルゴリズムの特徴③
分枝限定法による最高評価値のBB探索



Qの内容：検出状態
- 評価値
- ボウンディングボックス

評価値の上限が最大のBBを探索

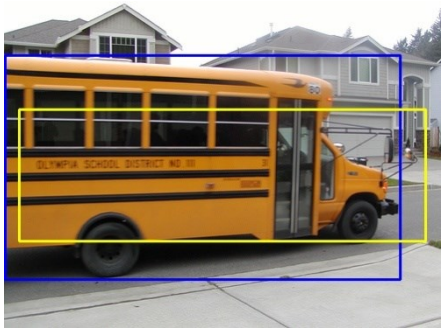
幅が0のBBの上限値
= ベストBBの評価値

実験内容

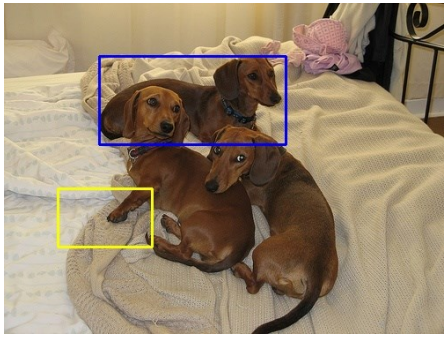
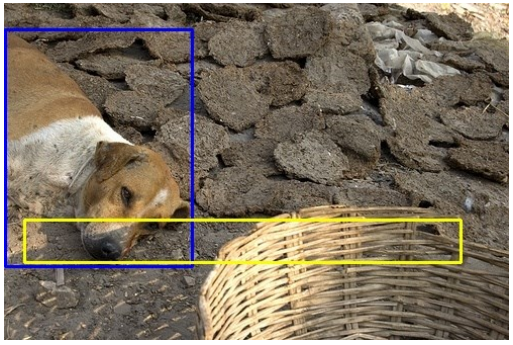
- 実験データ
  - Pascal VOC 2009のデータセット
    - 20種類のクラス(Aeroplane,Cat,Dog,Tvmo
    - 学習データ　749枚
    - 実験データ　750枚
- 実験環境
  - Xeon 2.66 dual core 2GB
  - SIFT特徴検出にSIFT++
  - SVMモデル作成にSVM_light
  - ESSの作成にESS-1_1

実験結果

- 成功例



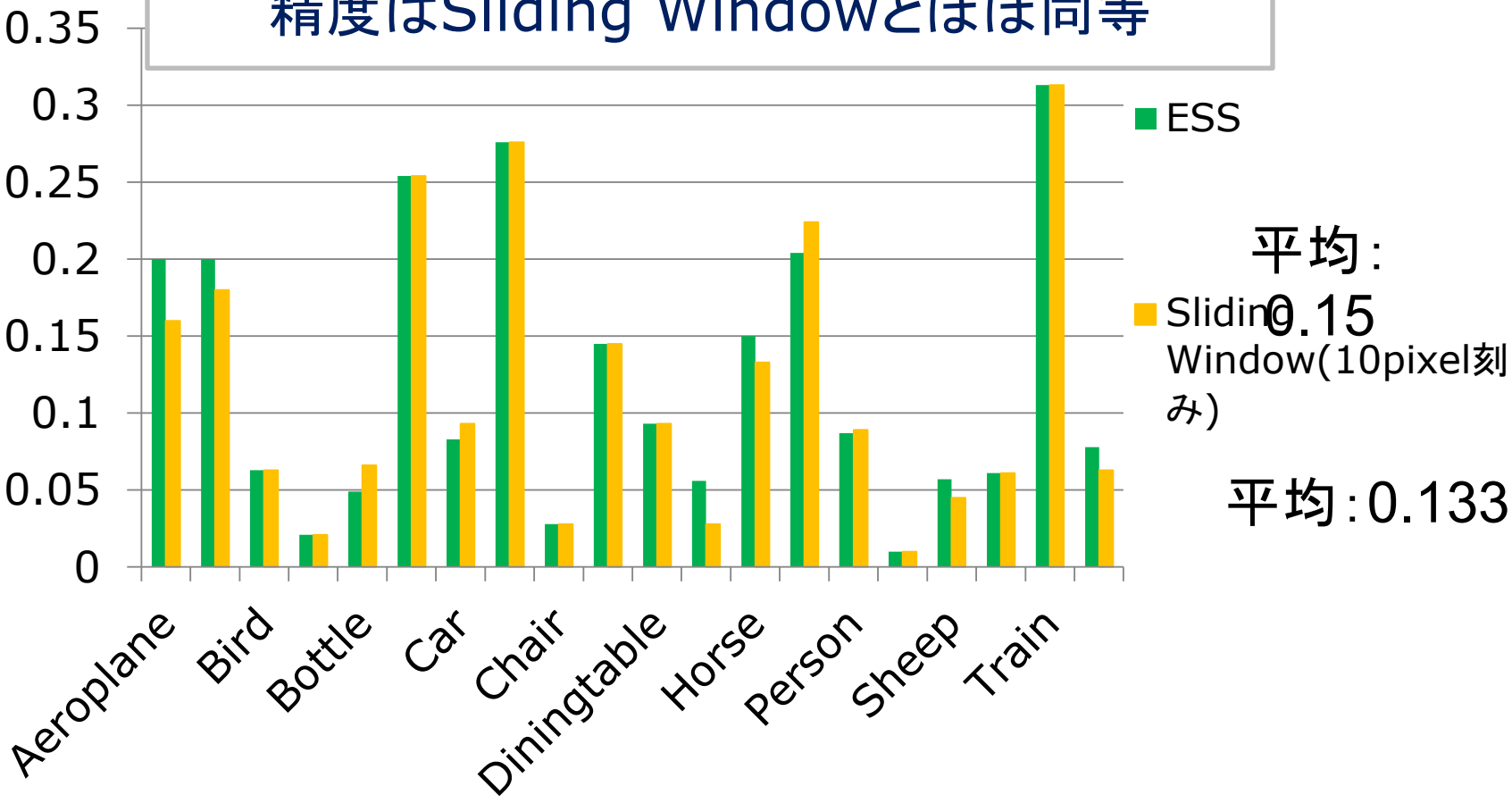


実験結果

- 各クラスの再現率(%)

精度はSliding Windowとほぼ同等

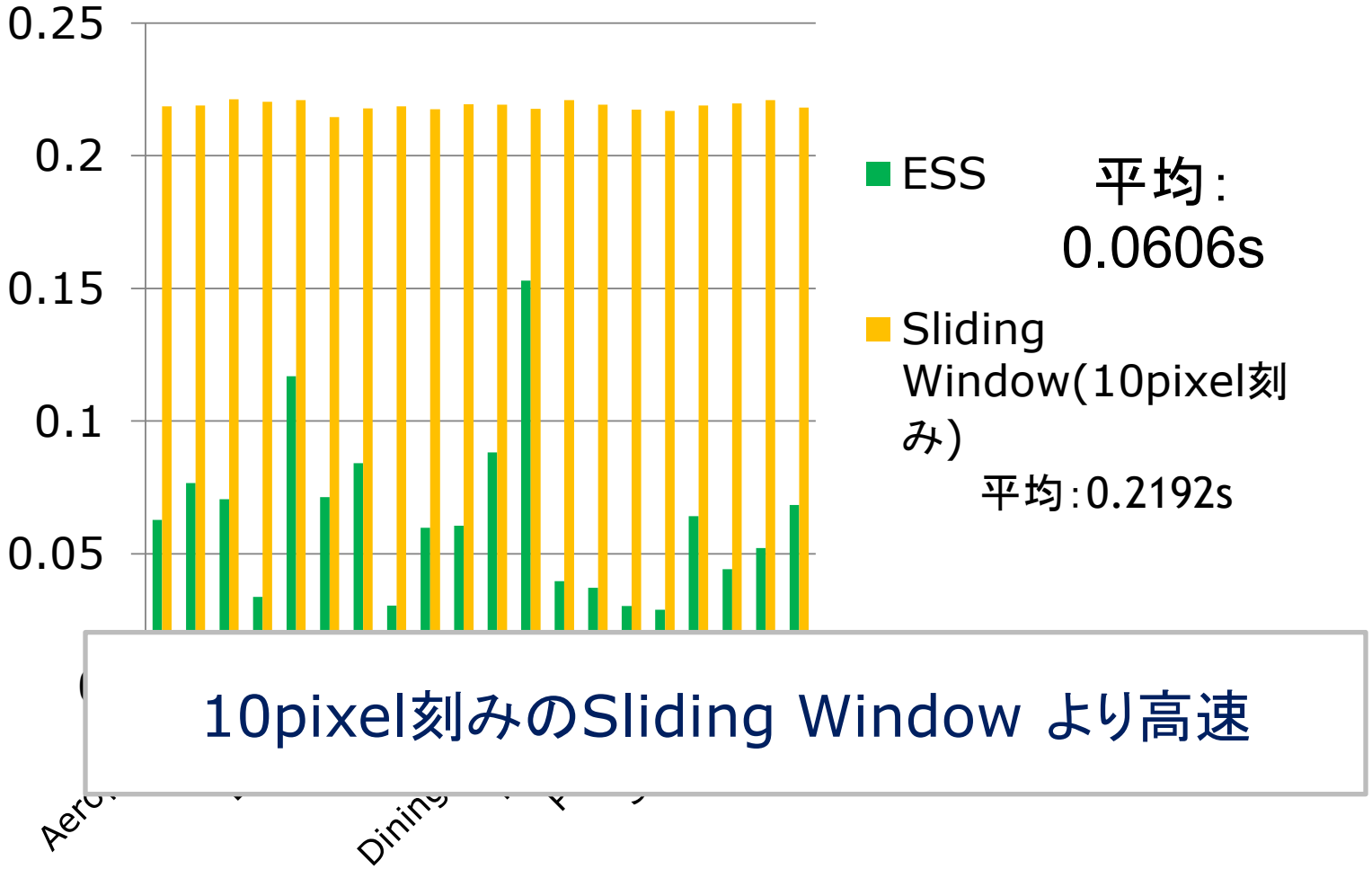


平均:0.15

平均:0.133

実験結果

- 平均時間(s)



10pixel刻みのSliding Window より高速

ESSの改良

- Bounding box の中心部分のVisual word を事前に照合し探索候補を削減し，高速化 [Lehmann et al. 2009]
- Bounding polygons [Yeh et al. 2009]

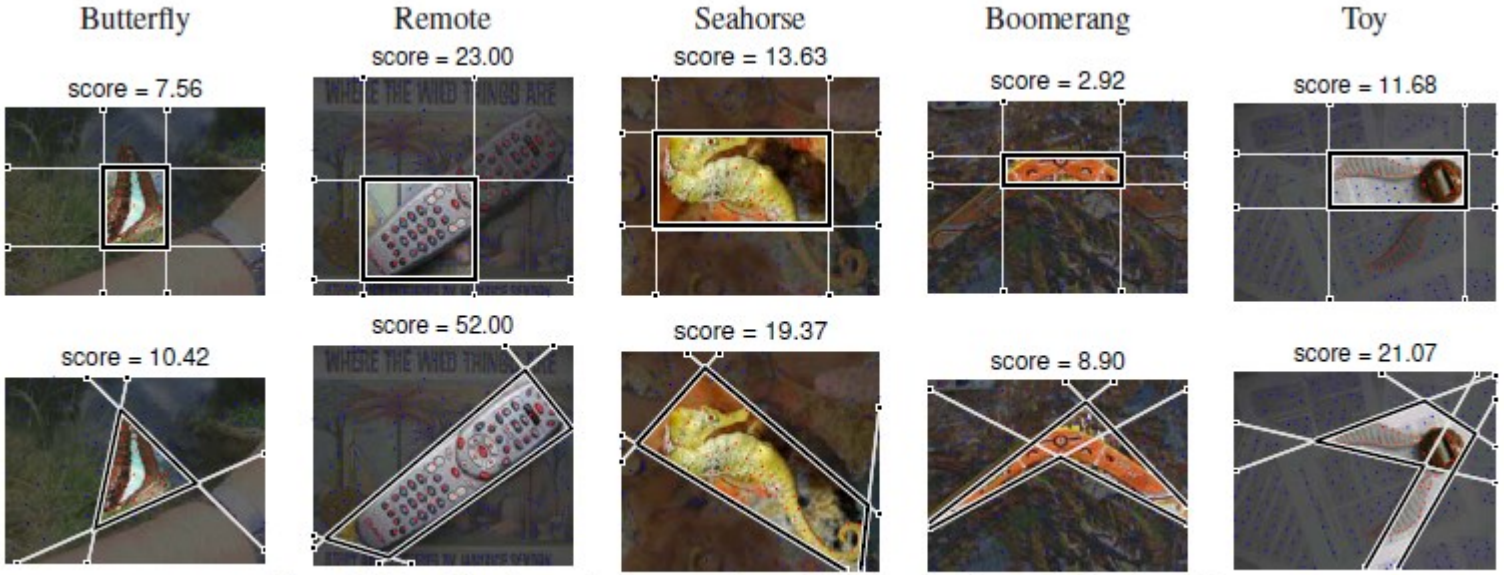


Figure 8. **Bounding boxes (top) versus bounding polygons (bottom)** (Section 3.2).

Sliding window 3段階手法 [Vedaldi et al. ICCV09]

Multiple Kernel Learningによる18種類の特徴の統合

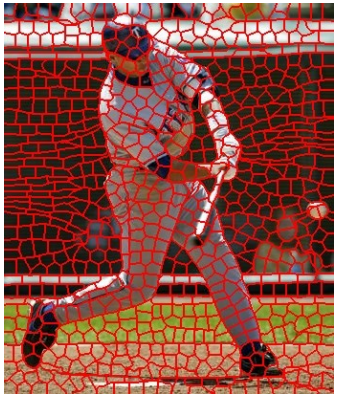
- 候補位置：discriminative visual wordsを含むbounding boxを多数生成.
- [1st step] MKL+線形カーネル で，物体体候補探索(2000程度)
- [2nd step] MKL+準線形カーネル $K(x,y)=\frac{1}{2}(1-\chi^2(x,y))$ の利用
  - 候補を100個程度に絞る
- [3rd step] MKL+カイ2乗RBF

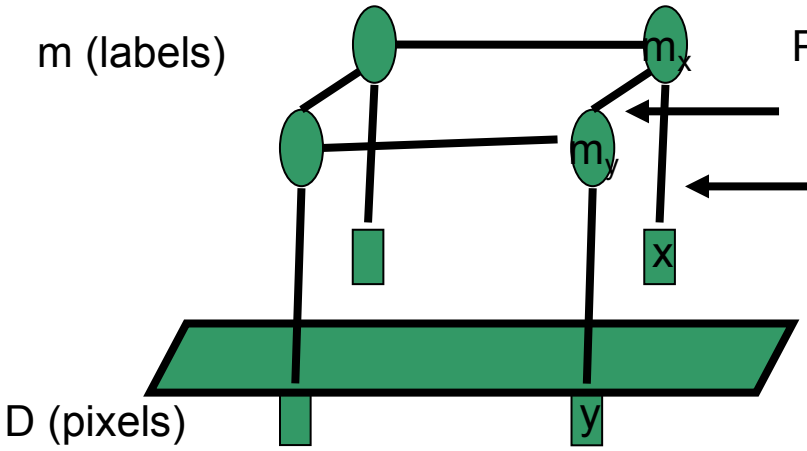
軽い処理 ⇔ 重い処理

低精度 ⇔ 高精度

カテゴリー領域分割

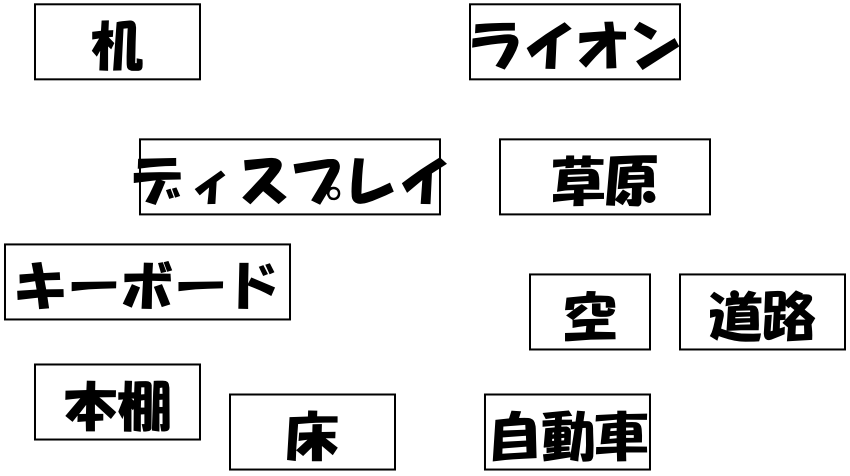
- 過分割領域（*superpixel*）＋ MRF（CRF）



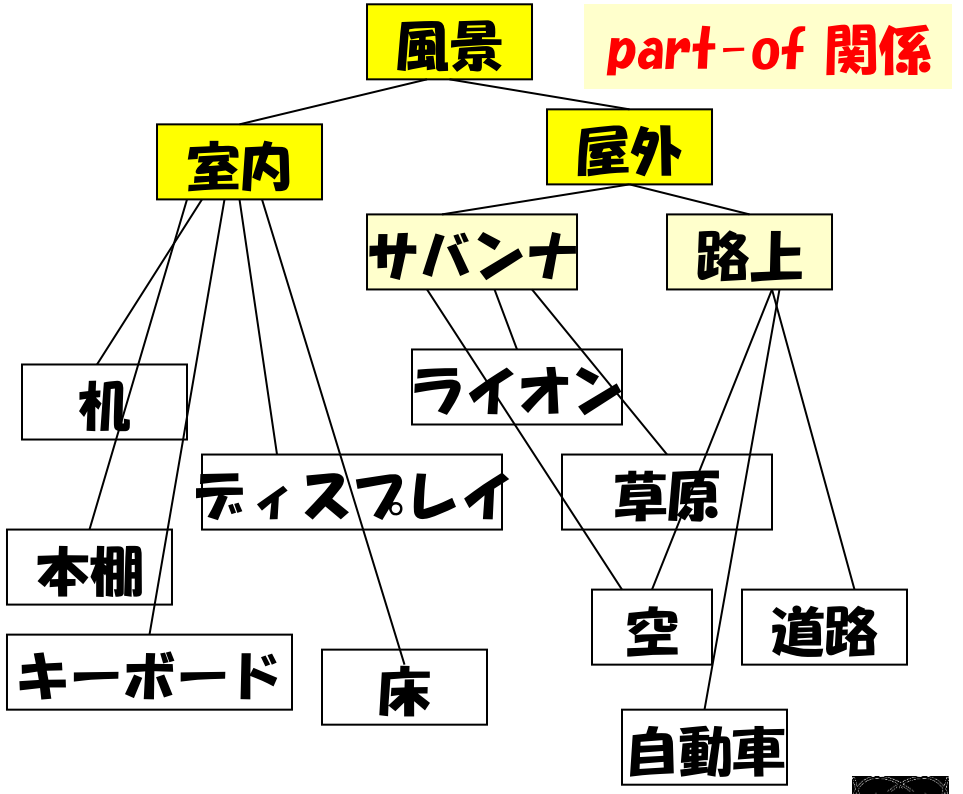


コンテキストの利用：人間は「常識」として持っている知識

- **共起関係：**
  共起の強さを確率で表現

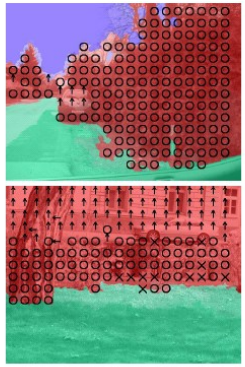


- **階層的認識：シーン認識**
  ＋物体認識（＋領域分割）

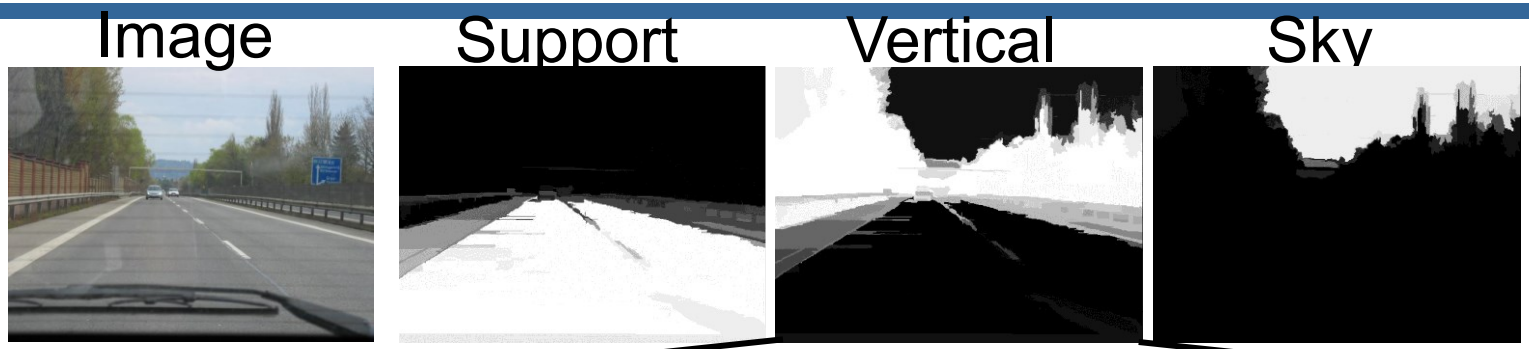


コンテキストの利用(2):
人間は「常識」として持っている知識

- 様々なコンテキスト情報
  - 共起関係
  - 相対位置関係　自動車は道路の上にある.ポストの上にはない.
  - 相対スケール　　机の上の車はミニカー
  - (重力に対する)支持関係　PCは机の上にある
  - 背景と前景の関係
- 外部コンテキスト情報
  - 写真の撮影日時，GPS情報
  - カメラのパラメータ(Exif情報)



例）Surface Estimation



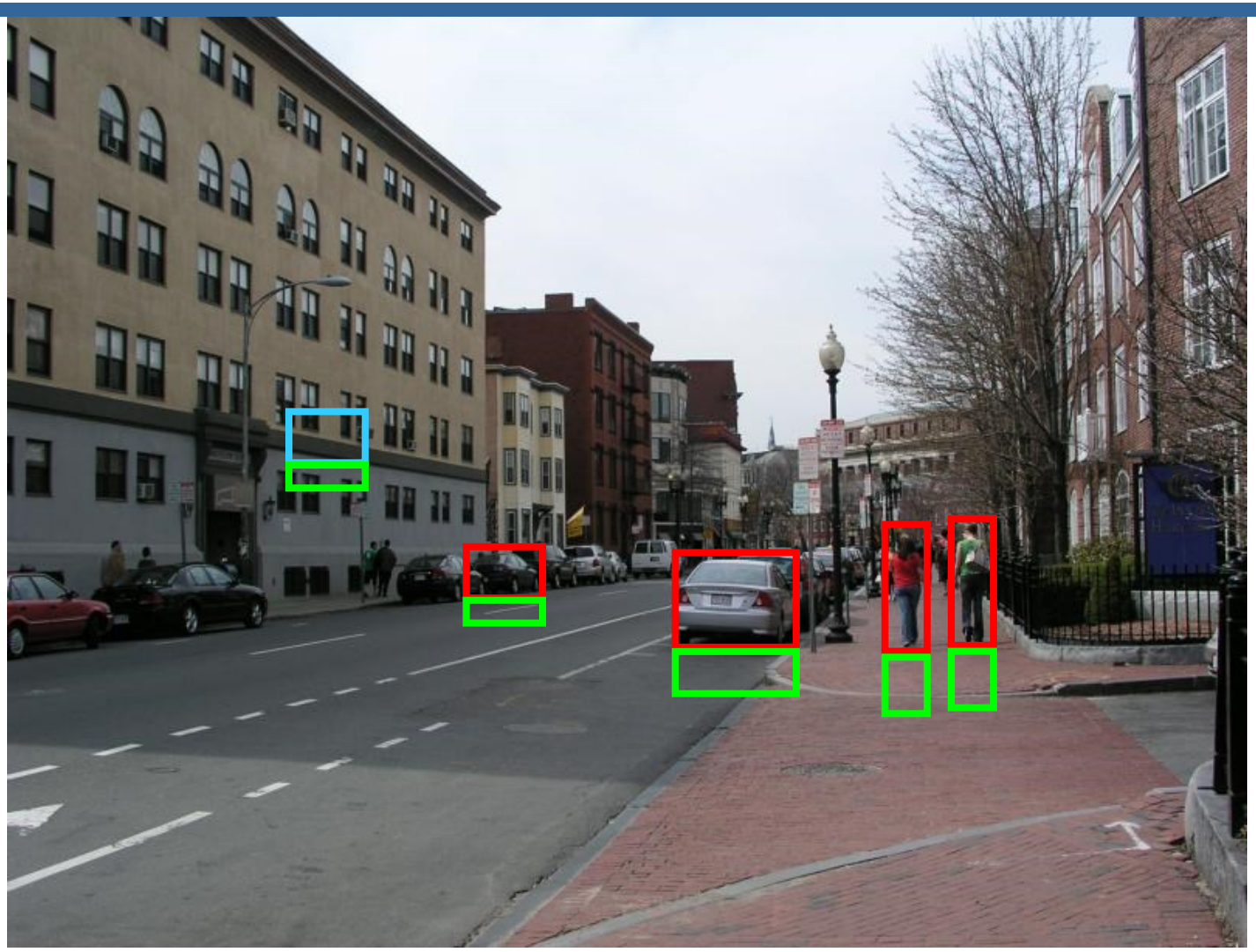
Object
Surface?

Support?

[Hoiem, Efros, Hebert ICCV 2005]

Slide by Derek Hoiem

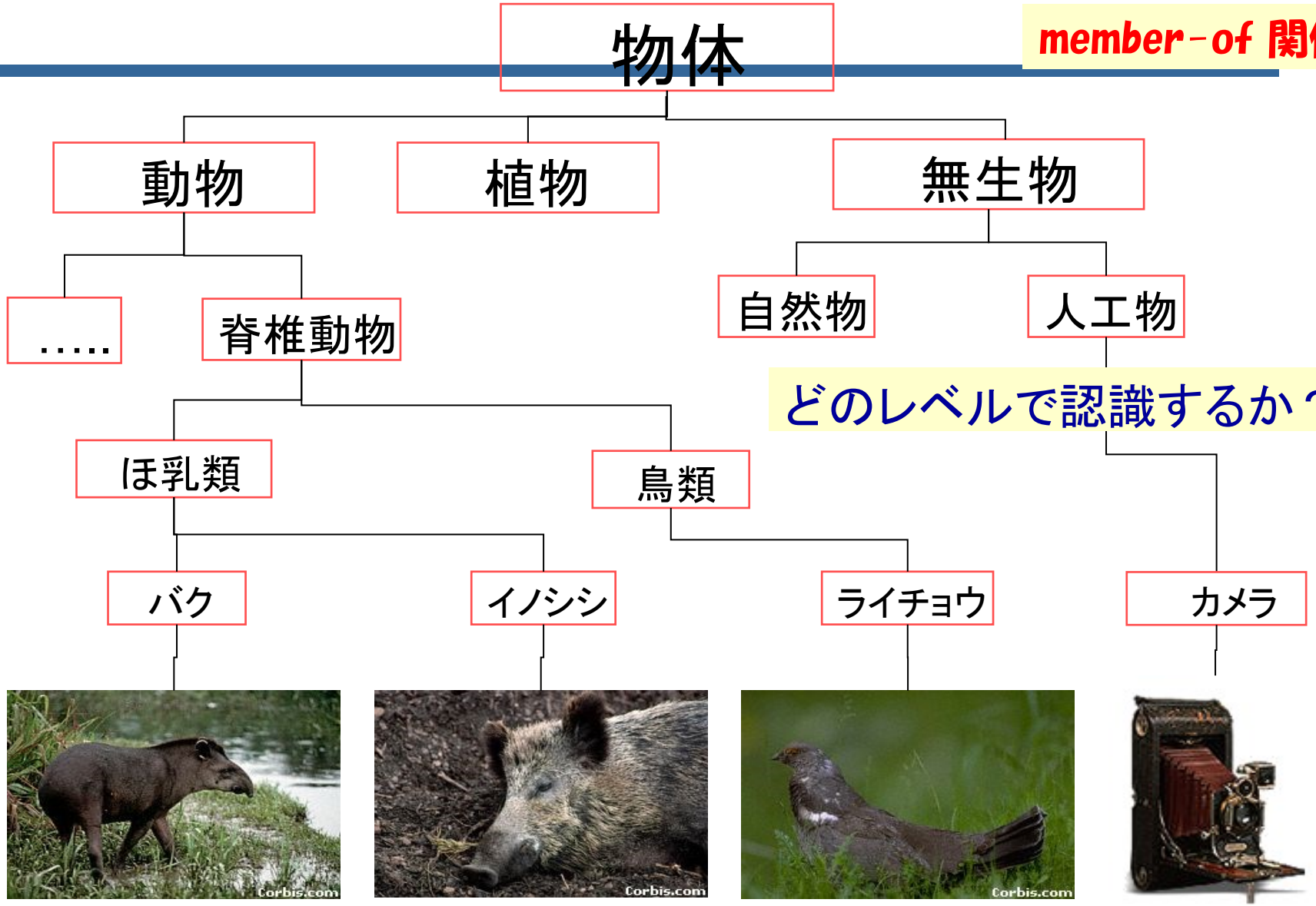
Object Support



Slide by Derek Hoiem

言語の階層的分類(タクソノミー)との関係

member-of 関係



5. 大量画像データ

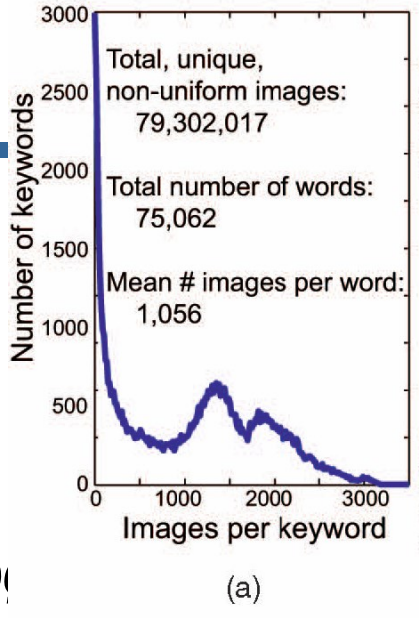
80 Million Tiny Images: A Large Data Set for Nonparametric Object and Scene Recognition

8000万枚のWeb画像による
類似画像検索による認識

8千万枚のWeb画像による認識

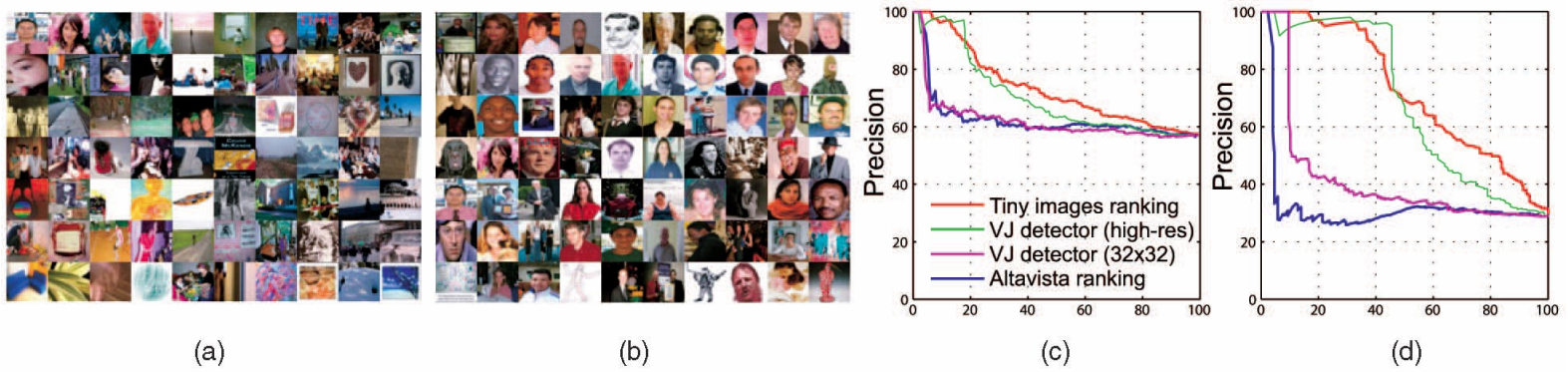
- データの量を増やすことで認識精度を上げる
  - 7900万枚の画像(巨大なデータセット)収集に1年
  - ノイズが含まれていても気にしない.
  - Nearest Neighbor法(簡単なアルゴリズム)
    - データからのアプローチ.
- 非常に多くのデータを扱う
  - 画像インデックス技術
  - イメージの低次元化(32×32×3), 760GB
- 最新手法の結果に匹敵する認識精度

3.1 画像収集



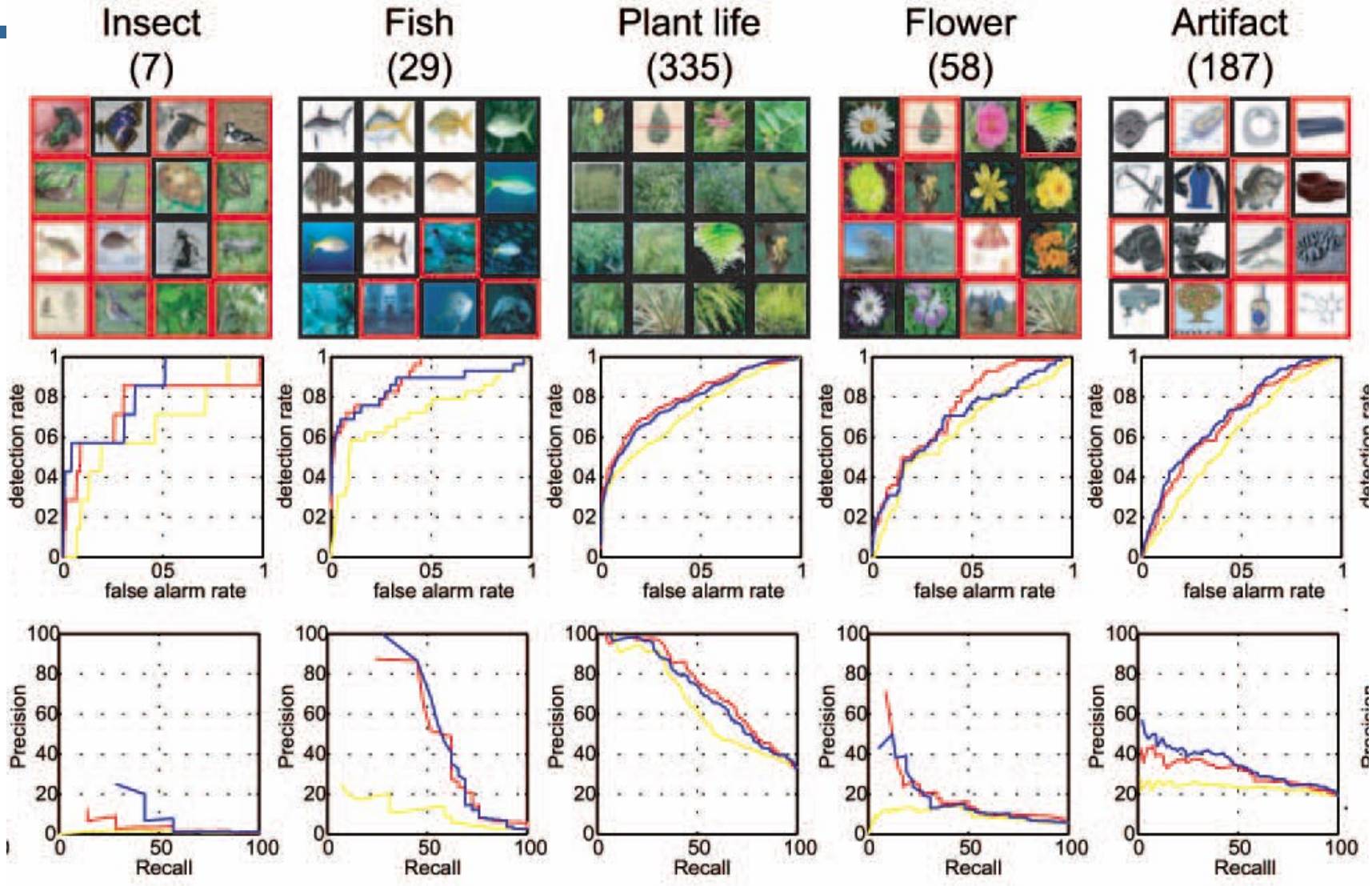
- キーワード
  - Wordnet[15]から75,062語
  - 抽象的でない名詞
- 画像検索エンジン
  - Altavista, Ask, Flickr, Cydral, Goo
    Picsearch, Webshots
- 収集枚数
  - 97,245,098枚
  - 重複画像を削除して，79,302,017枚
  - 1つのキーワードの画像上限は3000枚に設定
- 画像容量
  - 32×32のサイズで保存すると，760GB

「人」の認識



- 検索エンジン改良にも貢献できる
  - (a)はAltavistaでの検索結果のランク上位
  - (b)はWordnet Voting Schemeでのランク上位
- 高解像度画像を使った顔画像認識(OpenCV)と　同程度の精度を出すことに成功した

一般対象の認識


Insect
(7)
Fish
(29)
Plant life
(335)
Flower
(58)
Artifact
(187)
detection rate
false alarm rate
Precision
Recall

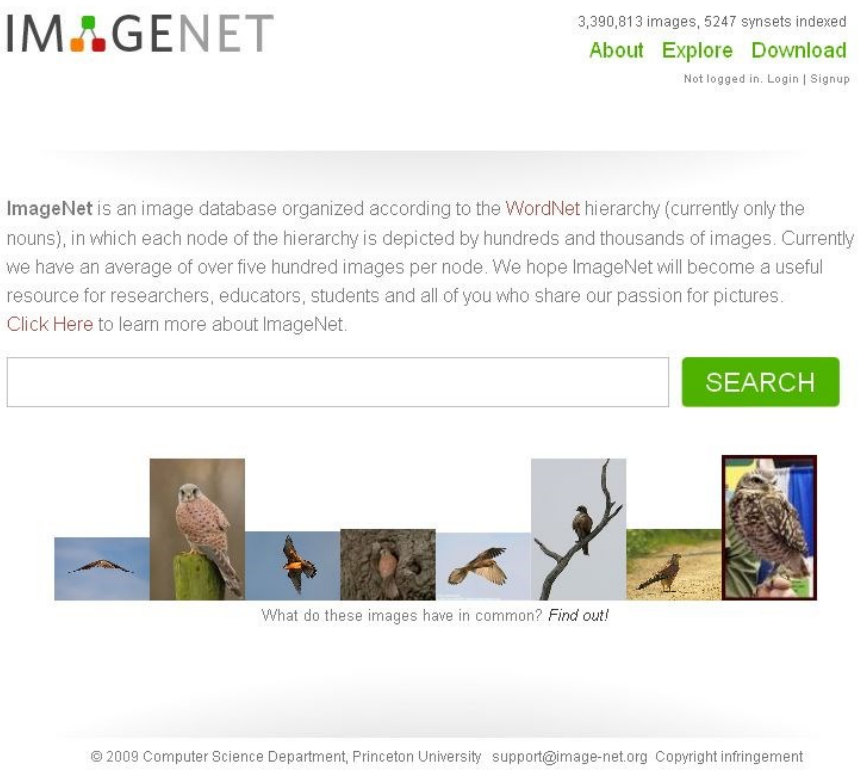
結論

- 画像が多いクラスは特定クラス検出器に匹敵する
- 一部のクラス以外は利用できる学習画像が少ない．それらのクラスには最近傍法は向かない
- 物体認識の二つの側面
  - モデルとデータ

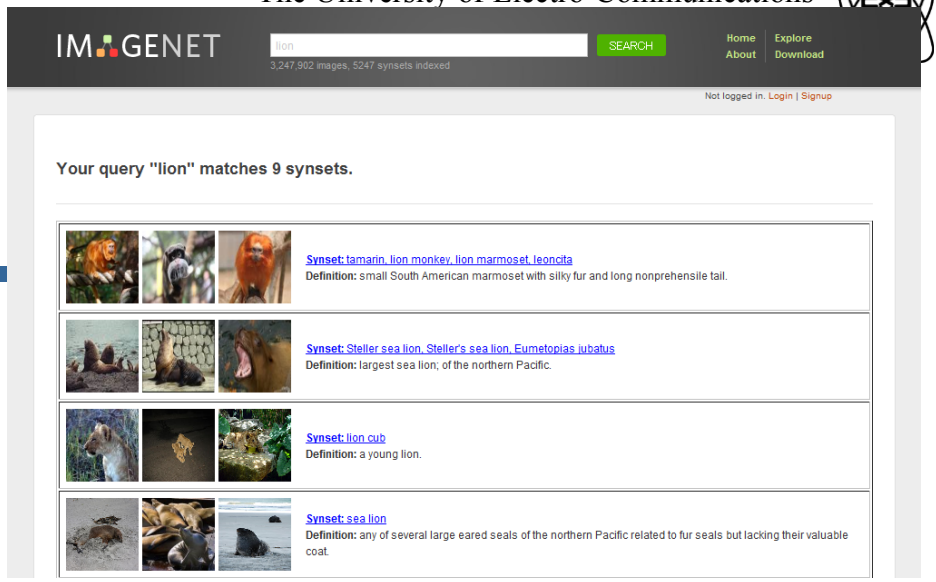
MSRAのXian-Sheng Huaが，
「4billion で実験したら，もっとパフォーマンスが出た．」
とICCV WS on LAVDで発言．　→ ARISTA [CVPR10]

Image Net

[2] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, Li Fei-Fei : ImageNet: A Large-Scale Hierarchical Image Database, CVPR, (2009).



ImageNet



- あらゆる言葉に関する画像をデータベース化. Wordnetの画像版.
- 現在5247語について320万枚収集

http://www.image-net.org/

- Amazon Mechanical Turk を利用して作成. (不特定多数の人の知識の利用)
  - 画像認識は補助的に利用. 最終的には人手
- 階層構造を持つ大規模画像DBの意義
  - 画像認識の学習データ. 評価データ.
  - 画像と意味の大規模分析

Amazon Mechanical Turk

- 大量の“人力”の利用 ： 集合知

ImageNetの概要

- ImageNetとは？
  - WordNet構造の背景に基づいて打ち立てられた画像の大規模な画像オントロジー
- ImageNetの内容
  - WordNet[9]の80,000synset(synonym set)に対応
  - 各synsetで平均500-1000枚のクリアでフル解像度の画像を用意(予定総数は5千万)
  - 画像認識研究用に全画像を公開.
  - Amazon Mturk を利用して人手で構築. ノイズなし

現在のImageNetの状況 (2009/11)

- 構築状況
  - 5247synsets (今年9月) ⇒14847 syn. (今日)
  - 320万枚の画像 (今年9月)⇒ 940万枚(今日)
  - 12のサブツリーの構築
    - 哺乳類, 鳥, 魚, 爬虫類, 両生類, 乗り物, 家具, 楽器, 地形, 道具, 花, 果物
- データベースの閲覧
  - http://www.image-net.org

図1　哺乳類と乗り物のサブツリー

ImageNetの今後

- ImageNetの完成
  - およそ50,000synsetsに広がる5000万の良質画像
  - 一般公開，オンラインですぐに利用可能
  - データの効率的な配布にクラウドストレージを利用
  - 位置検出，セグメンテーション，相互synset参照等の拡張
  - ImageNetコミュニティの育成および全ての人が貢献または利益を得られるオンラインプラットフォームの開発

写真共有サイトの登場

- Flickr, Picasa など
  - Consumer-generated media (CGM)と呼ばれる
  - 多くの写真にはユーザが付与してタグ(キーワード)が付いている
    - 写真共有サイト内のタグ検索は，Web画像検索より高精度．ただし，多義語の問題はある．
  - 最近では，位置情報もついている場合がある．
    - 位置情報付き画像，geotagged image と呼ばれる
  - 画像(visual feature)と単語(word concept)の関係分析に利用しやすい

これらの画像＋メタデータを利用することで，新しい研究が可能！

"Sushi" in Caltech 256

- (Probably) Collected in English keywords

"Sushi" in our own dataset

- Collected in Japanese keywords

Which do you like to eat ?

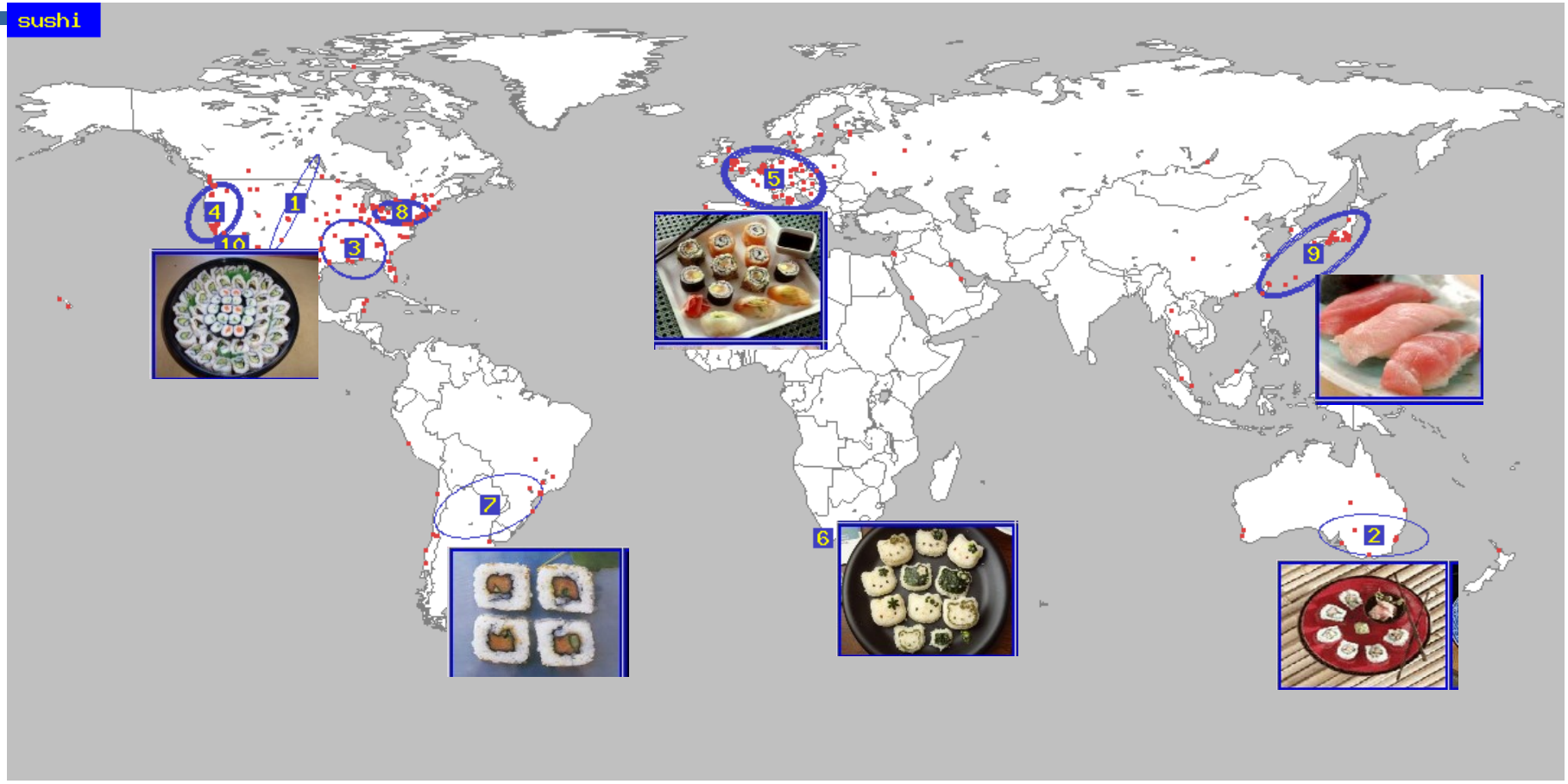
These two "sushi" image sets are surely different, although both are image sets associated with the "sushi" concept !

なぜCaltechの寿司は，まずいのか？

Caltech "sushi"

Japanese "sushi"

"Sushi" over the world

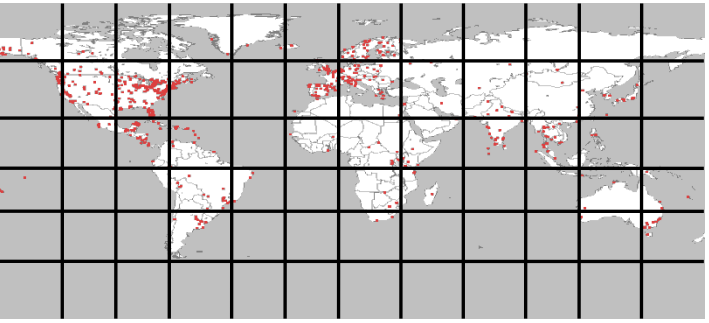


- 同一のカテゴリーに属する画像であっても，地域や文化によって変化する．カテゴリーはグローバルではない！

The University of Electro-Communications
Tokyo, JAPAN (UEC)

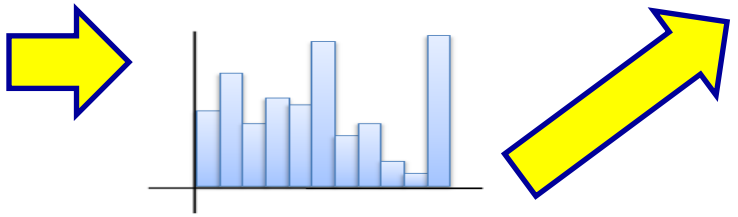
Image entropy と location entropy [Kawakubo et al. 09]

- Flickrから230の名詞について，それぞれ500枚の位置情報付き画像を収集．
- 画像領域エントロピーと位置エントロピーを求めて，コンセプト分析をする．



Geo-location entropy

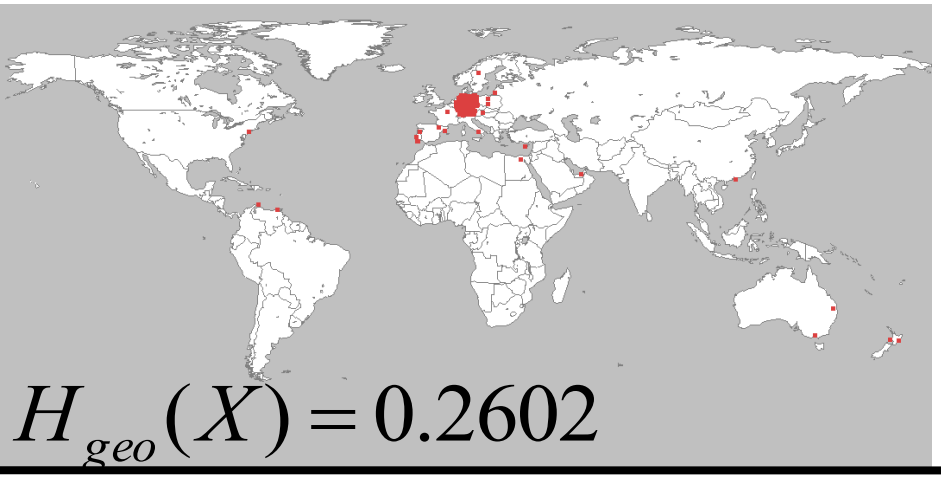
$$H_{geo}(X) = -\sum_i b_i \log_2 b_i$$



The University of Electro-Communications
Tokyo, JAPAN (UEC)

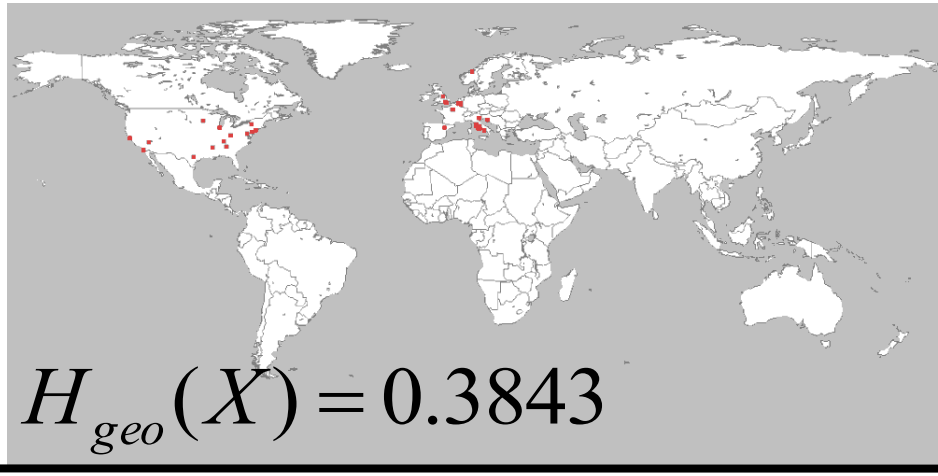
Geo-entropy $H_{geo}(X)$

Deutschland



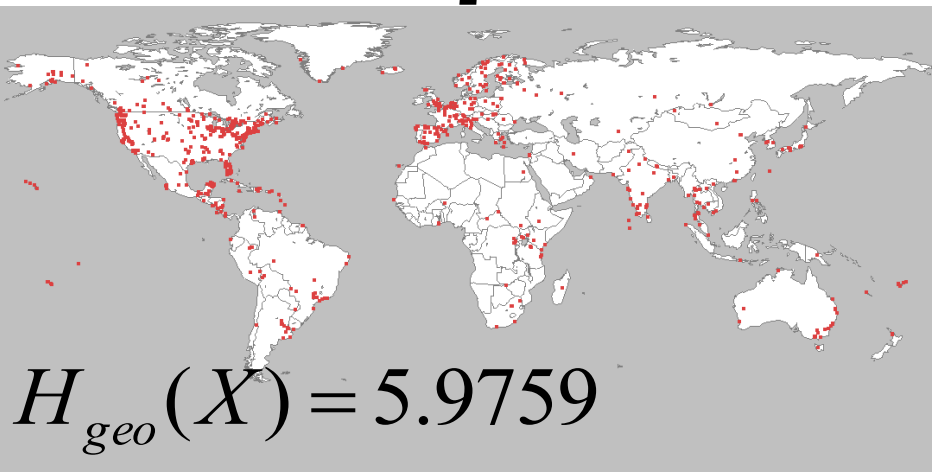
$H_{geo}(X) = 0.2602$

Rome



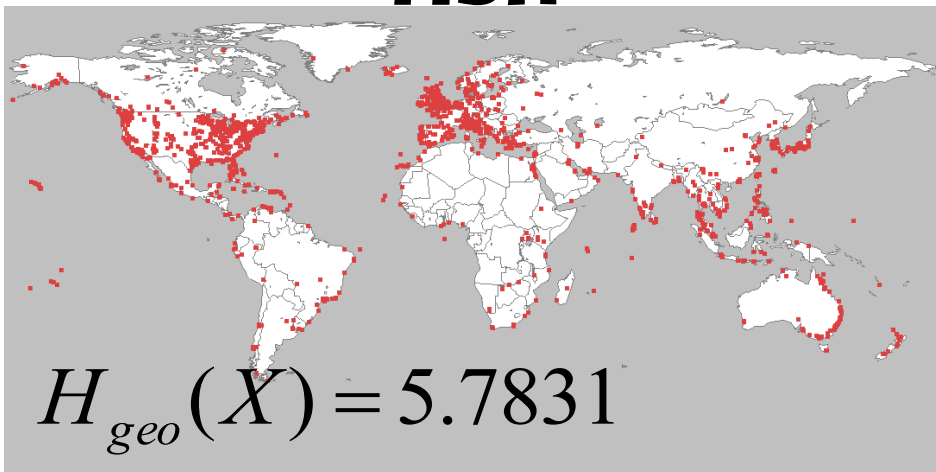
$H_{geo}(X) = 0.3843$

mosquito



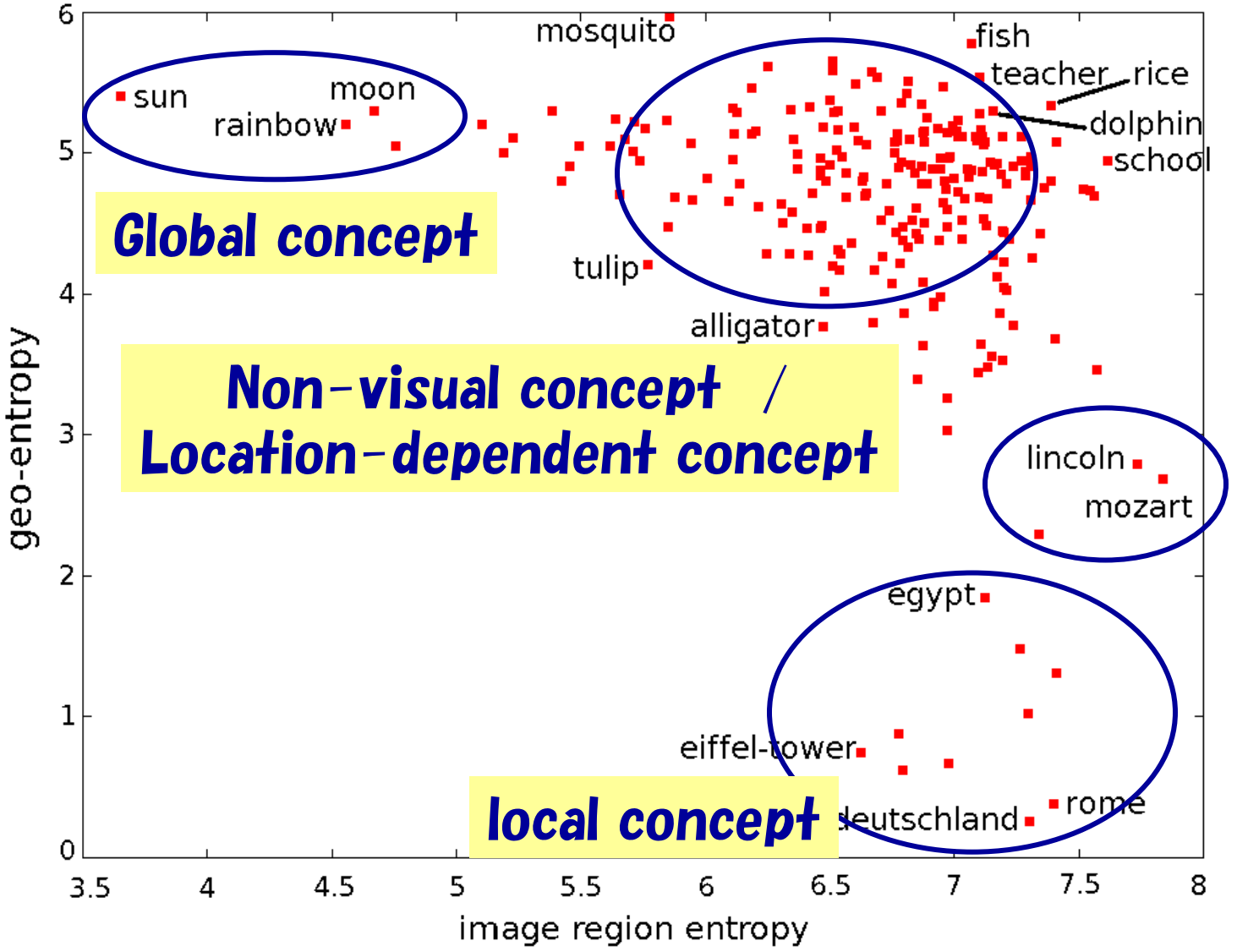
$H_{geo}(X) = 5.9759$

fish



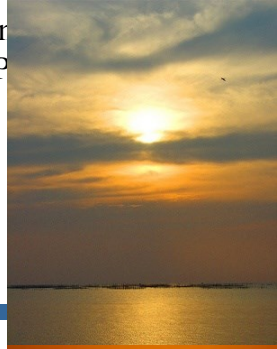
$H_{geo}(X) = 5.7831$

Image entropy vs. geo-entropy



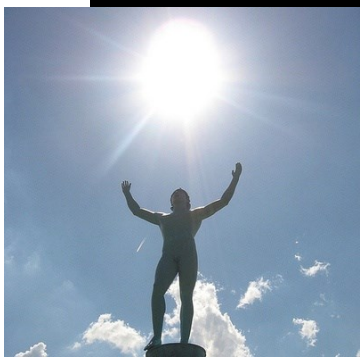
The University of Electro-Comm
Tokyo, JAP

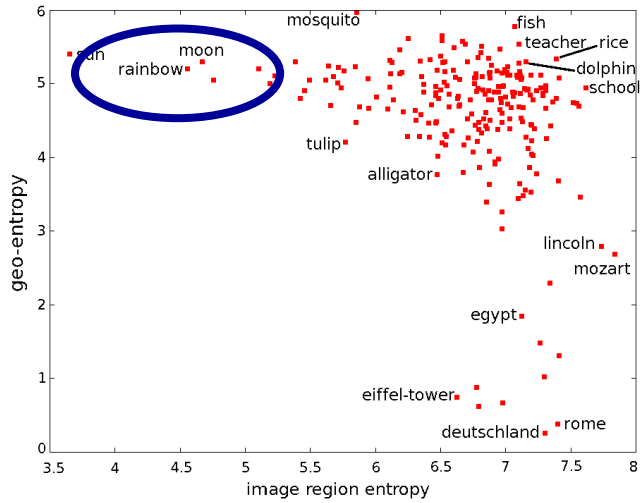
Sun, rainbow, moon

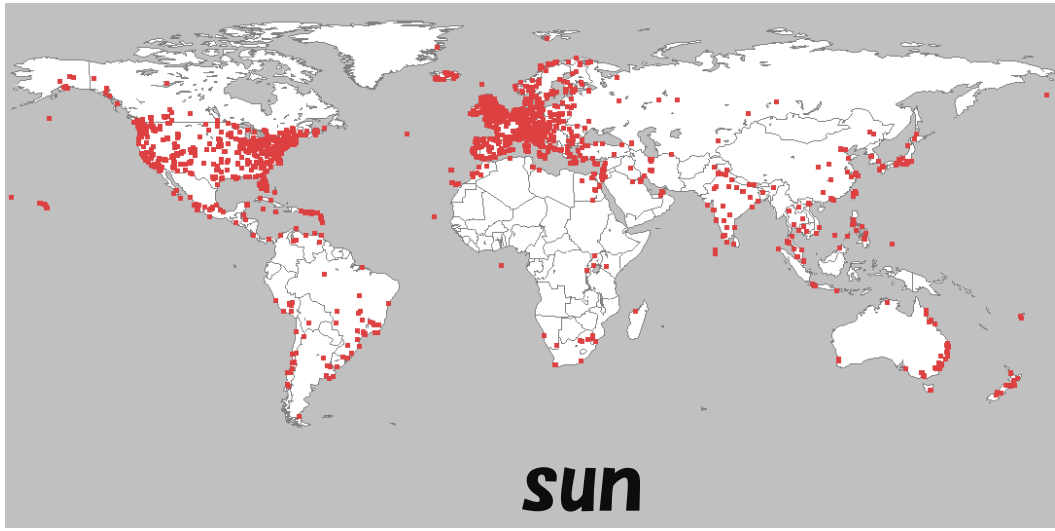


- Comcepts related to sky
  - Image region entropy : low
  - Geo-location entropy : high

They exists everywhere in the world, and the apperances are similar.





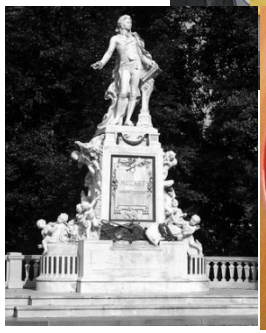


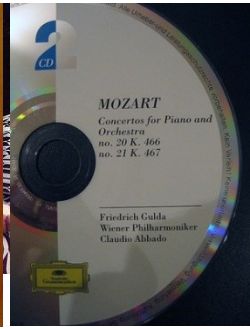
The University of Electro-Communications
Tokyo, JAPAN (UEC)

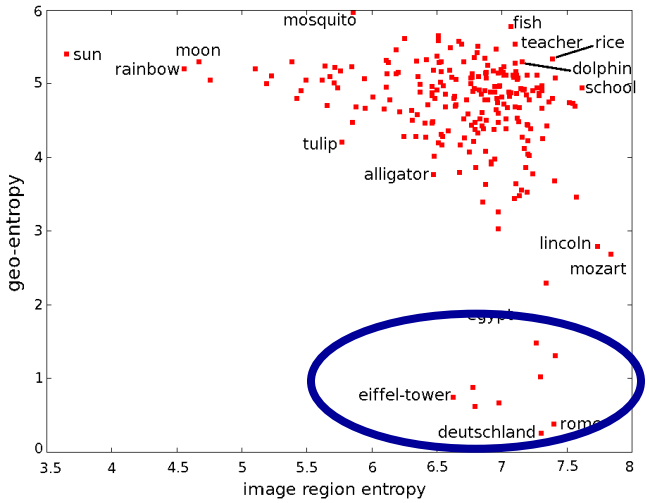
Rome, Deutschland, Mozart

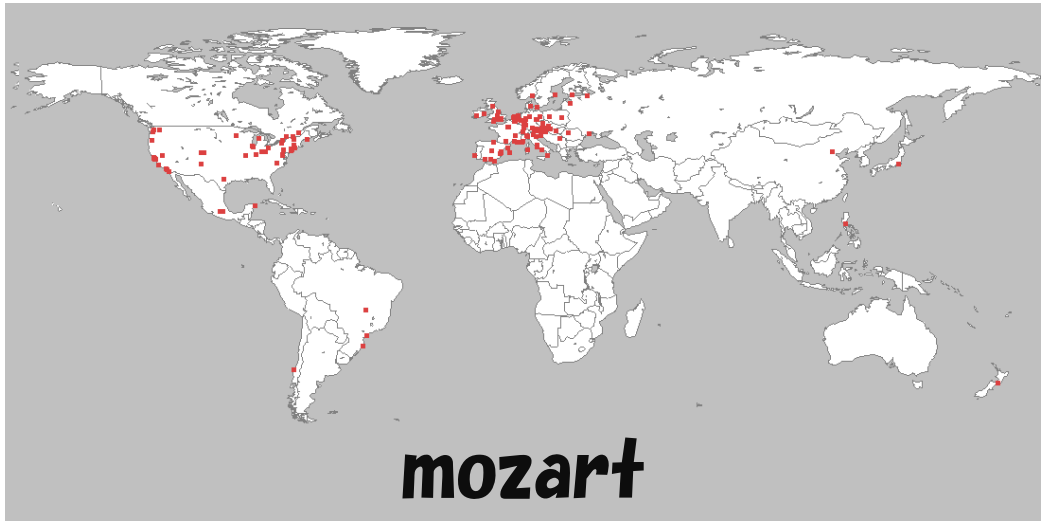
- Image region entropy: high
- Geo-location entropy : low

The geotags concentrates on specific areas. Their appearances are various.







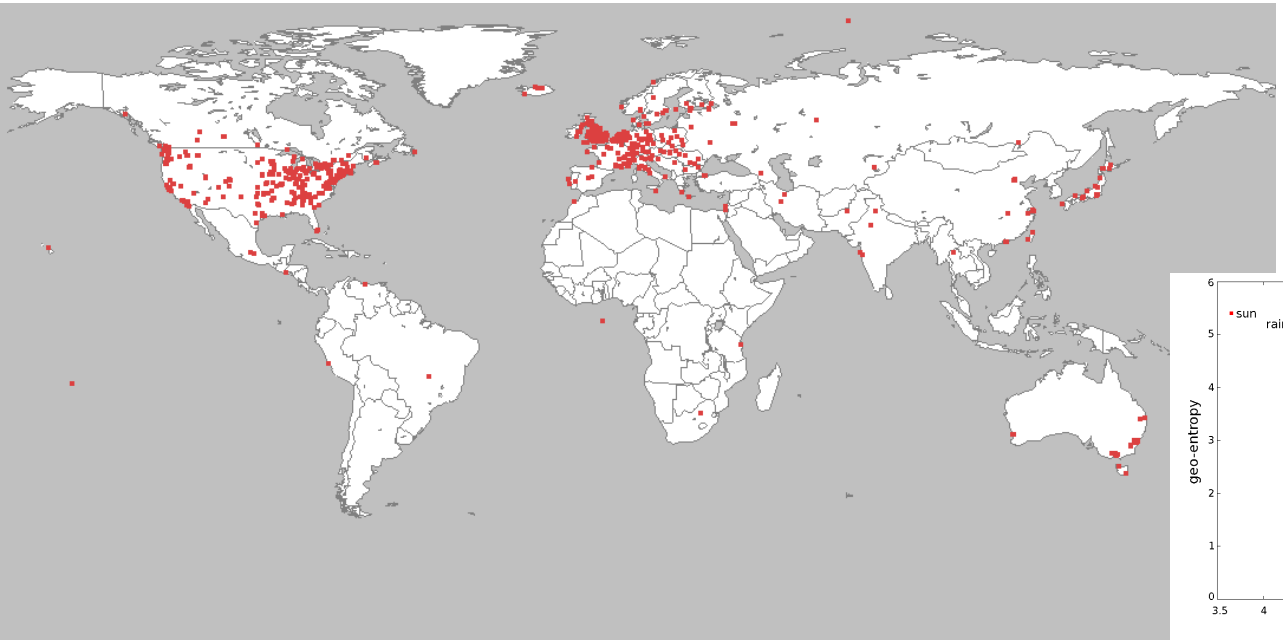


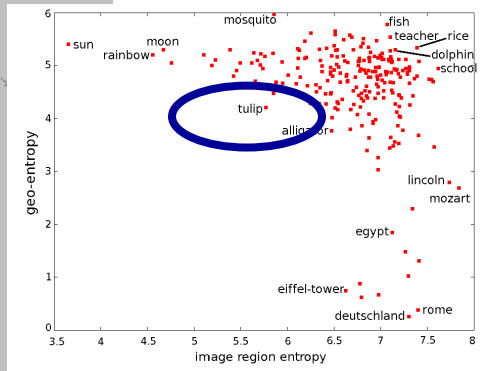
The University of Electro-Communications
Tokyo, JAPAN (UEC)

tulip

Image region entropy: low
Geo-location entropy : med.

- Variance of color did not reflect on image region entropy, since we use SIFT-based BoF representation.
- Holland and England are main areas.



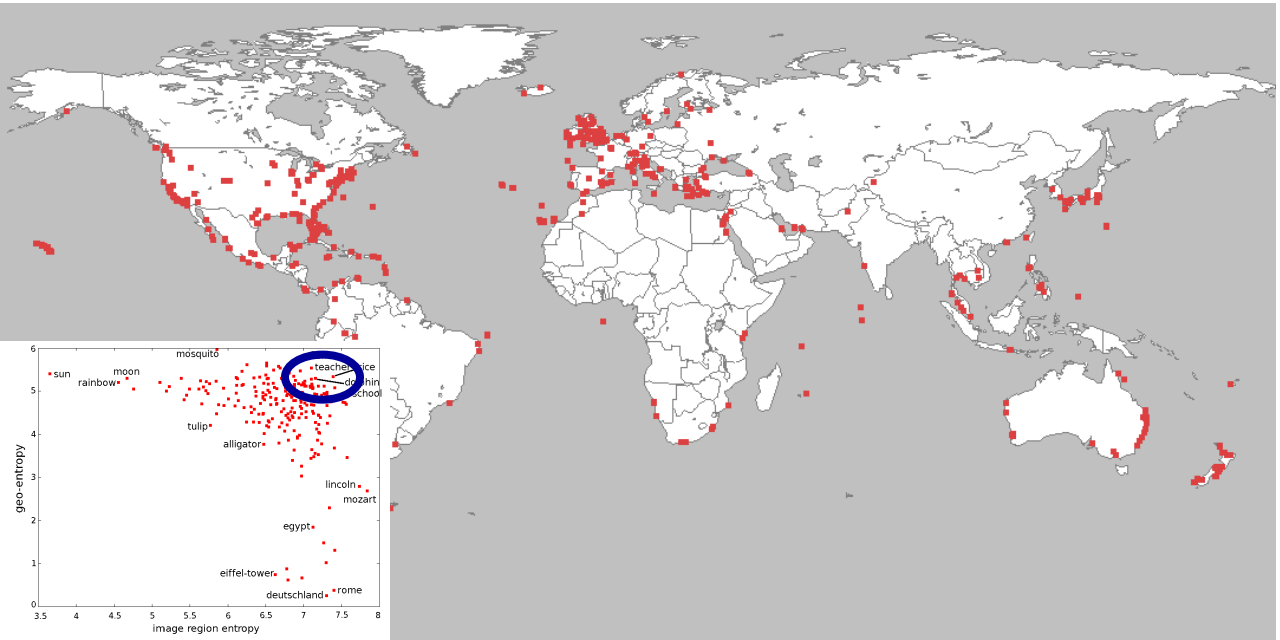


The University of Electro-Communications
Tokyo, JAPAN (UEC)

dolphin

Image region entropy: high
Geo-location entropy : high

- Most of dolphins are taken in sea or aquarium
- In seaside areas over the world



rice

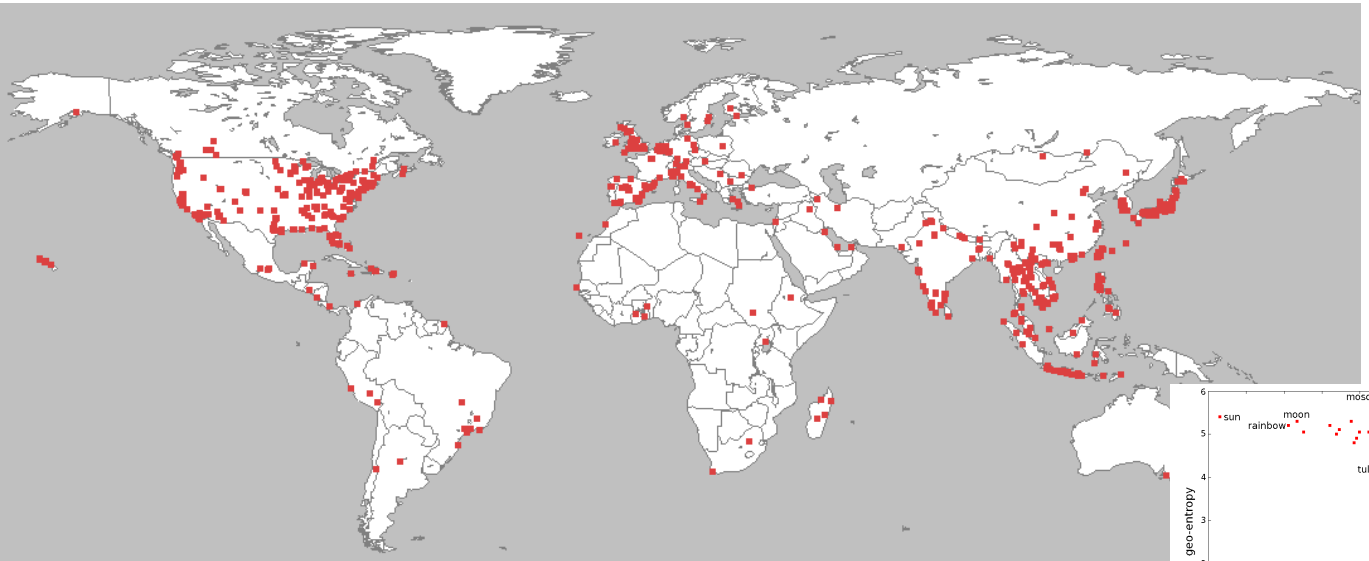
Image region entropy : high
Geo-location entropy : high

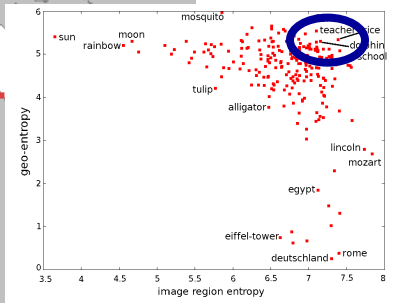










$H_{geo}(X)$ = 5.3425

６．今後の方向

まとめ

- 「一般物体認識」について紹介した.
  - 特定物体　と　一般物体　の違い
  - 現在どこまで出来るか？
- 基本的な手法：Bag-of-Features (BoF)
- 画像単位での分類
- 一般物体の位置検出
- Web上のデータによるデータセット
- 今後の展開
  - 特定物体認識が一般物体認識になる？

今後の方向（1）

- 80million から示唆されるように，
  量がブレークスルーになる可能性がある．
  - 特定物体認識（NN探索） が主役！
    - 一般物体認識の特定物体認識化．
    - メトリックの工夫が重要になる可能性．Metric learning．
  - でも，MS/Google以外には手に負えない可能性．．
    - 80millionの収集に1年掛かると，
      4billion⇒50年．．．

2003年

Google Image Search BETA

Advanced Search Preferences Search Tips

lion animal | Google Search | I'm Feeling Lucky

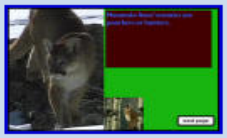
Search images Search the Web Mature content filter is On

Searched images for **lion animal**. (BETA) Results **1** - **20** of about **853**. Search took **0.05** seconds.

animal-lion.jpg
746 x 519 pixels - 88k
members.tripodasia.com.tw/Kasha/

lion.gif
164 x 158 pixels - 26k
www.akronzoo.com/animal.asp

lion watercolor.GIF
640 x 480 pixels - 254k
biology.ecsu.ctstateu.edu/HighSchool/animal.htm



mtnlion1.gif
519 x 348 pixels - 45k
www.beavton.k12.or.us/.../leahy/00-01/animals/mtnlion.htm

mtnlion2.gif
520 x 351 pixels - 64k
www.beavton.k12.or.us/.../leahy/00-01/animals/mtnlion2.htm
[More results from www.beavton.k12.or.us]

lion.jpg
400 x 400 pixels - 52k
www.billybear4kids.com/.../sliders/online/lion.html

lion-girl.jpg
400 x 400 pixels - 48k
www.billybear4kids.com/.../sliders/online/lion-girl.html

lion.jpg
277 x 292 pixels - 27k
maskmaker.com/animal.html

lion.jpg
452 x 335 pixels - 19k
www.ragsdalefinials.com/animalheads.htm

lion.jpg
288 x 229 pixels - 29k
www.uiuc.edu/unit/ATAM/conservation/iran.html

animal.JPG
512 x 384 pixels - 308k
www.trnty.edu/bookstore/

stuffed animals cougar puma mountain **lion** panther.JPG
350 x 434 pixels - 24k
www.kathyskreations.com/.../stuffed%20animals%20cougar.html

lion.jpg

AliGloves.jpg

image16.gif

lion.GIF

lion animal

2005年

Google イメージ

ウェブ イメージ ニュース グループ New! ディレクトリ デスクトップ more »

lion filetype

イメージ検索 オプション
表示設定

lion filetype:jpg animal の検索結果 約 **6,110** 件中 **1 - 20** 件目 (**0.14**

表示: **すべてのサイズ** - 大 - 中 -

n&h **lion** 1024
1024 x 768 ピクセル - 147k
www3.coara.or.jp/.../ n&h-lion-1024.jpg

lion 009L
500 x 341 ピクセル - 31k
www.stockley.co.za/ gallery/lion-009L.jpg

Animal Bites
200 x 200 ピクセル - 14k
www.traveldoctor.co.uk/ animals.htm

personnel tom_b 2004 **lion**
599 x 800 ピクセル - 145k
www.vetmed.wisc.edu/.../ tom_b/2004-lion.jpg

Lion
266 x 400 ピクセル - 18k
www.sdnhm.org/ exhibits/eyes/lion.html

Animal Posters
540 x 445 ピクセル - 56k
www.realtime.net/~raintree/ gallery/p-lion.jpg

LION
230 x 230 ピクセル - 18k
www.yamaha-motor.co.jp/.../ animal/world/lion/

Stuffed **Animal Lion**
300 x 299 ピクセル - 37k
www.certificatespecialists.com/ images/Consume...

2009年

Google 画像検索 lion animal 画像検索 画像検索オプション

セーフサーチ: 中 ▼

像 検索ツールを表示 検索結果 約 7,950,000 件中 1 - 20 件目（0.04 秒）


Lion -
800×600 - 89k - jpg
raidmyspace.com
類似の画像を探す


If you were an
470×324 - 31k - jpg
sodahead.com
類似の画像を探す


Lion Bait
792×768 - 99k - jpg
whohastimeforthis...
類似の画像を探す


Animal picture -
601×480 - 204k
animalpicturegallery.net
類似の画像を探す


Pictures of African
448×299 - 30k - jpg
pictures-of-african-anim...
類似の画像を探す


lion animal
312×432 - 14k - jpg
shop-safely.com


lion animal photo
800×600 - 71k - jpg
graphicshunt.com


African **Animal**
400×300 - 26k - jpg
animals.howstuffworks.com


Lion
625×450 - 105k - jpg
animal.discovery.com
類似の画像を探す


Area ライオン、トラ、ヒョウ
484×530 - 47k - jpg
area2.sakura.ne.jp
類似の画像を探す


Wild **Animal** Park
450×338 - 56k - jpg
randomsandiego.com
類似の画像を探す


australian **animals**
417×297 - 31k - jpg
wildanimalsplanet.com


Starring horse-riding
468×366 - 26k - jpg
letstalkug.net
類似の画像を探す


Wild **animals**, as
479×450 - 62k - jpg
adamthinks.com
類似の画像を探す


Clipart - **lion**,
300×208 - 18k - jpg
fotosearch.com


Cub **Lion** Picture
360×299 - 41k - jpg


Lion Face
600×465 - 75k - jpg


Nature **Animal Lion**
350×460 - 210k - jpg


A LION ANIMAL
350×504 - 28k - jpg


Lion on Horseback
468×393 - 39k - jpg

Google 画像検索 [ ] 画像検索 画像検索オプション
セーフサーチ: 中 ▼

2009年　類似画像検索

検索結果 約 556 件中 1 - 20 件目（0.04 秒）


350×504 - 28k - jpg
news.com.au
類似の画像を探す


272×480 - 130k
anythingbutipod.com
類似の画像を探す


400×600 - 98k - jpg
jpbutler.com
類似の画像を探す


355×508 - 77k - jpg
klub.bgdcafe.com
類似の画像を探す


400×600 - 119k - jpg
nadiroba.ucoz.ru
類似の画像を探す


2592×3888 - 675k
commons.wikimedia.org
類似の画像を探す


781×557 - 41k
animalpicturesarchiv...
類似の画像を探す


400×600 - 87k - jpg
buzztexas.com
類似の画像を探す


220×330 - 44k - png
camera-africa.com
類似の画像を探す


433×600 - 53k
sciencephotogallery.com
類似の画像を探す


290×263 - 23k - jpg
travel.msn.co.nz
類似の画像を探す


299×349 - 55k - jpg
missaovidanova.com.br
類似の画像を探す


577×884 - 156k
donna9507095...
類似の画像を探す


168×293 - 9k
d230.org
類似の画像を探す


781×1000 - 143k
brianhamptonphotogra...
類似の画像を探す

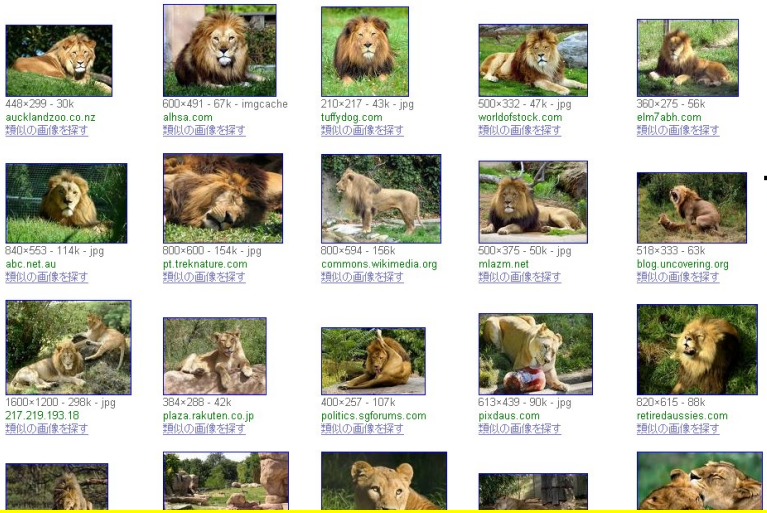
201?年

一般物体認識問題は，超大量データ＋特定物体認識 で解決？

Google™ 画像認識

80 billion image DB

NN search（特定物体サーチ）



surrounding text, tag解析

「ライオン」

「動物園」

「多摩動物園」

量が解決する可能性．さあ，我々はどうする？

今後の方向 (2): 別の方向を向いてみる.

- 認識するのは「物体カテゴリ」だけじゃない！
  - ICCV 2009: "Image Sequence Geolocation with Human Travel Priors"
    - 画像(写真)の「撮影位置」を推定.
    - ストリートビューから位置推定もできそう.
    - では, ラーメン画像から店を推定するのは？
  - 「Attribute」の認識　(○○ness, ○○ability…)
    - 顔認識から, 年齢, 性別推定へ発展. 性格推定？病気推定？人は見かけによる？よらない？
    - 食べ物のカロリー推定, 旨さ推定. 写真の上手さ判定.
    - 物体の古さ推定. お宝鑑定. 競馬の勝ち馬予測(?). . .
    - New Challenge : "One image tells many things !"
      - どこまで「画像認識」が可能か？ 画像から分かること. 分からないこと.


おわり