

一般物体認識における 機械学習の利用

< 画像認識と機械学習 >

電子情報通信学会
第1回 情報論的学習理論と
機械学習研究会 (IBISML)

2010年 6月15日

電気通信大学 大学院情報理工学研究科
総合情報学科

柳井 啓司

機械学習(ML) と パターン認識(PR)

■ 「パターン認識と機械学習」 ビショップ7本

パターン認識(PR)は工学を起源とするが、
機械学習(ML)は計算機科学の分野から生じている。
同じ分野を2つの側面から見たものとみなせ、...

PR \doteq ML

ML: アルゴリズム志向

PR: 応用志向



パターン認識(PR)の代表的な応用例 ⇒ 「画像認識」



「画像認識」研究とは？

■ いろいろなタスクがあります。

■ 基礎：画像処理, 特徴抽出, 領域分割

■ **パターン認識** 基礎理論

■ 光学的解析, 画質改善

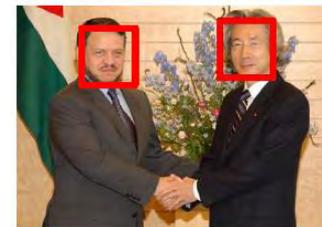
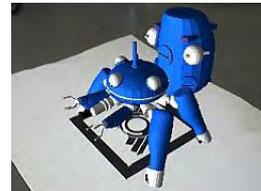
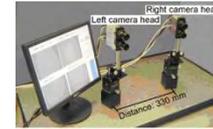
■ 3次元復元, ステレオ復元.

■ トラッキング.

■ CGとの融合. Augmented Reality.

■ **物体認識, 画像検索** ⇒ 狭義の「画像認識」

■ 実用化システム(ITS, 医用画像, セキュリティなど)



(狭義の)「画像認識」の基本的な処理

学習



ラベル付き学習データ

特徴抽出

機械学習モデルの学習

3つの基本要素

- 学習データ
- 特徴抽出
- 機械学習手法

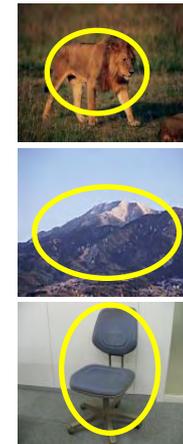
分類



未知の画像

特徴抽出

学習済モデルによる分類



「ライオン」

「山」

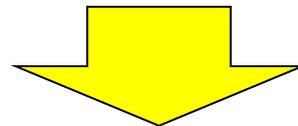
「椅子」

認識結果 (名称と存在位置)



画像認識 = 特徴抽出 + 学習データ + 機械学習

- 特徴抽出の決定版の手法の登場 (2004年)
 - Bag-of-Keypoints (Bag-of-Features(BoF) もしくは Bag-of-Visual-Words(BoVW) ともいう.)
 - 一般物体認識に有効であることが示されている。
- Webからの学習データの容易化



画像認識 = 特徴抽出 + 学習データ + 機械学習

「画像認識」研究における、
「機械学習」の相対的重要性が増大！！



アウトライン

1. [導入] 一般物体認識とは？

- 特定物体 と 一般物体
- 一般物体認識の例.

2. [特徴抽出]

基本特徴表現: Bag-of-Features (BoF)

3. [機械学習手法]

- 画像単位 / 複数ラベル / 領域単位 / スライディングウィンドウ / 領域抽出

4. [学習データ] 大規模データセット (Flickr, Mturk, ImageNet), ベンチマークデータ

5. [時間があれば] 研究紹介

1. 一般物体認識とは？

【参考文献】

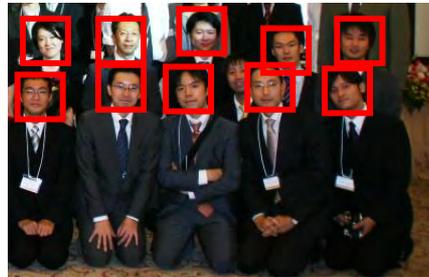
柳井啓司. 一般物体認識の現状と今後. 情報処理学会論文誌: コンピュータビジョン・イメージメディア, Vol.48, No. SIG16 (CVIM19), pp. 1-24, 2007.

「物体認識」とは？

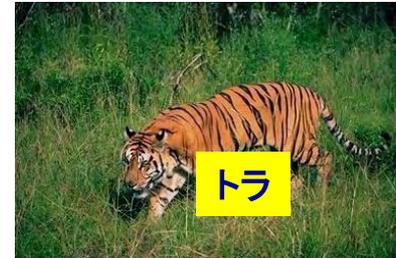
■ 画像中の「物体」を認識する技術, 研究分野



文字認識



顔画像検出



カテゴリー認識 シーン認識



3Dモデル物体認識



顔画像認識



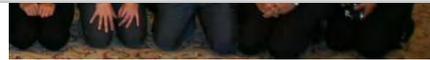
特定物体検出

「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



カテゴリ認識



シーン認識

カテゴリ認識



同一物体(インスタンス)の認識



顔画像認識



特定物体検出



3Dモデル物体認識



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

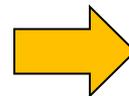


■ 同一物体認識

- 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

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.



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



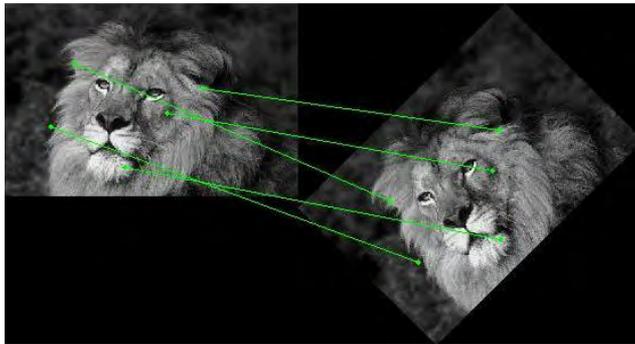
見た目 (appearance) が
まったく同じ物体を検出



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.

特定物体認識の技術的背景

■ 局所不変特徴量による特徴点マッチング



■ 大量の特徴点に対する検索

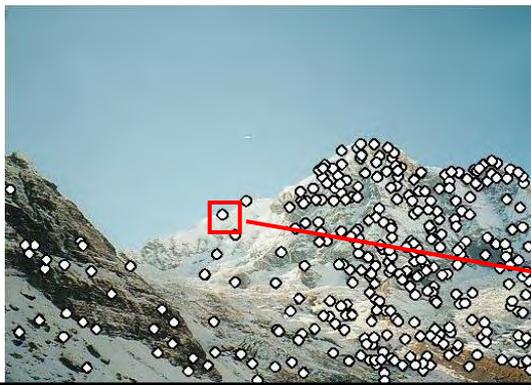
- 1枚の画像から数百点 ⇒ 50万枚画像で1億点
- Approximate Nearest Neighbor (ANN)
- Locality Sensitive Hashing (LSH)
- Visual Words + Inverted Index

機械学習という
よいも“検索”



局所特徴：対応点探索のための特徴

- 画像の特徴的な部分を検出し(**特徴点検出**), その部分の周辺パターンをベクトル化(**特徴点記述**)
- **特徴点**: 画像の変化(回転や拡大など)があっても常に検出される特徴的な場所(山の山頂, 色の境目など)

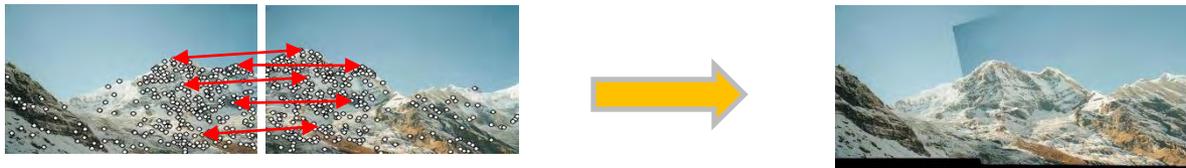


特徴点ベクトル: 回転, 拡大縮小があっても, 同じ見た目のパターンはほぼ同じベクトル値になる

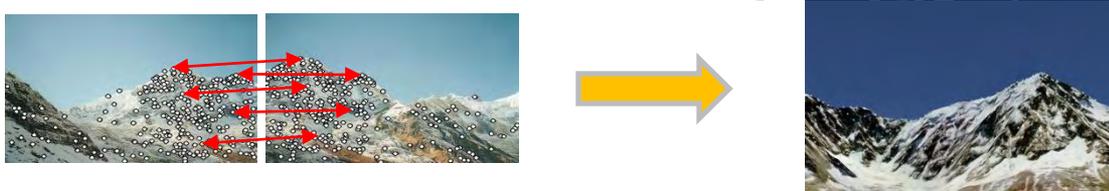


局所特徴量による対応点探索

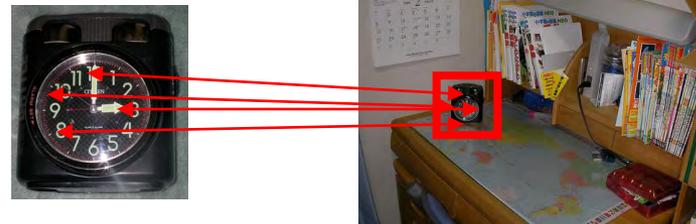
- 2枚の画像の点の対応が分かると..
 - パノラマ画像生成。(複数画像からの大画像生成)



- 3次元計測。(2枚の対応と相対撮影位置が分かれば, 3次元ステレオ計測が可能.)



- 特定物体認識。(同一物体の対応付け.)



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

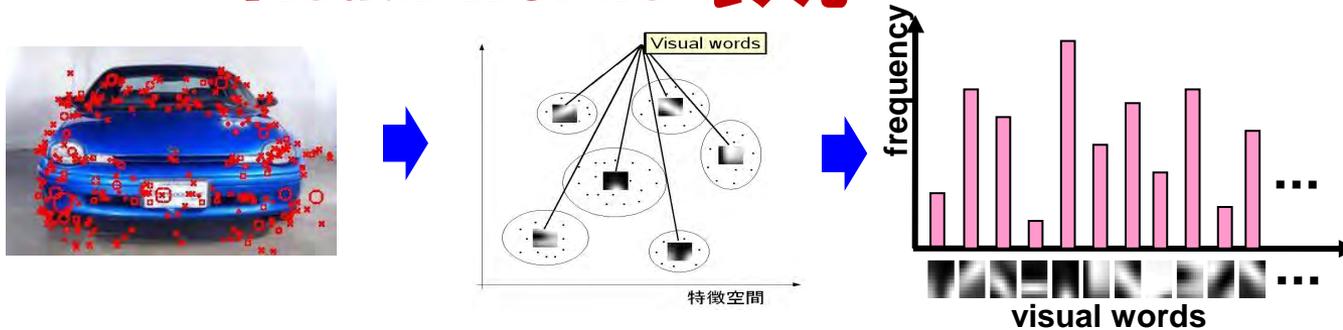
静止画像に対して, その中に含まれる
物体もしくはシーンの一般名称(カテゴリー)を認識

究極的には人間以上に幅広く詳細な認識

一般物体認識の技術的背景

■ Bag-of-Features表現 (BoF)

- 特徴点の不変特徴ベクトルのベクトル量子化表現
⇒ **Visual words 表現**



■ SVM + BoF向きカーネル による分類

- カイ2乗RBFカーネルや
ヒストグラムインターセクションカーネルの利用

$$k_{CHI}(x, y) = \exp\left(-\frac{1}{\gamma} \sum \frac{|x_i - y_i|^2}{|x_i + y_i|}\right) \quad k_{HI}(x, y) = \sum \min(x_i, y_i)$$



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な画像の認識技術の必要性

画像の意味的处理. 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」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



画像認識でよく利用される 機械学習手法の例

■ SVM

■ といえずSVM.

- ・ libSVMとSVMLightのお陰!

■ カーネルを工夫する

- ・ *Spatial pyramid Kernel, Multiple Kernel learning*

■ Boosting

■ デジカメの顔画像認識は, Boosting によるもの

■ (テキスト向け)確率トピックモデル

■ PLSA, LDA, HDP, それらの拡張(ノンパラベイズなど)

■ あと, Nearest Neighbor ⇒ 最近, 再注目!



一般物体認識の分類 (1)

■ 画像全体の 카테고리 分類



→ クマ



→ トラ



→ ソウ

■ 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

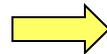


→ ソウ
キバ
空
草
草原



一般物体認識の分類 (2)

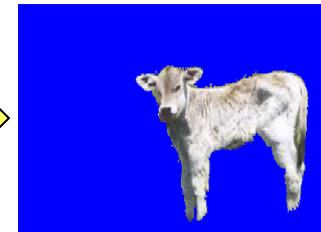
■ 画像ラベリング: 領域分割 → 分類



■ 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

物体カテゴリー認識

空

建物 / ビル

木 / 桜

木 / 桜

木 / 桜

建物

外灯

バス

信号機

自転車

道路

自動車 / バン

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?



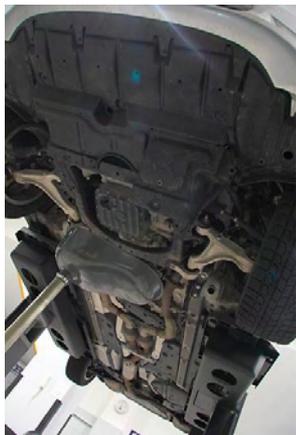
様々な「ライオン」

カテゴリー内変化(1) :

いろいろな「椅子」



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な方法。**

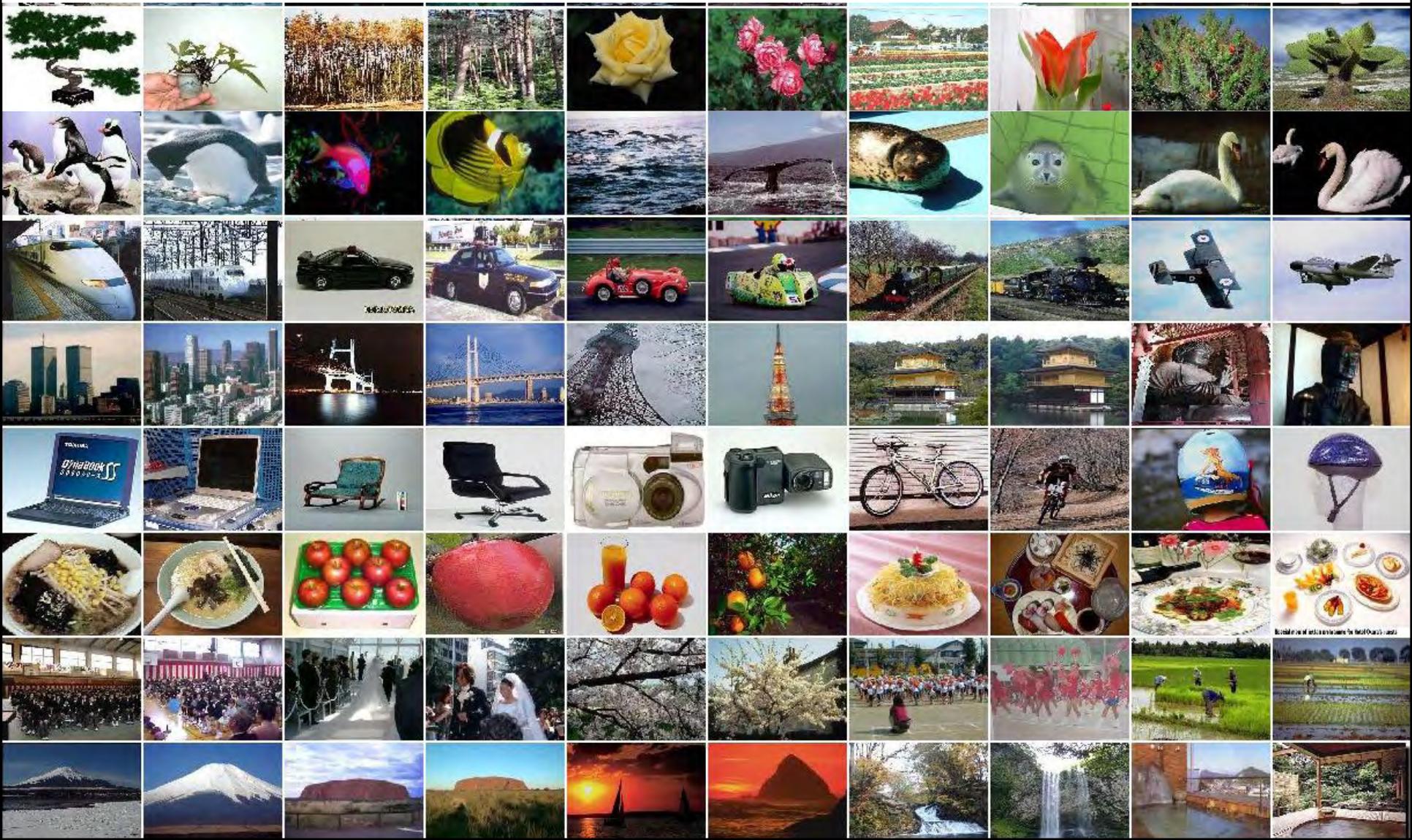


様々な視点からの見た目の学習が必要



カテゴリー数が多い：

多様なカテゴリー



「カテゴリー問題」

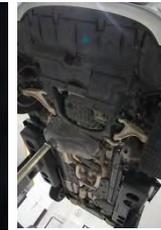
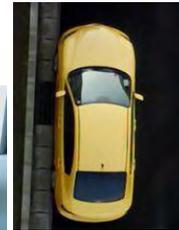
[1] どの程度のカテゴリー内変化に対応すべき？

■ どちらが「バイク」？



■ 一般的な視点とは？

データセットの作り方によって
難易度がいくらでも変わる。



[2] 何種類のカテゴリーに対応すべき？

■ 人間は、数万種類に対応。

・ 256？ 1024？ 辞書に出てるだけ(数万)？ 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」

- 実世界に対応する知能システムを構築するには、あらゆる可能性を考慮する必要がある。 e.g. 椅子の形は無限にある。



■ 「椅子」認識システムは構築不可能！？



- DBに含まれる「椅子」のみを認識できればといあえずいいにことにして、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」

- **問題点**：標準データセット作成者の主観でカテゴリー内変化やカテゴリー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(1万語)すべてについてデータセットを作成 [Tinyimage 2009][imagenet.org 2009]
 - ・ Web(Flickr)から数百万枚収集して画像アノテーション

「Webに存在する知識 ≒ 人間の一般的知識」とみなすことにより対処。



一般物体認識は「人工知能問題」

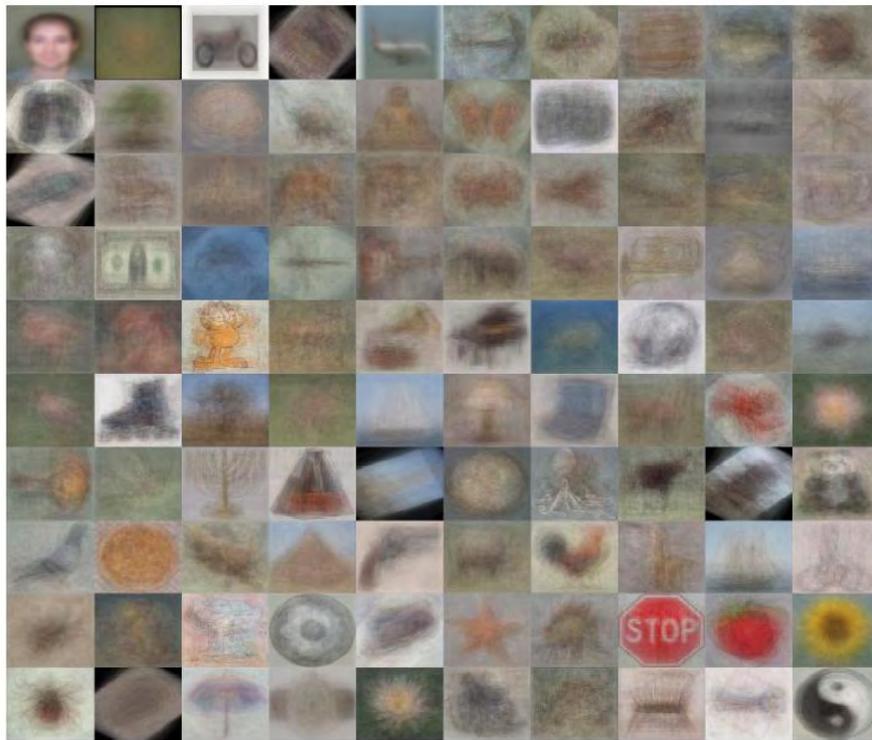
- 究極的には、人間が「○○」と思う物は、計算機も同様に「○○」と認識するべき。
 - 人間の言語と対応したカテゴリーの認識。
 - 特定物体認識の場合は、カテゴリー(クラス)がない。
- 一般物体認識：人間が基準。人工知能問題。
 - クラス定義があいまい。フレーム問題。
- 特定物体認識：全く同一の物を認識。
 - 問題が明確。アルゴリズムの問題。

理想の一般物体
認識システム



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)



各カテゴリーの平均画像

平均画像からカテ
ゴリが分かる! →
カテゴリー内変化が小

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.)

242.watermelon



171.refrigerator



093.grasshopper



162.picnic-table



014.blimp



257.clutter

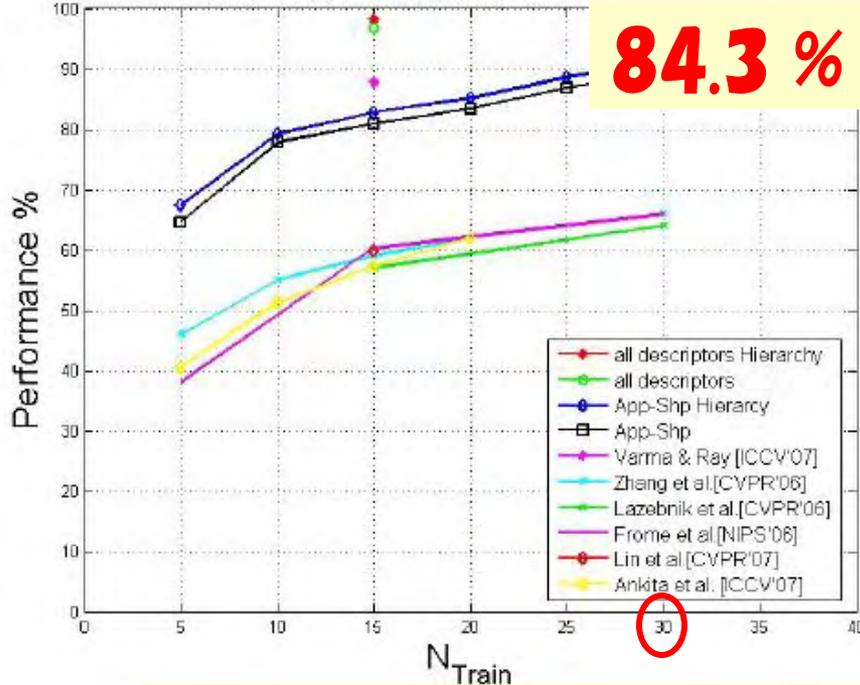




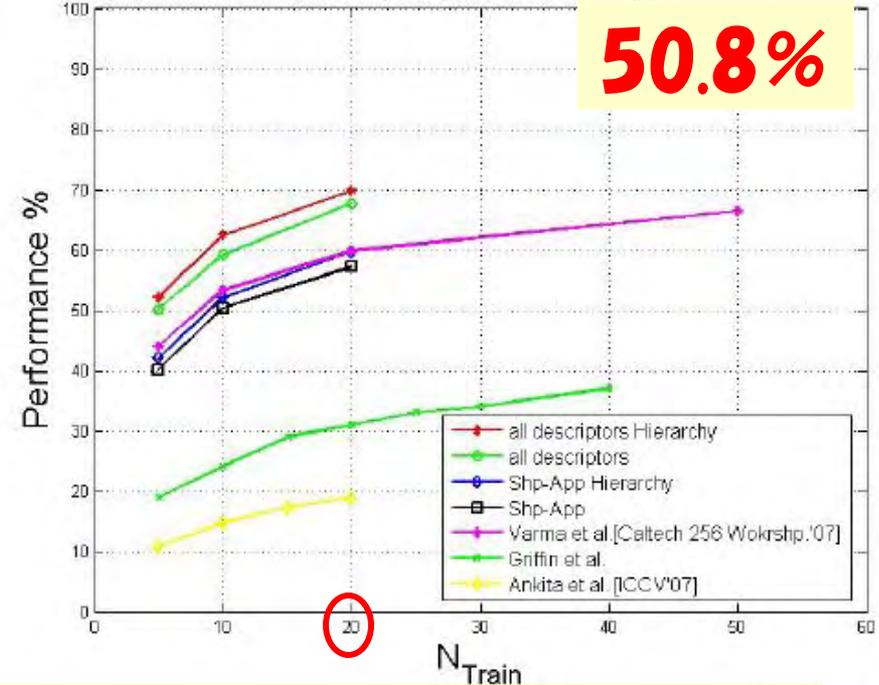
最新の一般物体認識の分類精度

■ Caltech-101 / 256 [Varma et al. 2007]

state-of-art - Caltech 101



state-of-art - Caltech 256



Caltech-101 / 256 の認識精度は、年々少しずつ上昇。

BoF+MKL が現在の最高精度 → 新しいブレイクスルーが必要!



実際の研究用データセット(3)

- Caltech-256などの大規模セットの問題点
 - 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れている。

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!?

beer mug coffee mug



airplane
fighter-jet



実際の研究用データセット(3)

- Caltech-256などの大規模セットの問題点
 - 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れる「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!?





PASCAL Visual Object Challenge

(競争型のワークショップ)

2005年～

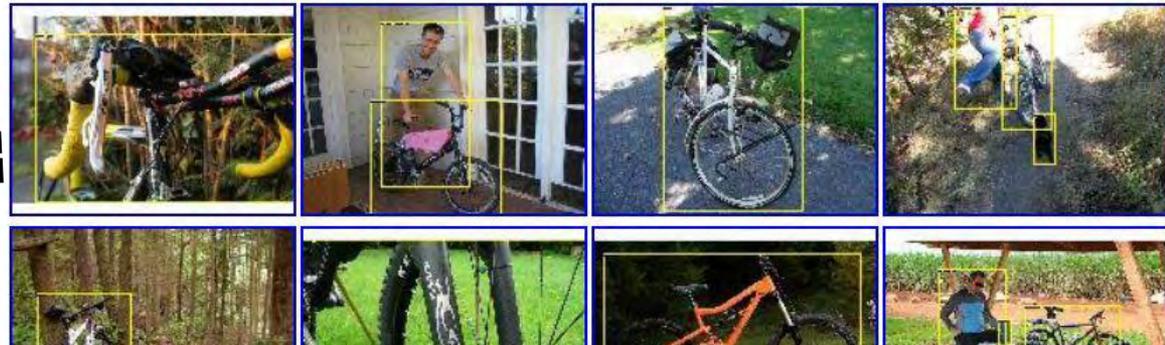
Aeroplanes - all images contain at least one aeroplane.

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。

Bicycles - all images contain at least one bicycle.



Caltechよりも多様な画像。主にFlickrから収集した画像

機械学習の期待される役割

- 少ない学習サンプルから高い汎化性能
 - 特徴空間における「椅子空間」の範囲を少ないサンプルから学習



「これは椅子？」



「椅子空間」

2. 特徵表現： Bag-of-Features (BoF)

【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.

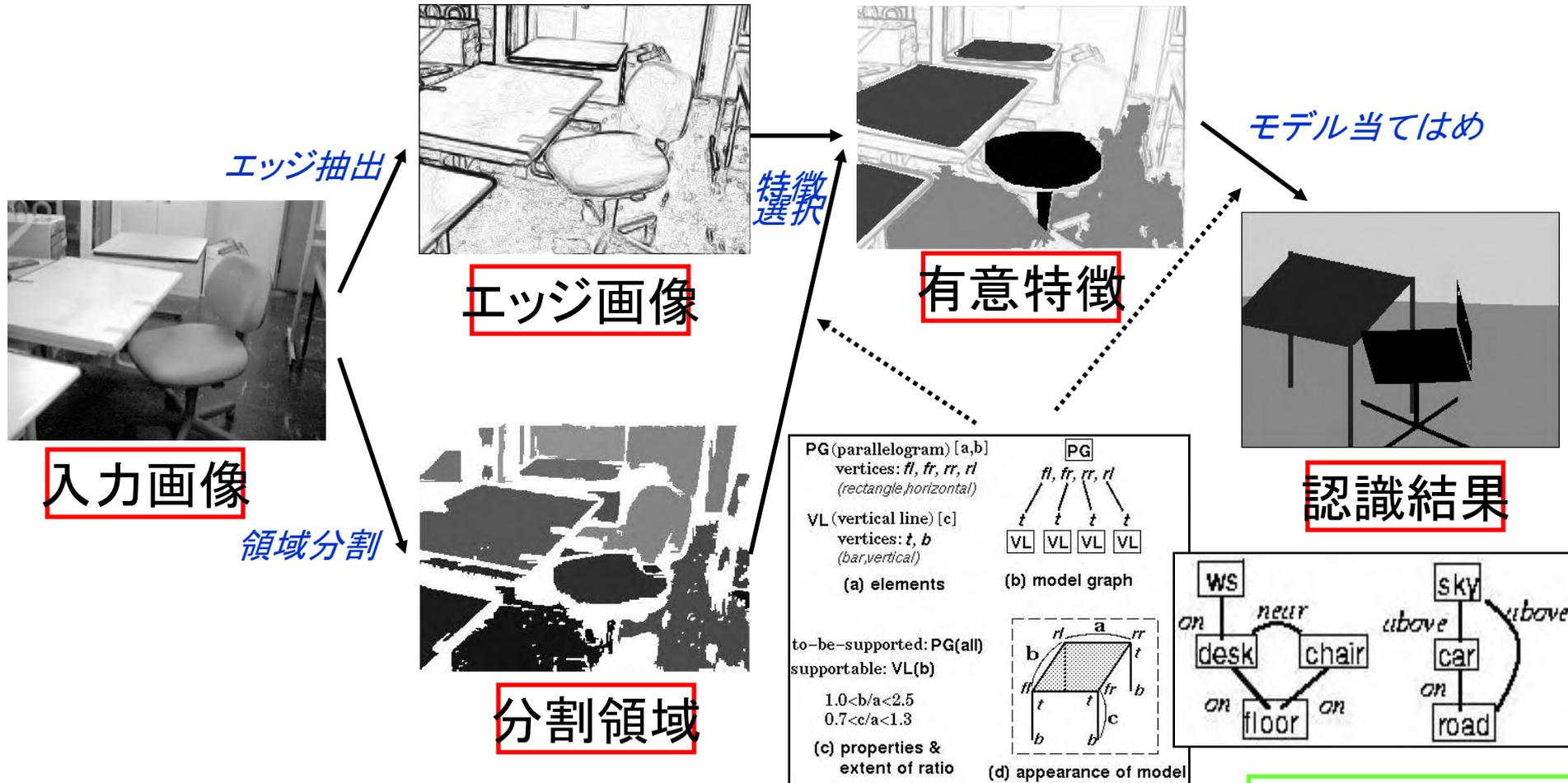
- 顔画像認識(Eigenface)の成功。固有空間法。
- 画像DBにおける画像の意味的分類。

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)



昔は「学習」がなく、人手によって記述されたルールで認識。



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 search) **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 kernels

BoF + CRF 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

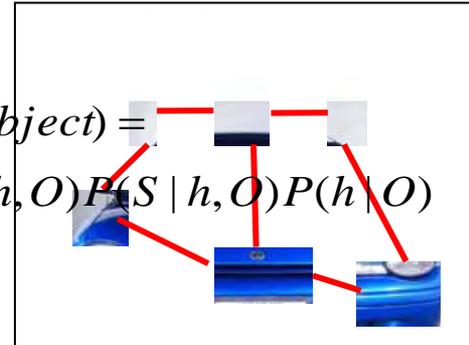
- 従来の認識: 認識対象の全体を利用
 - 固有空間法, 領域分割を用いた方法



- オクルージョンや変形に弱い。

Constellation model

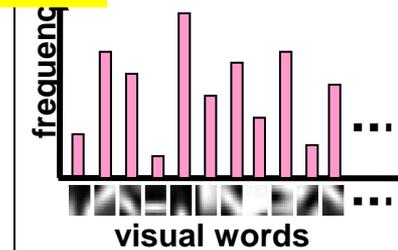
$$P(\text{Parts, Locations, Scale} | \text{Object}) = \sum P(P | L, S, h, O)P(L | S, h, O)P(S | h, O)P(h | O)$$



- Part-based手法の登場:
複数の部分の組み合わせで認識

位置関係を考慮すると, 視点変化に弱い!

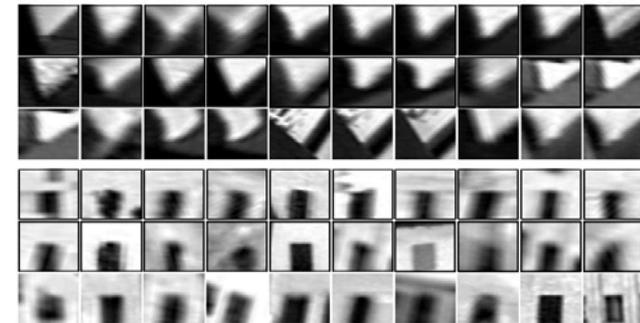
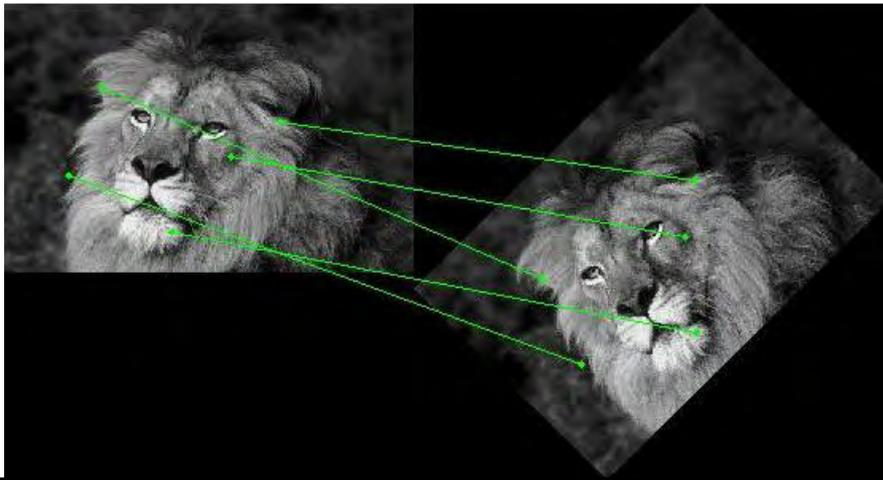
- 局所パターンの分布に基づく認識 (*bag-of-features*)



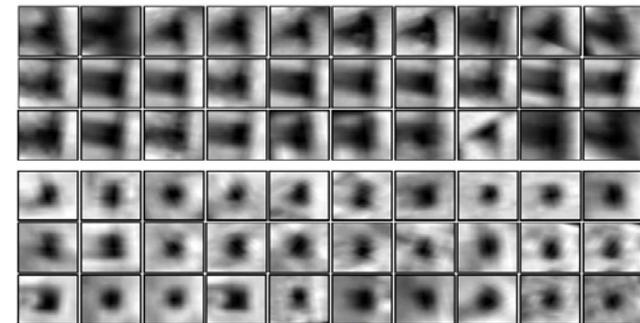
Bag-of-featuresモデル

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, 回転・拡大・位置によらず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.



(a)



(b)



局所特徴による特定物体検索



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:

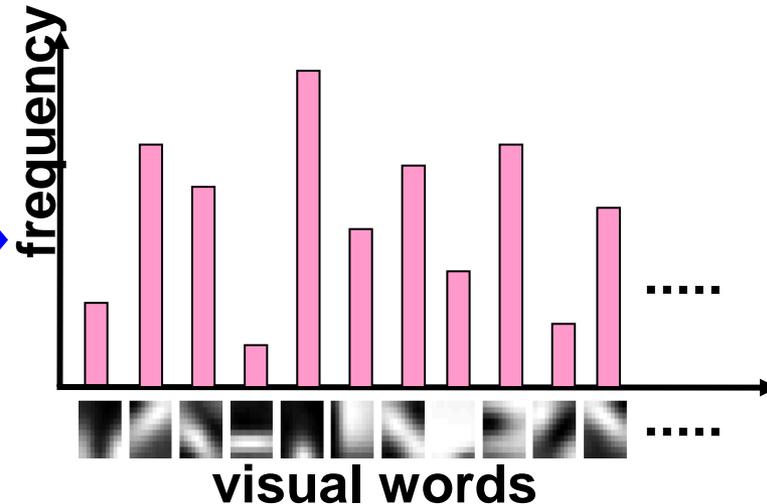
bag-of-features表現への変換

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。
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。

SIFT法
(特徴点抽出+記述)

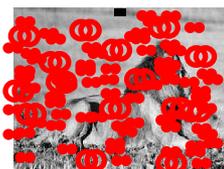
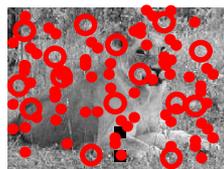
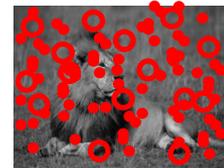




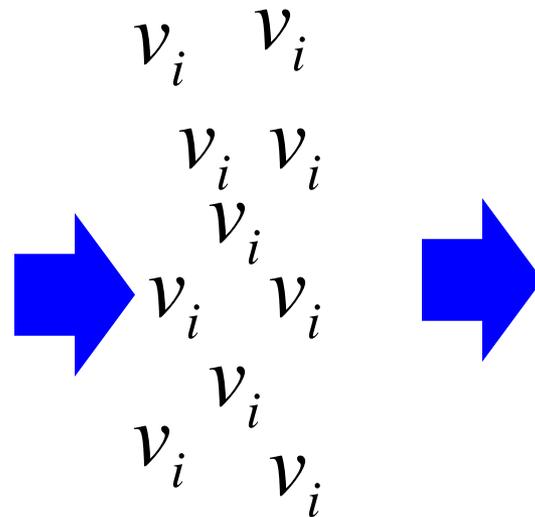
Visual words の求め方

特定と同じ

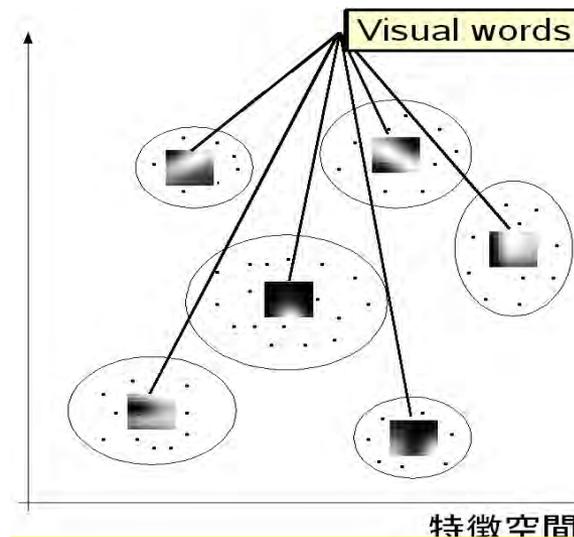
- 学習画像(正例, 負例)を用意し, SIFT特徴ベクトルを全画像から抽出 (枚数が多い場合は, ランダムサンプリング)
- k-means クラスタリングを実行



各クラスタの中心が “visual words”



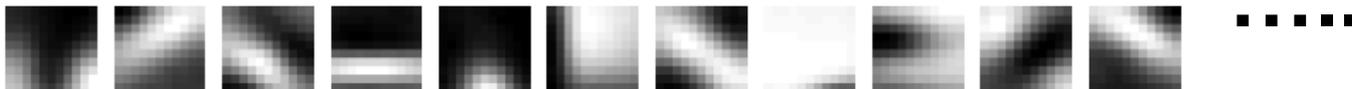
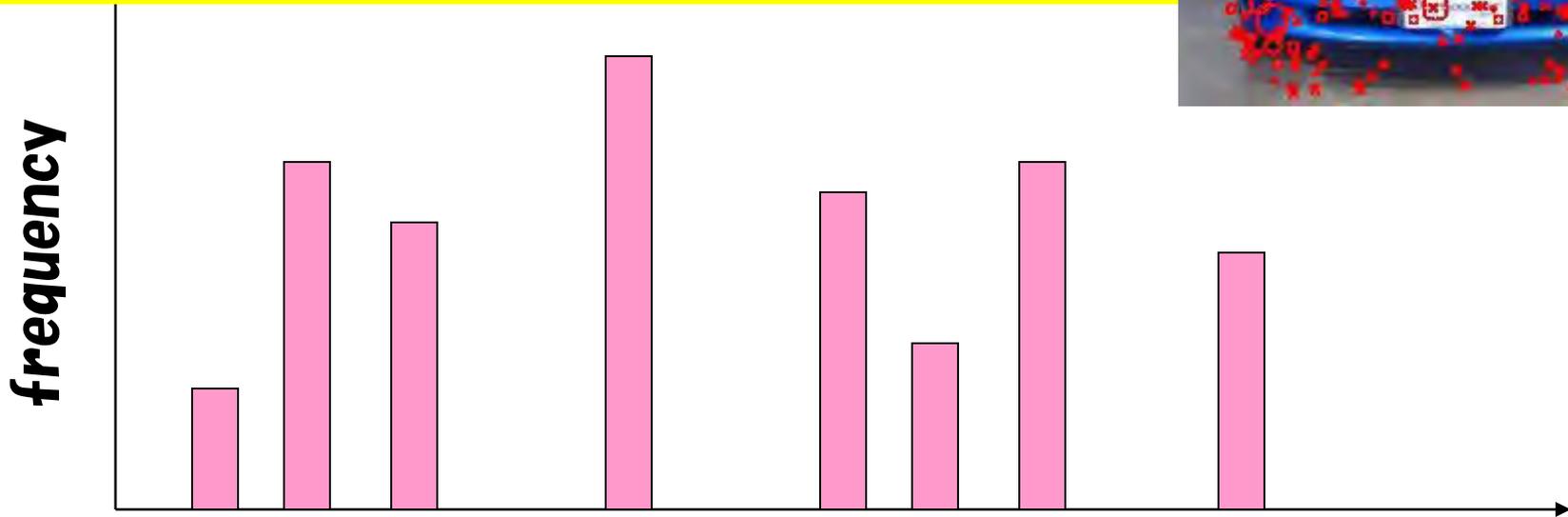
SIFT vectors



“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: 代表パターンの集合による表現

Object



Bag of 'words'



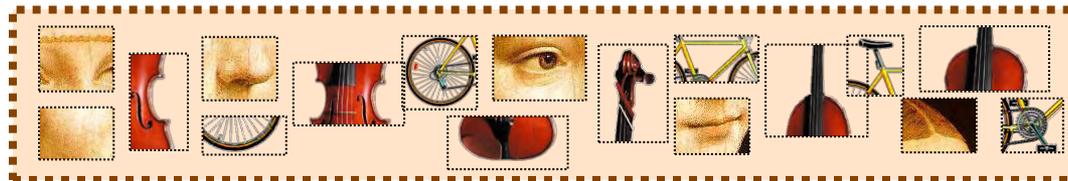
[ICCV 2009 tutorial スライドより]





同一カテゴリ：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





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

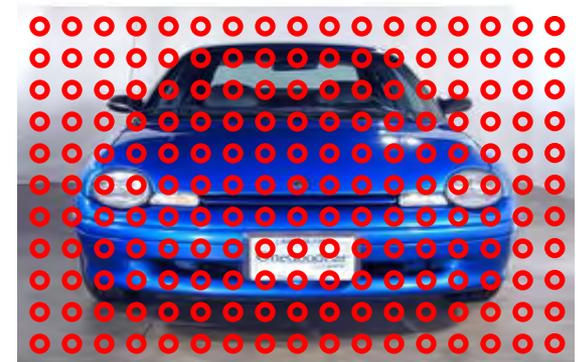
- Difference of Gaussian(DoG) **SIFT法の方法** sparse sampling と呼ぶ
 - Random sampling
 - Grid sampling
- } dense sampling と呼ぶ



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。



Bag-of-features (BoF) 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1](#), [ジャンプ2](#))



[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 lib⁷⁸)

3. 機械学習手法

【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ベクトルの分類

- **機械学習手法(特にテキスト向き)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ますます。
 - ・ カイ2乗RBFカーネルが, 最も良いと言われている。
 - ・ 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. **グリッドからランダムで, 特徴点を決定**
3. **OPENCVのExtractSURF関数で特徴点記述**
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.**



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

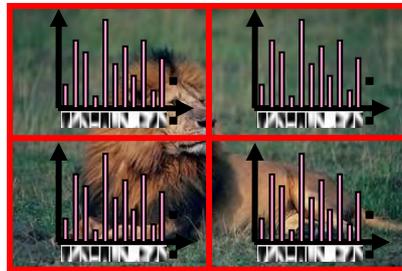
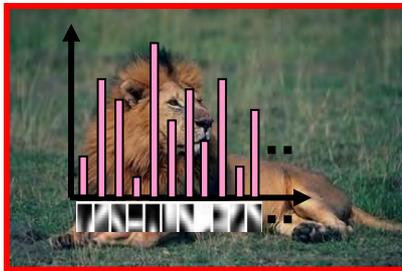
- ・ Supervised clustering
 - ・ Information bottleneck (IB) 法など
- ・ Discriminative visual words のみの利用
- ・ 空間分割と分類を交互に行う
 - ・ Unifying Discriminative Visual Codebook Generation with Classifier Training [Yang et al. CVPR 2008]
- ・ 他にもいろいろ多数. カバーし切れません!



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

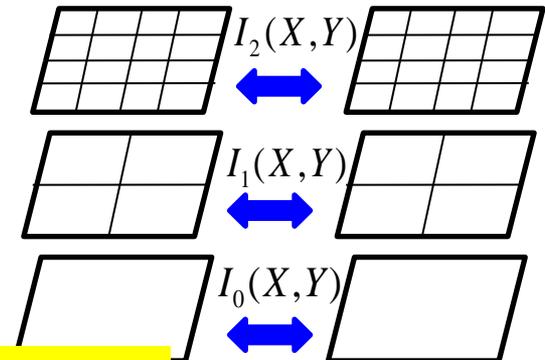


- 各レベルごとにヒストグラムインターセクションを求め、レベルごとに異なる重みで統合. SVMのカーネル関数とする.

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

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

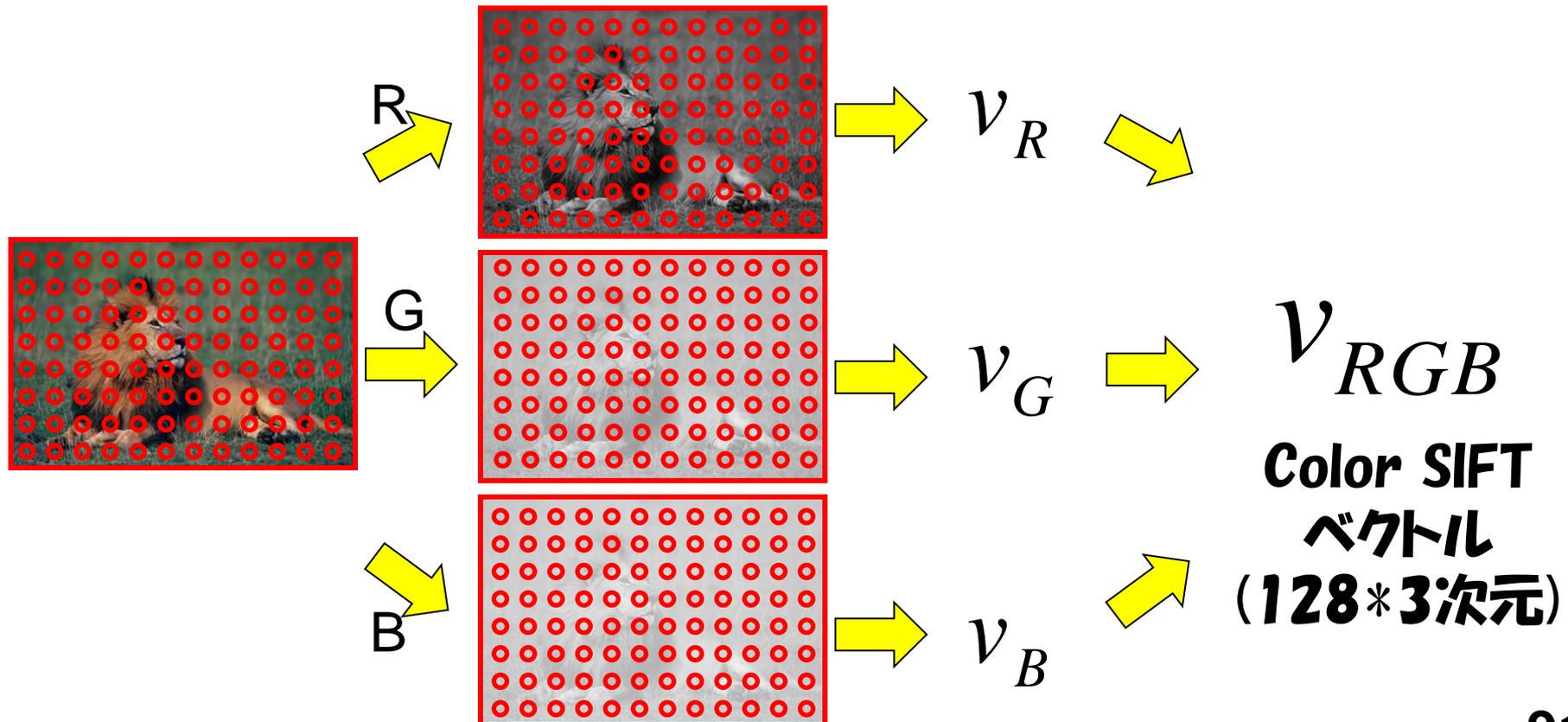
I_l : Histogram intersection in level l



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 K_1, \dots, K_{N_k} とすると,
統合カーネルは,
$$K_{\text{opt}} = \sum_k d_k K_k$$

ただし, $d_i (i = 1, \dots, N_k)$ 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



画像認識方法

- **Multiple Kernel Learning(MKL) [Lanckrietら,2004]**
を用いた特徴統合によるマルチクラス分類手法
(SHOGUNライブラリの利用, 1-vs-restによる)
- **統合する画像特徴**
 - **局所特徴のBag-of-Keypoints(6通り)**
 - **カラーヒストグラム**
 - **ガボール特徴のヒストグラム(2通り)**



局所特徴

- SIFTを使用
- 特徴点検出は3タイプ
- 画像はBag-of-keypoints表現を用いて局所特徴の出現頻度のヒストグラムで表現
 - 1000次元と2000次元で表現



DoG検出



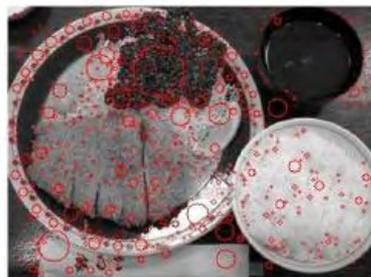
グリッド検出



ランダム検出



SIFT



Bag-of-keypoints表現



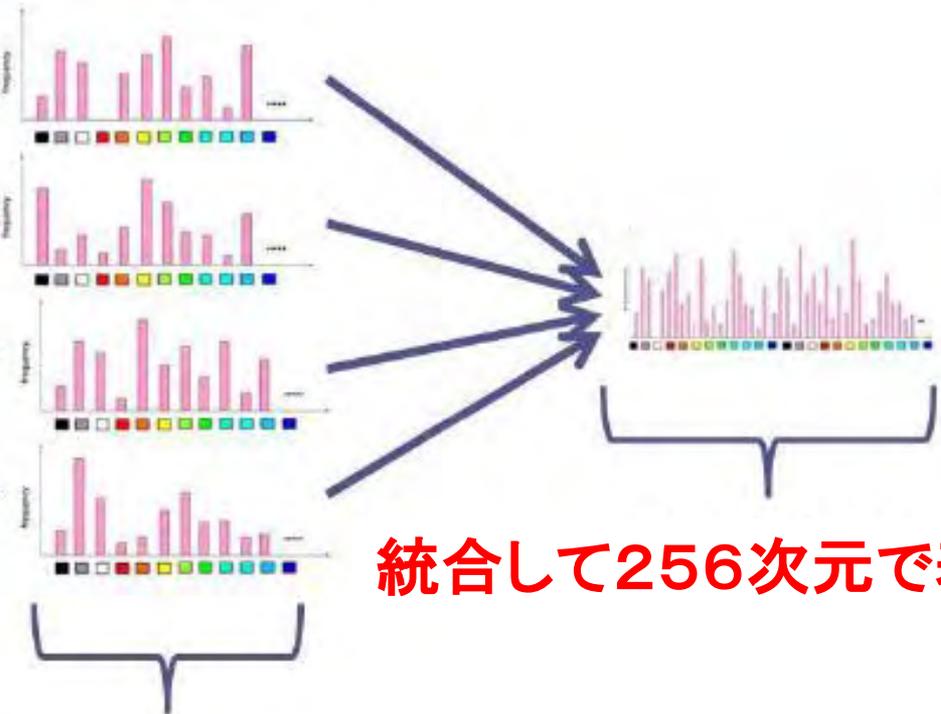
ヒストグラム

カラーヒストグラム

■ RGB色空間を使用



画像を2x2に分割



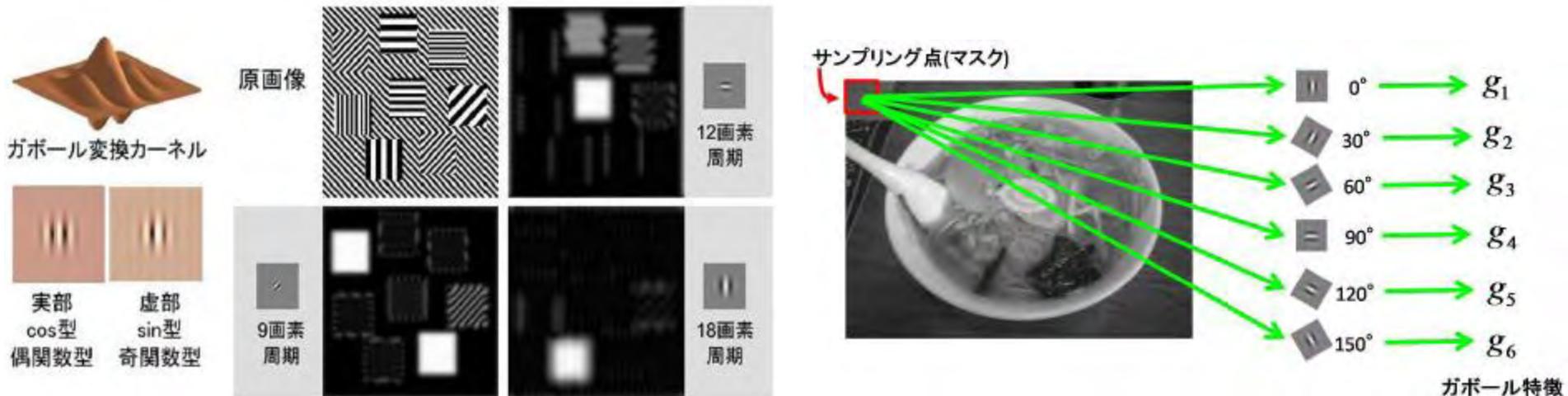
統合して256次元で表現

色空間を64色に量子化



ガボール特徴

- 画像化から局所的な濃淡情報の周期と方向を表現した特徴
- 6方向、4周期のガボールフィルタで特徴抽出
- 色特徴と同様に画像を3x3と4x4に分割
 - 216次元と384次元のベクトルで表現





評価実験

- **データセット**: 50種類食べ物画像セット
- **画像特徴**: 3タイプの画像特徴(計9種類)
- **手法**: 各特徴単独のSVM
全特徴によるMKL-SVM
 - MKLの実装は機械学習ライブラリ SHOGUNを利用
- **カーネル**: χ^2 カーネル
$$K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{1}{2\sigma^2} \sum_i \frac{\|x_i - y_i\|^2}{x_i + y_i}\right) \gamma$$
 - **スケールパラメータ γ の設定方法**
 - ① Cross-validationにより, 最適な値を探索
 - ② すべての学習データ間の χ^2 距離の平均の逆数

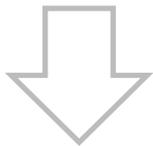
[Zhang, 2007]





実験結果

MKLによる特徴統合方法
と
特徴単独で用いた手法
の50種類の平均分類率



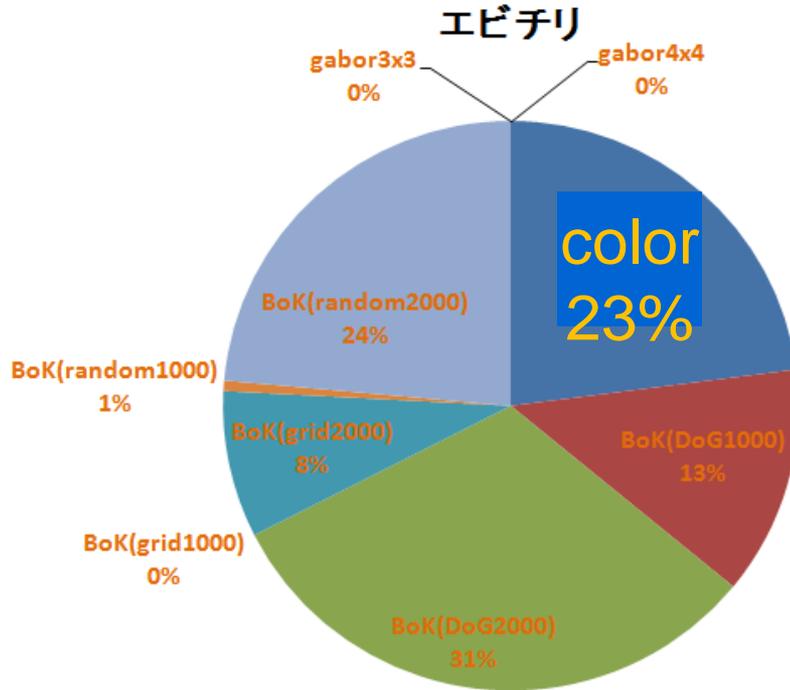
大幅に精度向上した

特徴	平均分類率
BoK(dog1000)	26.52%
BoK(dog1000)	27.48%
BoK(grid1000)	26.10%
BoK(grid2000)	27.68%
BoK(random1000)	28.42%
BoK(random2000)	29.70%
gabor3x3	31.28%
gabor4x4	34.64%
Color	38.18%
MKL(特徴統合)	61.34%

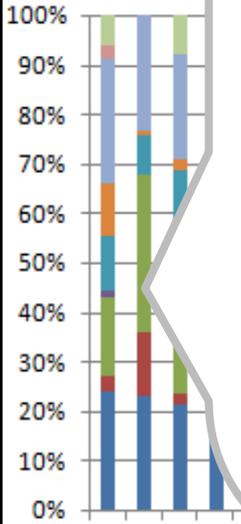
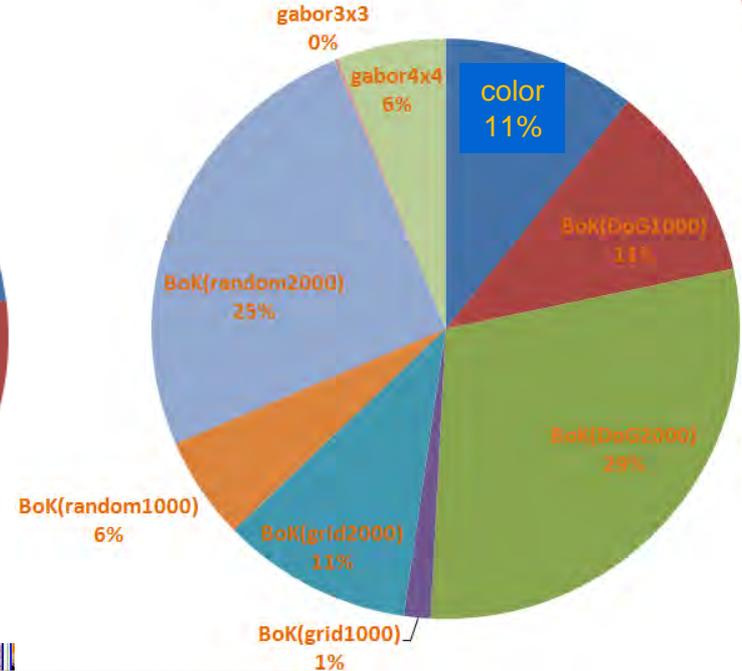


MKLで学習した特徴の重み

エビチリの結果



平均の重み

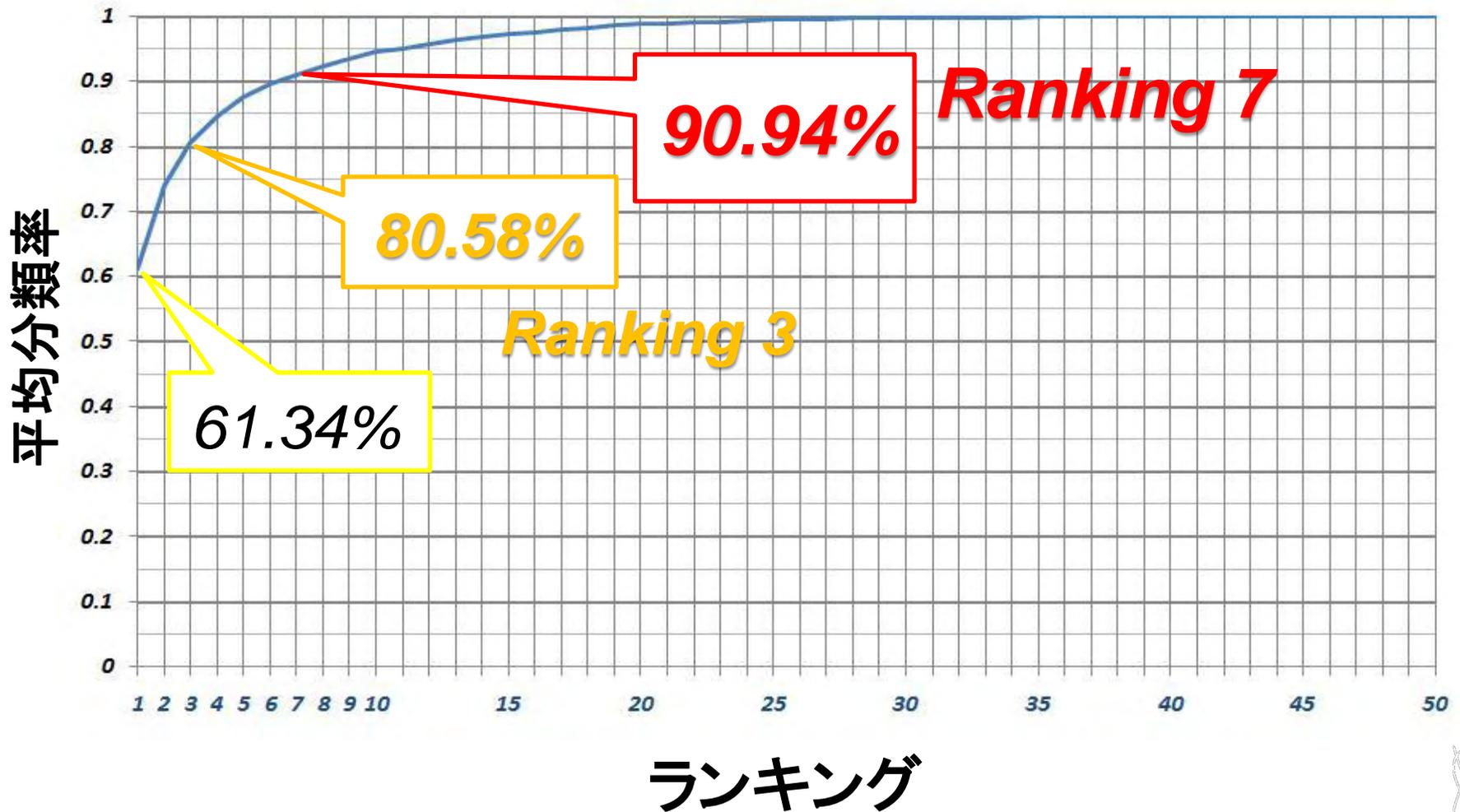


- オムライス
- エビチリ
- 味噌汁
- ポターシユ
- ビーフカレー
- つな重
- ピザ
- みるそば
- シヤオロース
- やし中華
- アキ
- 鮮井
- エウマイ
- ハゲティ
- かし寿司
- ルキャベツ
- い焼き
- バーガー
- 可
- 前煮
- サ
- ーメン
- 焼き
- ロッパン
- ルパン
- バーグ
- じゃが
- エー
- 子丼
- ンドウィッチ
- ツカレー
- アハン
- 麦焼き
- 肉ヨロケ
- ースカッ
- 丹
- チ
- フライ
- ーストチキン
- きそば
- き鳥
- ン丼
- ラーメン
- 煮
- ースト
- 好み焼き
- ーポ豆腐
- 春巻き
- ピラフ
- たに焼き
- 平均の重み



実験結果

許容クラス数を変化させたときの平均分類率の変化

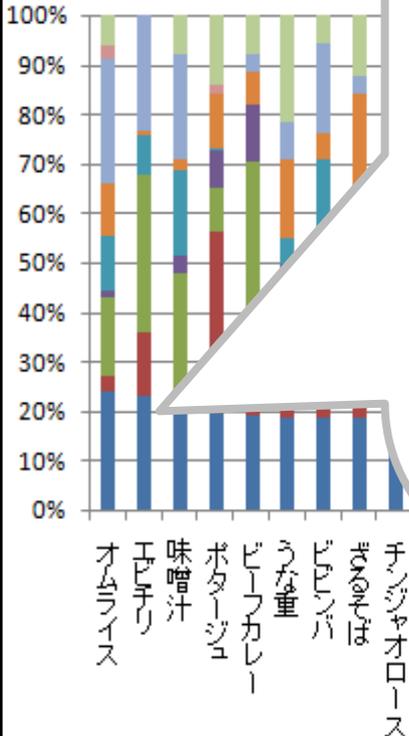
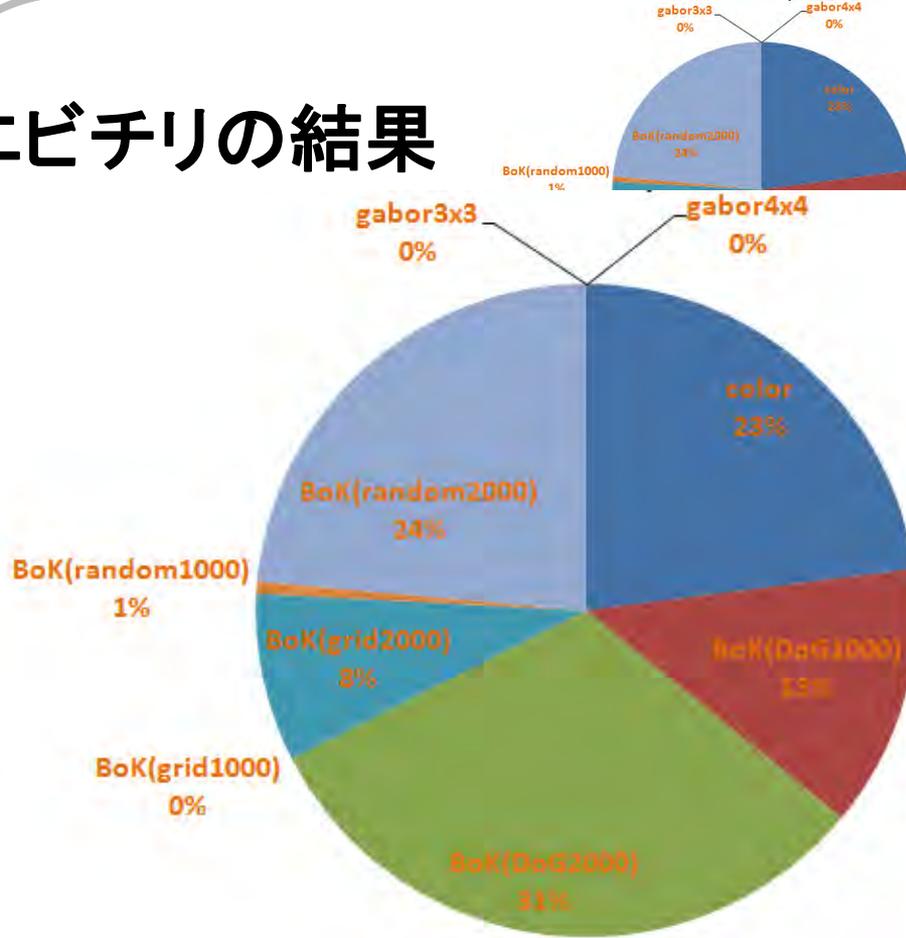




MKLで学習した特徴の重み

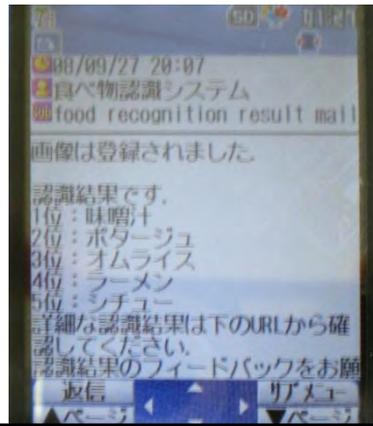
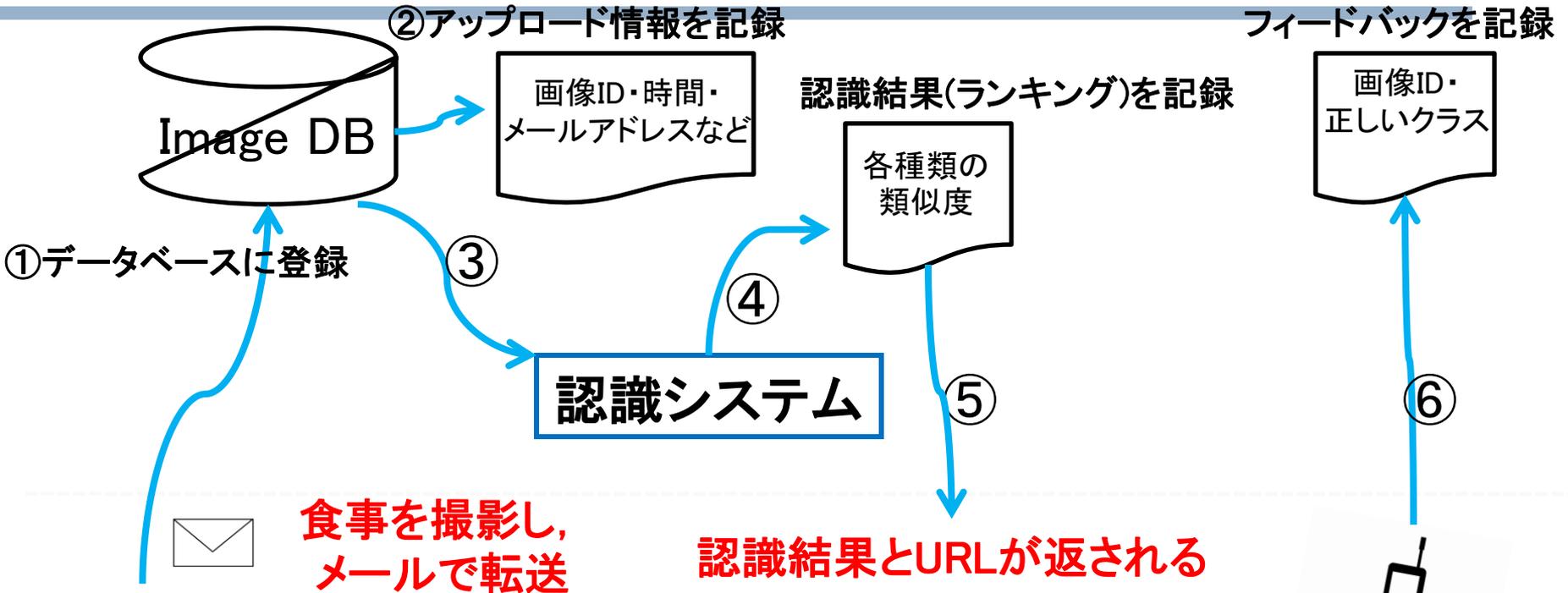
エビチリの結果

- gabor4x4
- gabor3x3
- BoK(random2000)
- BoK(random1000)
- BoK(grid2000)
- BoK(grid1000)
- BoK(DoG2000)
- BoK(DoG1000)
- color





③ 食事画像認識システム: 携帯版認識システムの構成



正しい食べ物の種類を選択





食べ物認識システム

- 画像を添付して, 送るだけ!!!

shokuji@mm.cs.uec.ac.jp

手法:

- Bag-of-features 表現 + 色ヒストグラムによる画像表現
- Multiple Kernel Learning法による特徴の複数種類の混合





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)**



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:

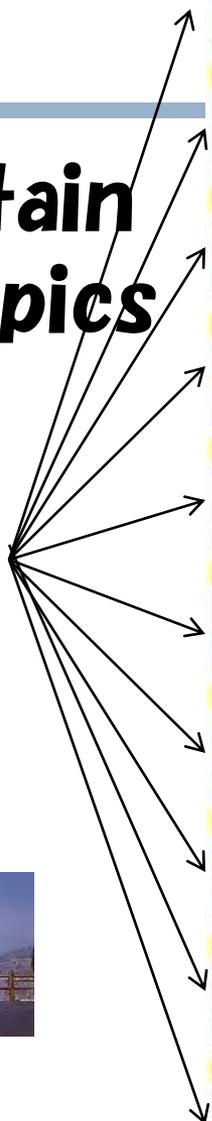
Mountain 10 topics



正例



負例



mountain
file: /plsa_grid_out/mountain_500_10.pdz (9)
list: /plsa_grid_out/mountain_500_data.list (4678)
Keypoints: GRID Method: pLSA codebook size: 500 # clusters: 10 word: mountain

P(Mountain|topic)



GMMによる認識と同じことができる. 多クラス化が容易.

確率的トピック

Mountain 10 topics

probabilistic clustering



周辺化によって mountain 確率を求める。

$$P(\text{mountain} | \text{img})$$

$$= \sum_{i=1}^k P(\text{mountain} | z_i) P(z_i | \text{img})$$

$P(\text{topic} | \text{img})$

0.15



0.24



0.03



0.05



0.10



0.01



0.01



0.01



0.05



0.30

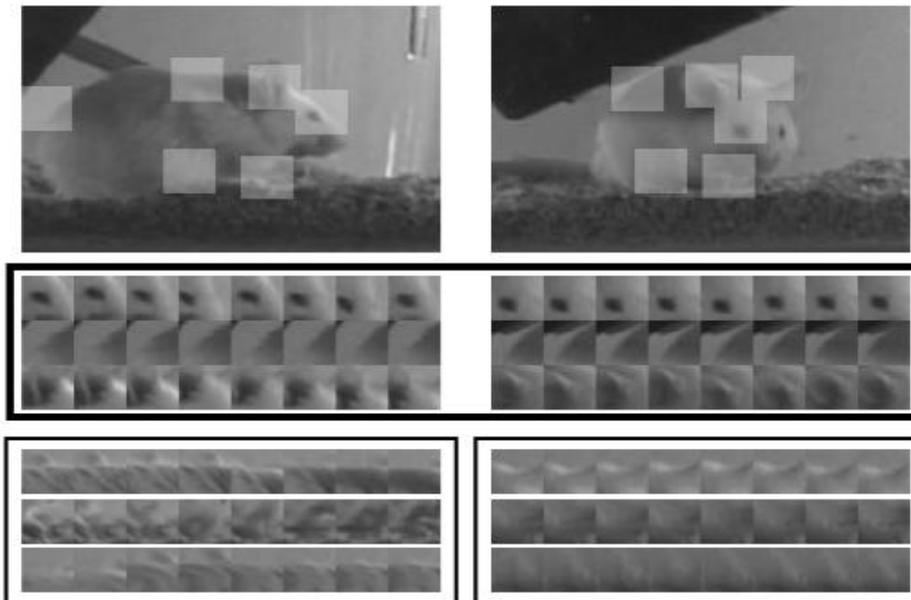




時間方向(動画)への拡張:

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

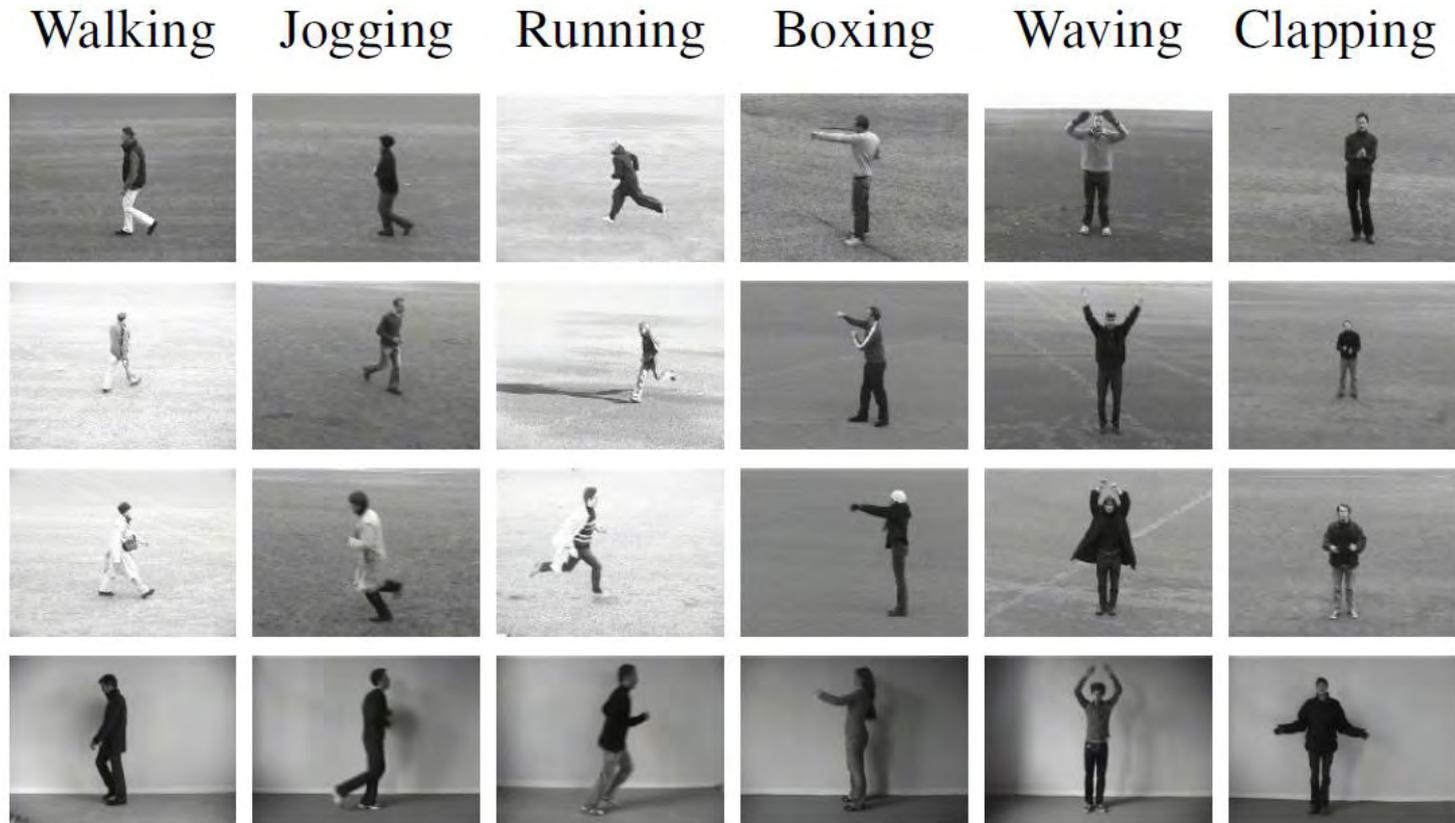


Figure 8. Sample frames from the KTH actions sequences. All six classes (columns) and scenarios (rows) are presented.



最新結果 [Lap08]

■ 分類は SVM + χ^2 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95%
- 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

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

Table 4. Confusion matrix for the KTH actions.



4. 一般物体位置検出

【参考文献】

P. Duygulu, K. Barnard, J.F.G de Freitas, and D.A. Forsyth: Object recognition as machine translation: Learning a lexicons for a fixed image vocabulary, ECCV, pp.97-112, (2002).

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

一般物体認識の分類 (1)

■ 画像全体の 카테고리 分類



→ クマ



→ トラ



→ ソウ

■ 画像アノテーション: 複数ラベルの付与



→ クマ
草
水



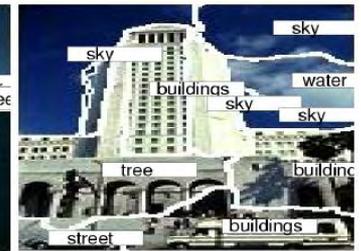
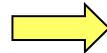
→ トラ
草
草原



→ ソウ
キバ
空
草
草原

一般物体認識の分類 (2)

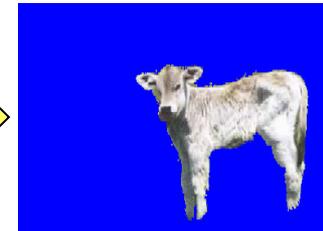
■ 画像ラベリング: 領域分割 → 分類



■ カテゴリー物体検出: ウィンドウ探索

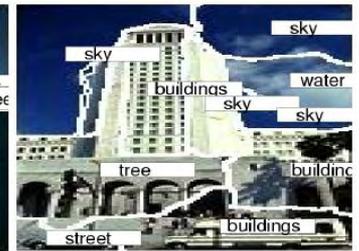


■ オブジェクト領域抽出: 認識 + 領域分割



一般物体を検出

■ 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。



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)

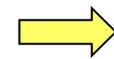
- 大量の2ヶ国語テキスト(対訳コーパス)を学習データとし、確率モデルを学習し、翻訳を行う。文法知識が不要。

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

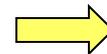


lion,
grass



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



領域への確率モデルに基づくラベリング



領域分割

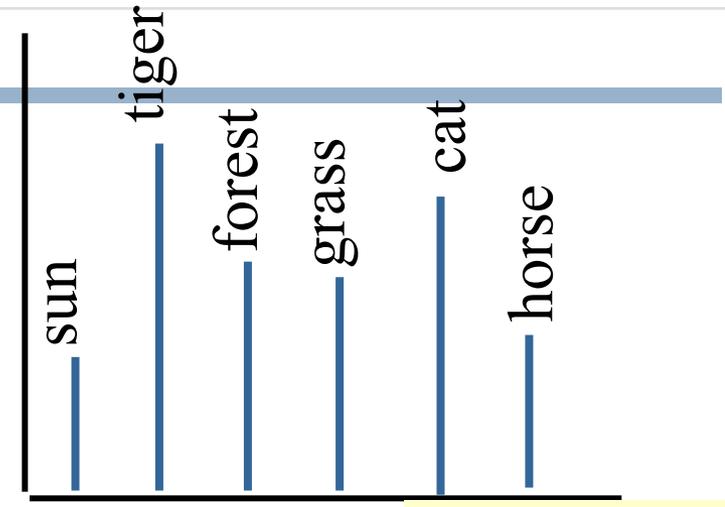


領域毎に特徴ベクトルを生成
 $P(\text{word}|\text{region})$ をGMMで表現し, 学習

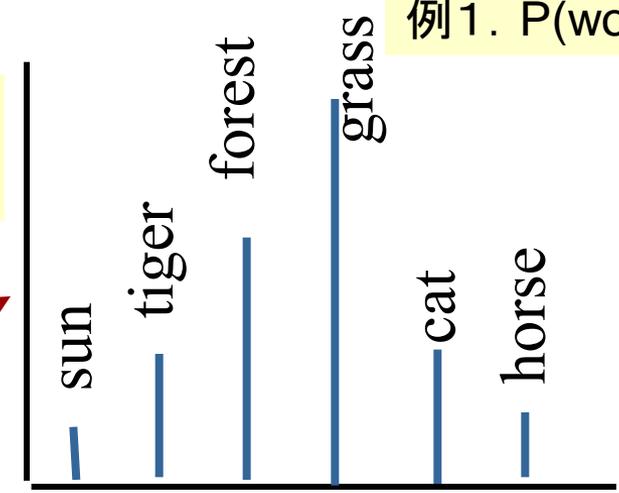


未知画像の領域の特徴ベクトルを
GMMに代入し, $P(\text{word}|\text{region})$ の値を計算.

領域毎に,
テクスチャ,
色, 形状
などの
特徴量
を抽出



例1. $P(\text{word}|\text{region}1)$

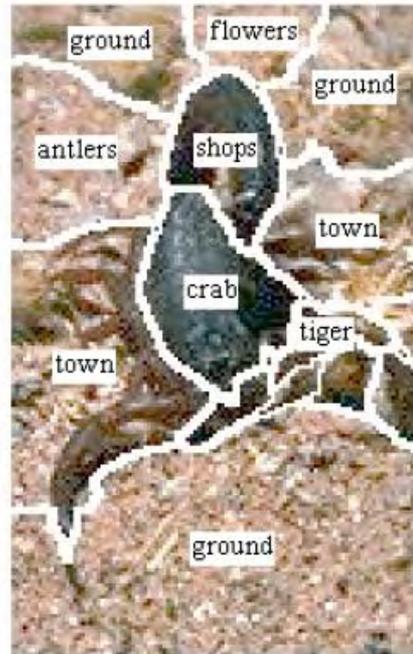


例2. $P(\text{word}|\text{region}2)$

80語でpre:0.2, rec:0.3程度 138



画像ラベリングの例





118011
WATER HARBOR
SKY CLOUDS

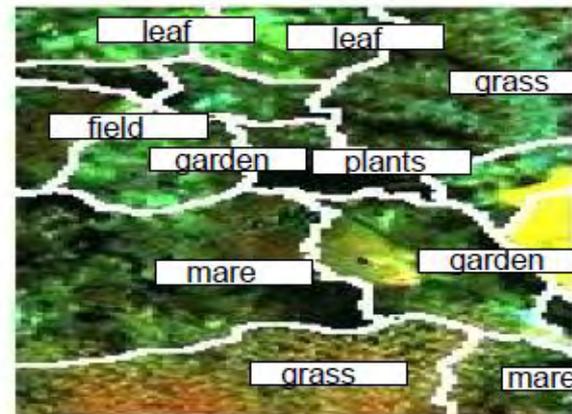
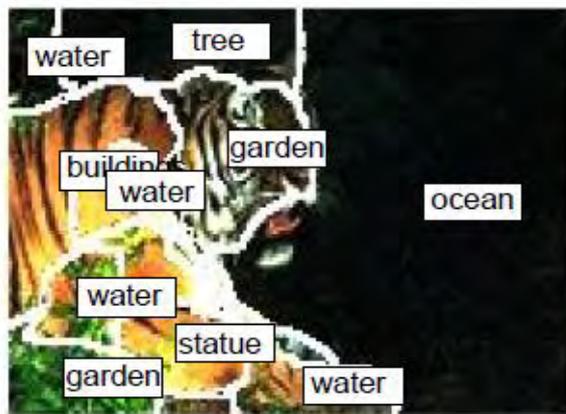
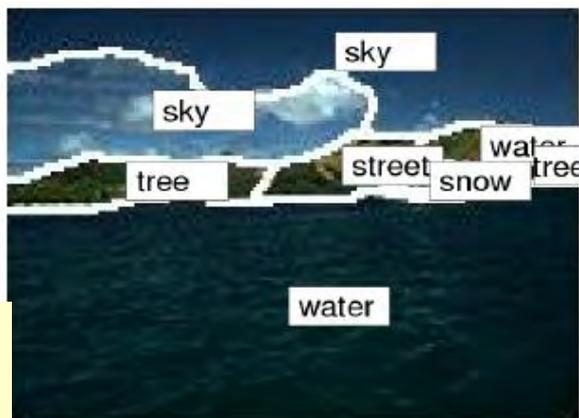
TIGER CAT WATER GRASS

143078
MOUNTAINS TREES
aspens VALLEY

1015
SUN TREE
PLAIN SKY

■ 学習画像: キーワード付きCorel画像 (4500枚)

実行結果





Translation Model の特徴

- 多数の種類 of 単語を同時に扱うことが可能 (387語)
- 個々の領域に直接, 単語を付けられる.
 - 統計的機械翻訳のモデルの利用により, 単語が多数の場合に対応.
- 精度は高くない. 精度2~3割.
- 結果が初期領域分割の結果に依存する.
 - 領域分割が容易でない画像以外ではうまくいかない.
- 大量 of 画像, 多数 of 単語を同時に扱うことができるため, visionよりもmultimediaのコミュニティーで発展研究が行われている.
 - ACM Multimedia, ACM SIGIR, IEEE ICMEなど



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

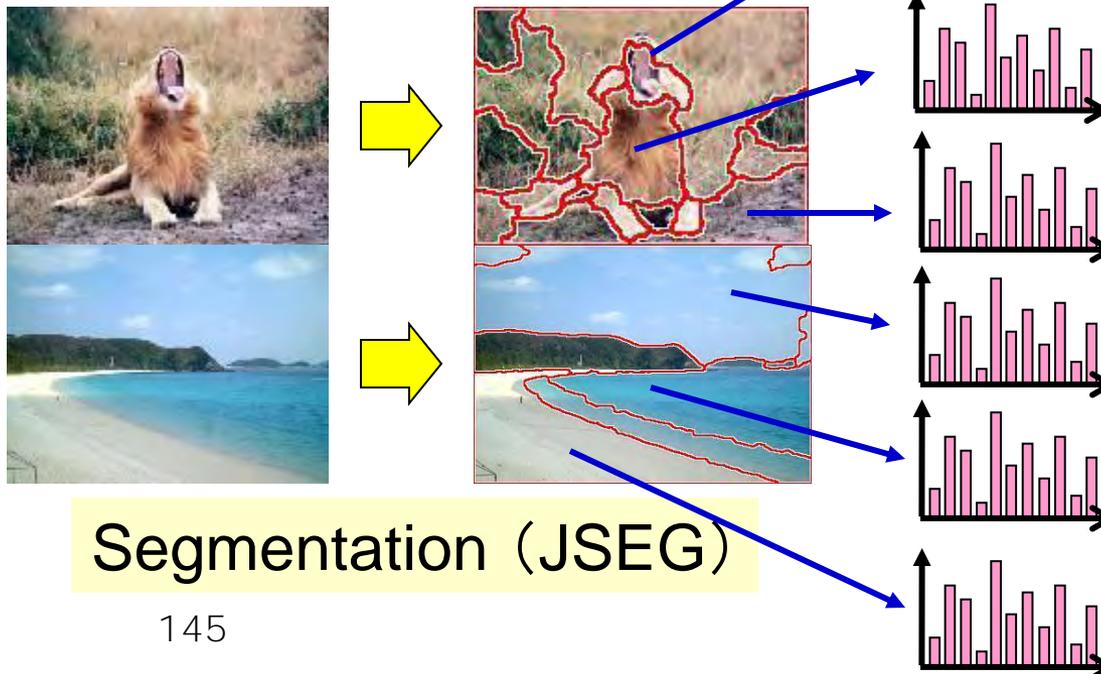
ライオンの
positive instance
はない。

- **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)**



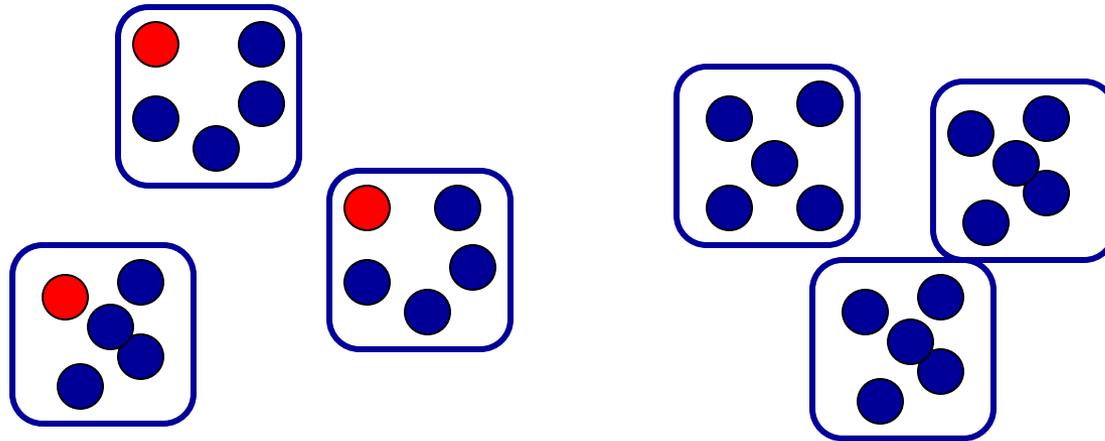
**Bag-of-features
(BoF) histograms
(1000-dim)**





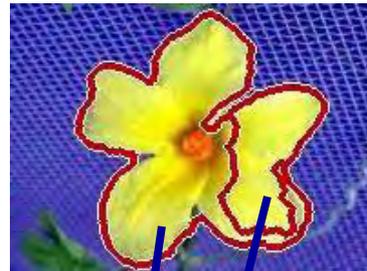
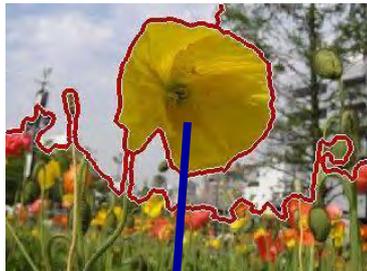
Multiple Instance Setting

■ 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

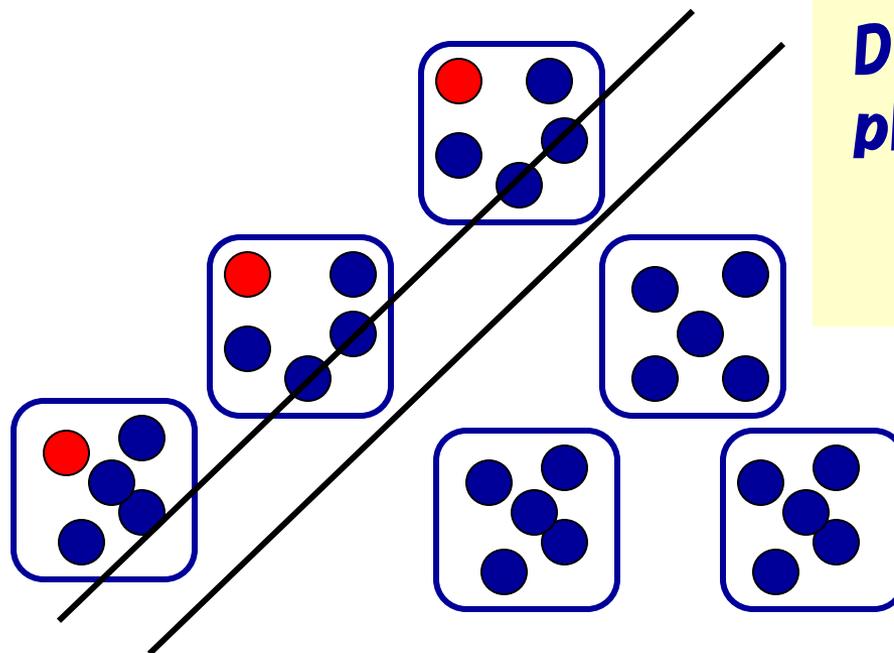




mi-SVM

[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.

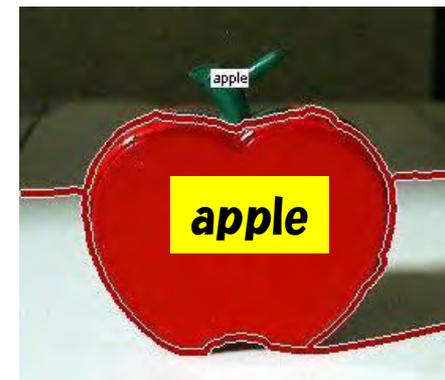
- **positive ins. (foreground)**
- **negative ins. (background)**





Final Image Re-ranking

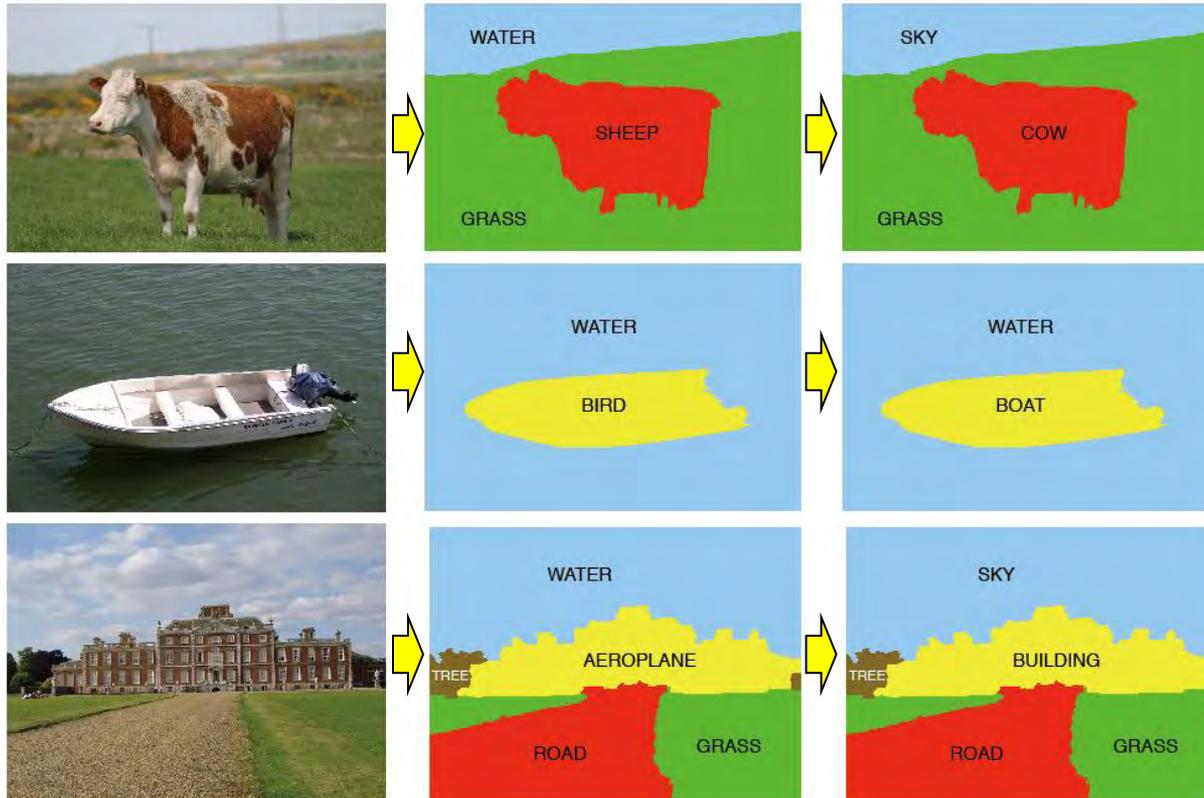
- Regard the *best SVM output score of the regions within an image as the score of the image*
 - An image having one *positive region at least is a positive image !*
- Rank images *based on the scores*





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

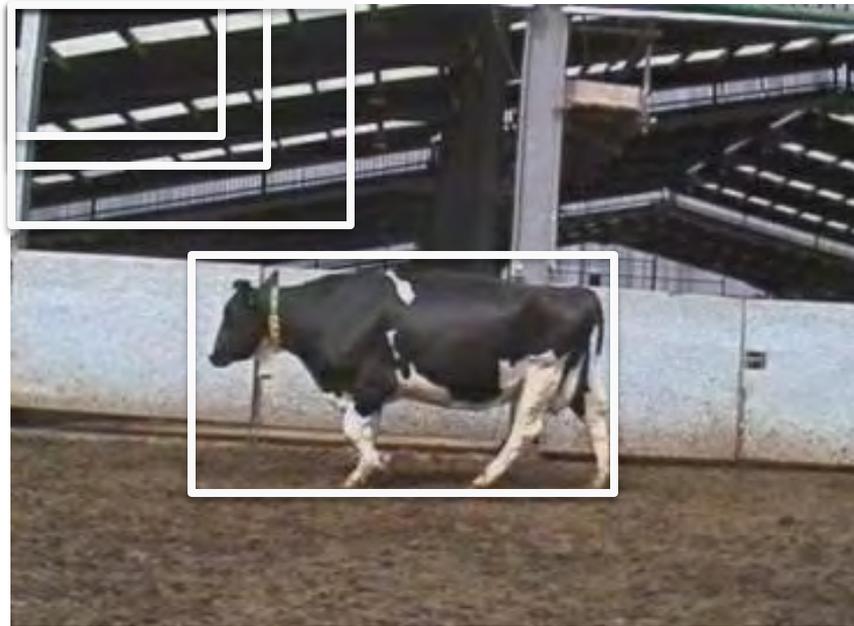
領域分割は,
Normalized Cuts.
を利用.

共起関係は
確率モデル(CRF)
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

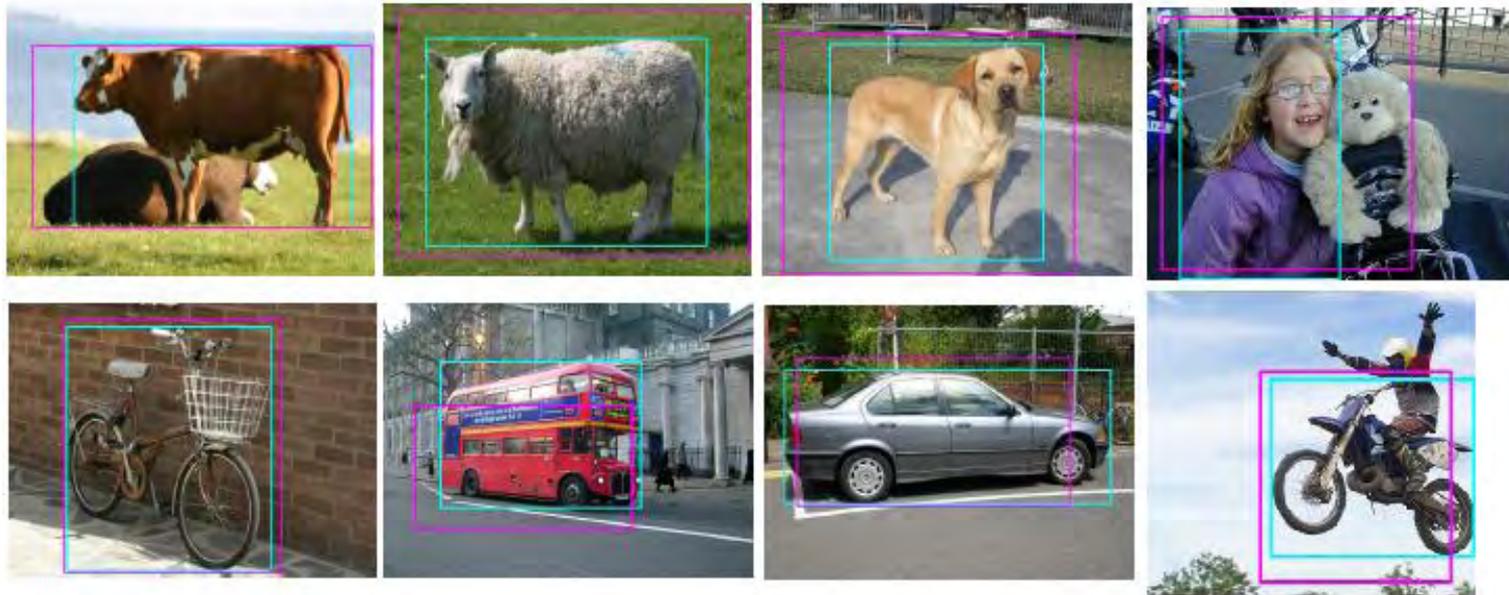


各windowに対しては画像全体に対する手法(ただし学習画像はBB付き)でよいが, 計算コストは非常に大きい!!



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.**



高速Windowサーチ: Efficient Subwindow Search [Iam08]

- **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
- **バウンディングボックス(BB)の表現**
 - 座標でなく幅を持ったBBとして表現→ $[T, B, L, R]$
 $T = [t_{min}, t_{max}]$, $B = [b_{min}, b_{max}]$, $L = [l_{min}, l_{max}]$, $R = [r_{min}, r_{max}]$
- **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



$$f(x) = 0.6$$



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, Tvm)
- ・ 学習データ 749枚
- ・ 実験データ 750枚

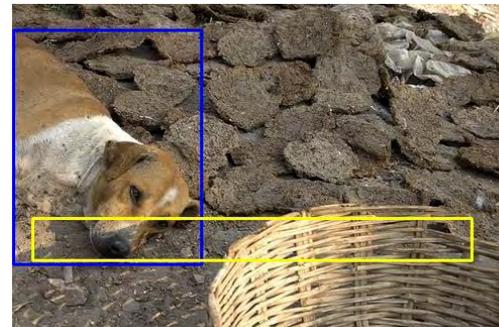
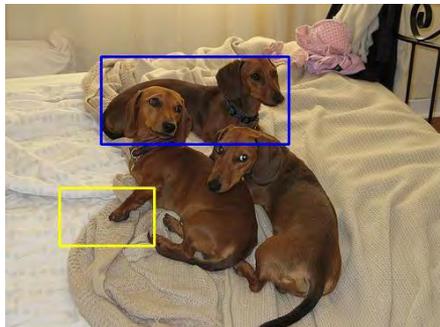
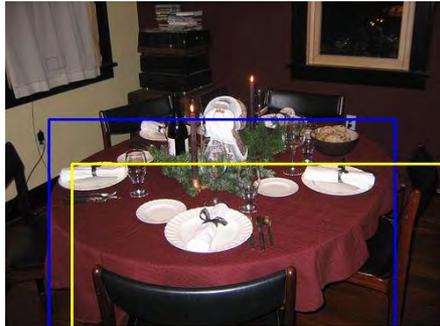
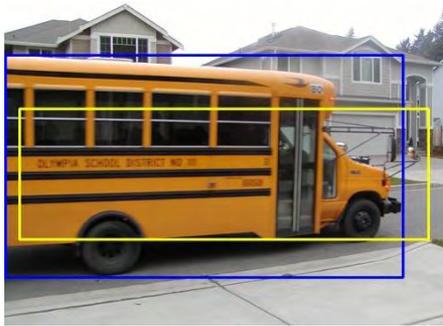
■ 実験環境

- 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

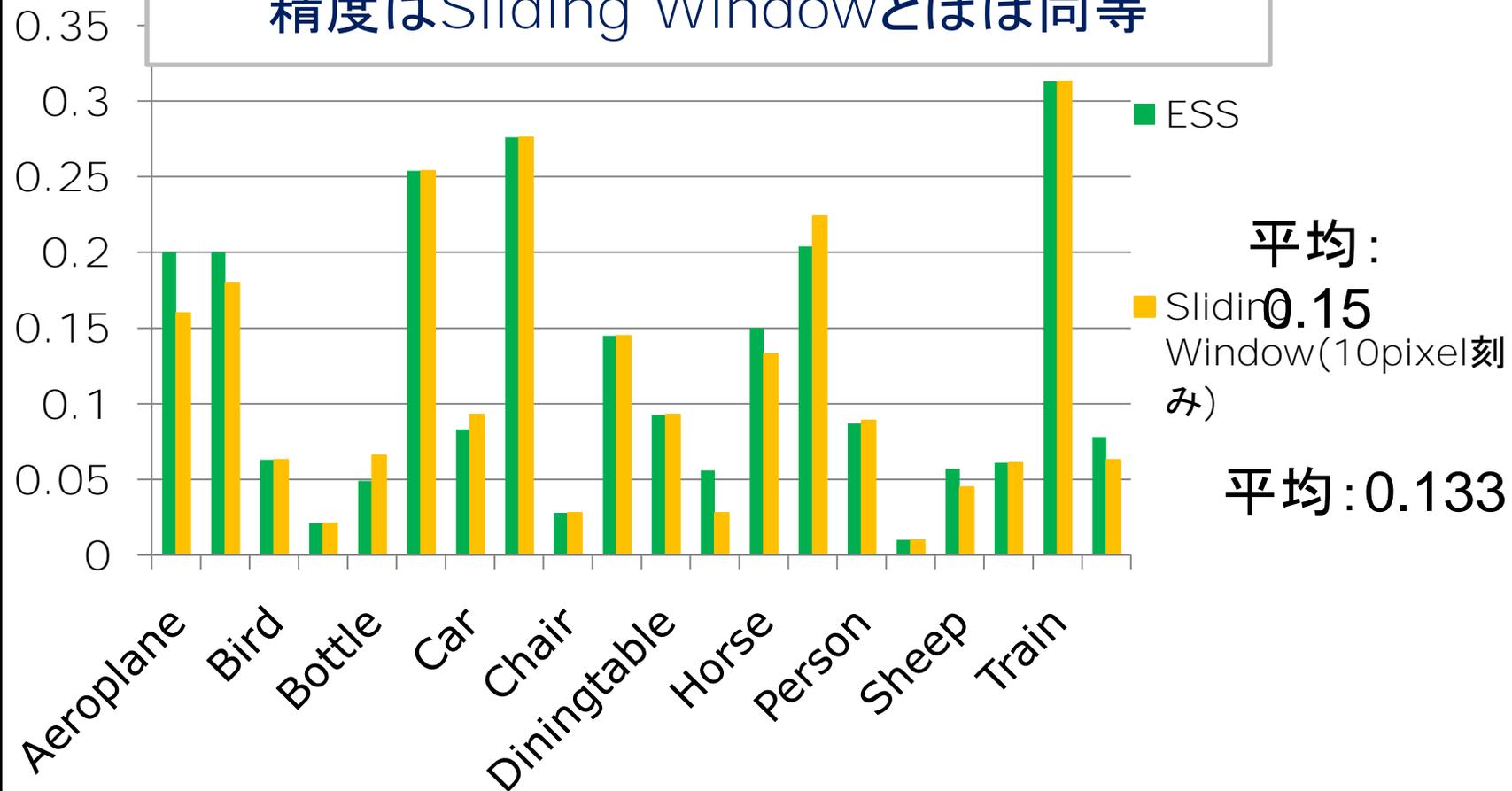




実験結果

■ 各クラスの再現率(%)

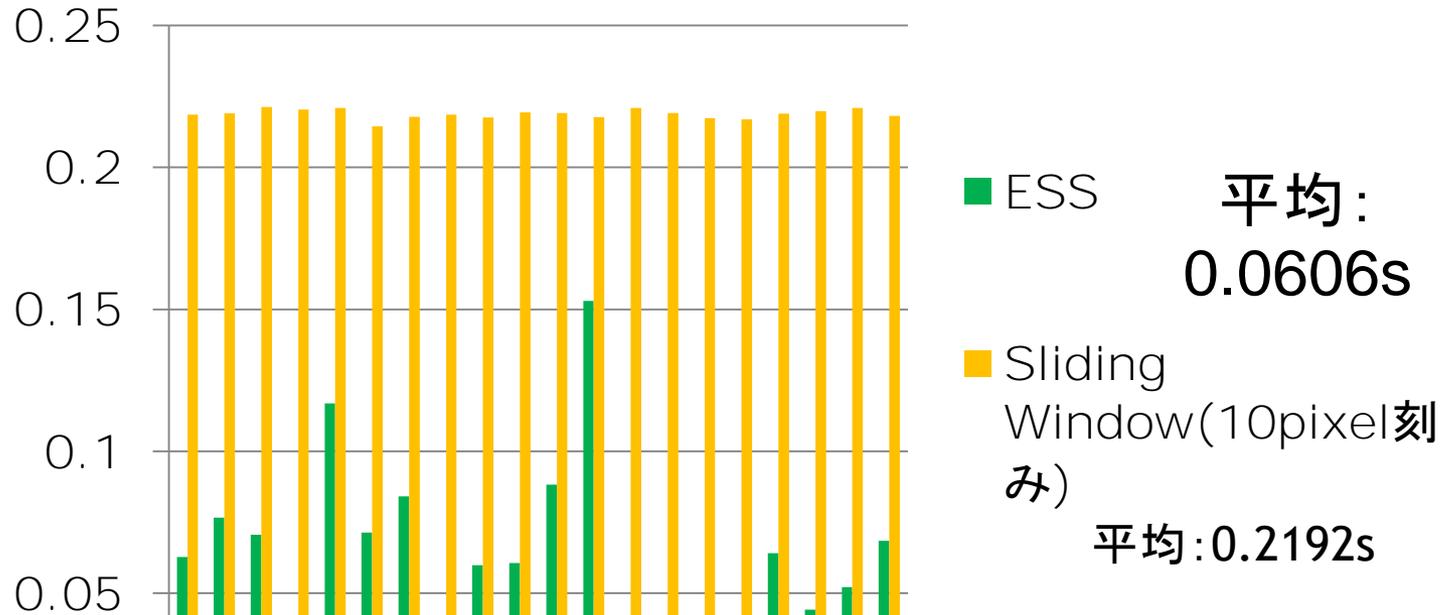
精度はSliding Windowとほぼ同等





実験結果

■ 平均時間(s)



10pixel刻みのSliding Window より高速

Aer

Dinili



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]**

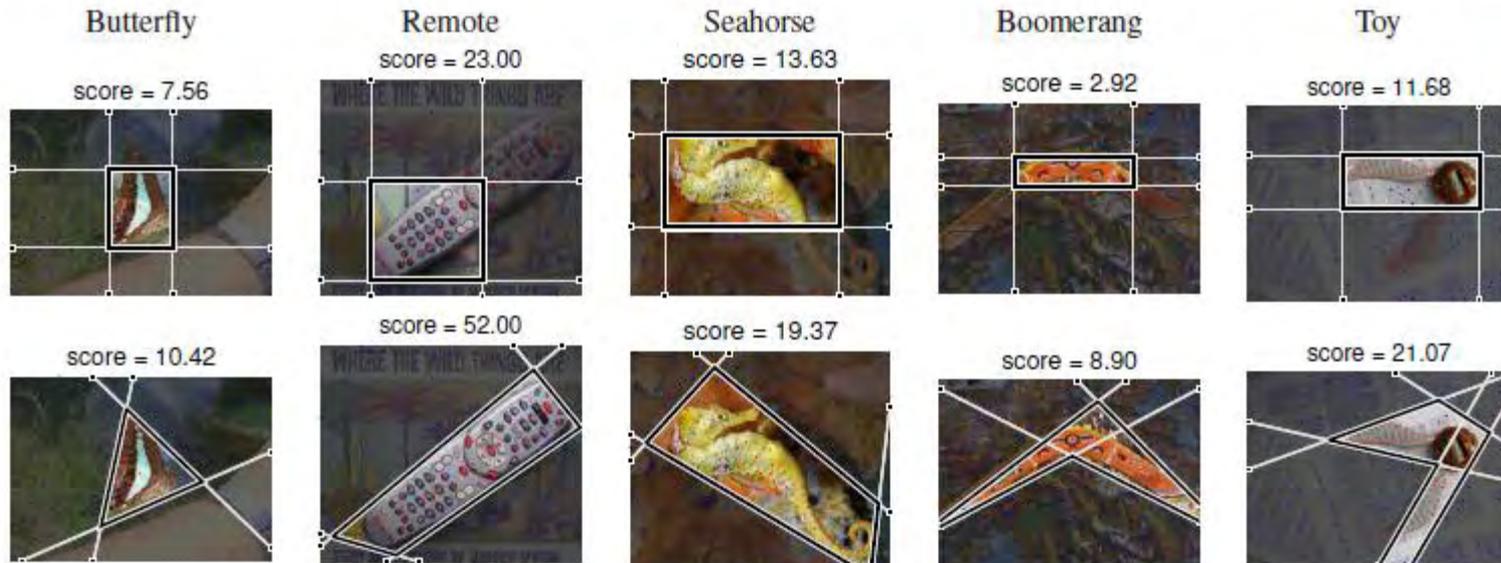


Figure 8. Bounding boxes (top) versus bounding polygons (bottom) (Section 3.2).

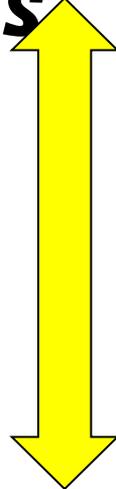


[Vedaldi et al. ICCV09]

Multiple Kernel Learningによる18種類 の特徴の統合

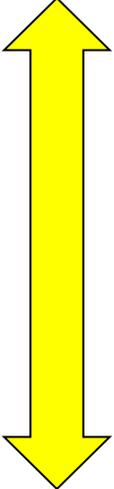
- 候補位置: discriminative visual wordsを含むbounding boxを多数生成.
- [1st step] MKL + 線形カーネルで、物体候補探索(2000程度)
- [2nd step] MKL + 準線形カーネル
 $K(x, y) = \frac{1}{2} (1 - \chi^2(x, y))$ の利用
 - 候補を100個程度に絞る
- [3rd step] MKL + カイ2乗RBF

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) + CRF

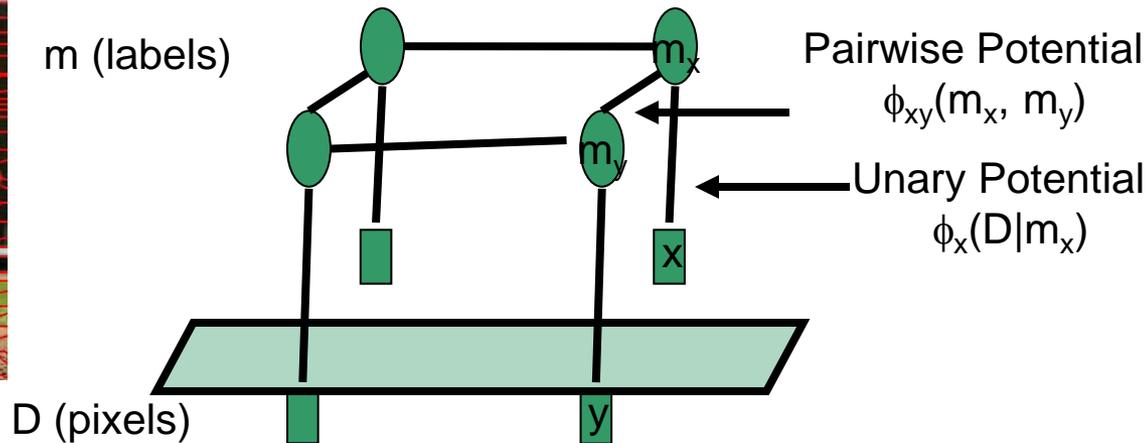
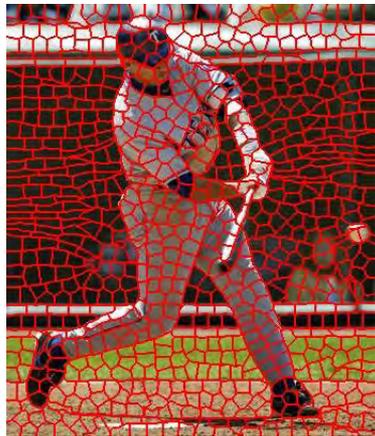
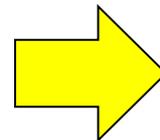


Image Plane

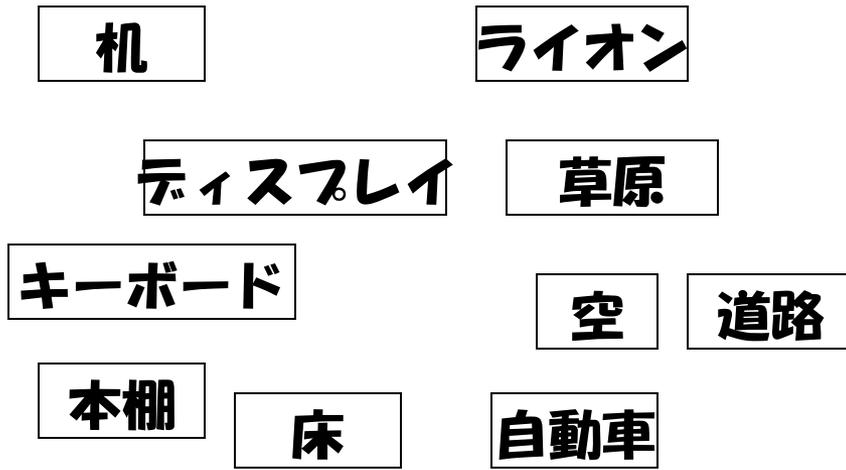




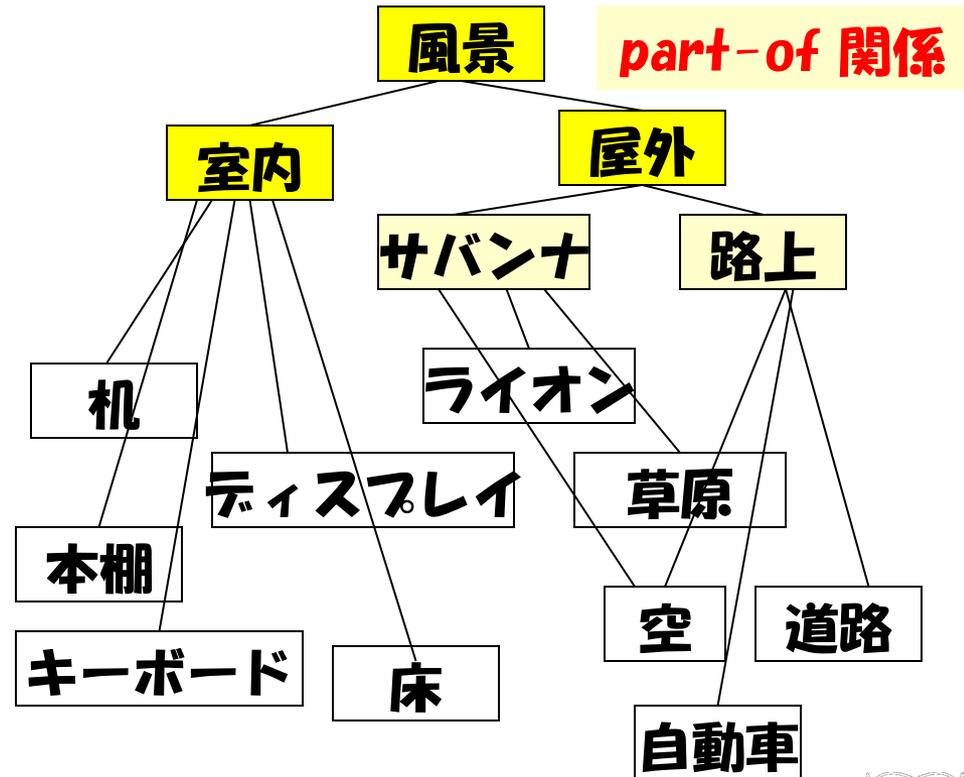
コンテキストの利用:

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

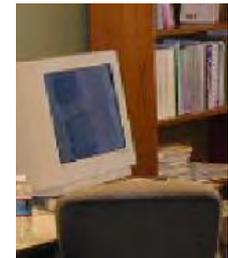
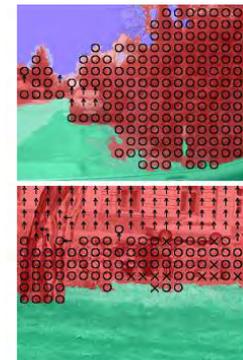
■ (重力に対する)支持関係 PCは机の上にある

■ 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

Image



Support



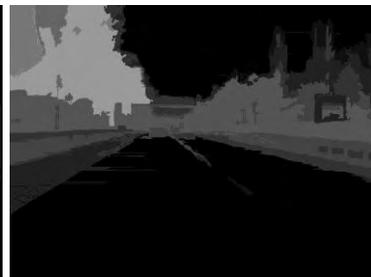
Vertical



Sky



V-Left



V-Center



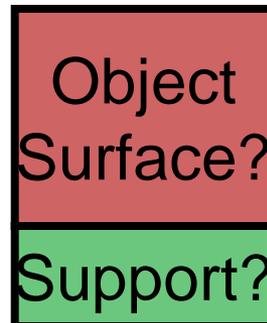
V-Right



V-Porous



V-Solid

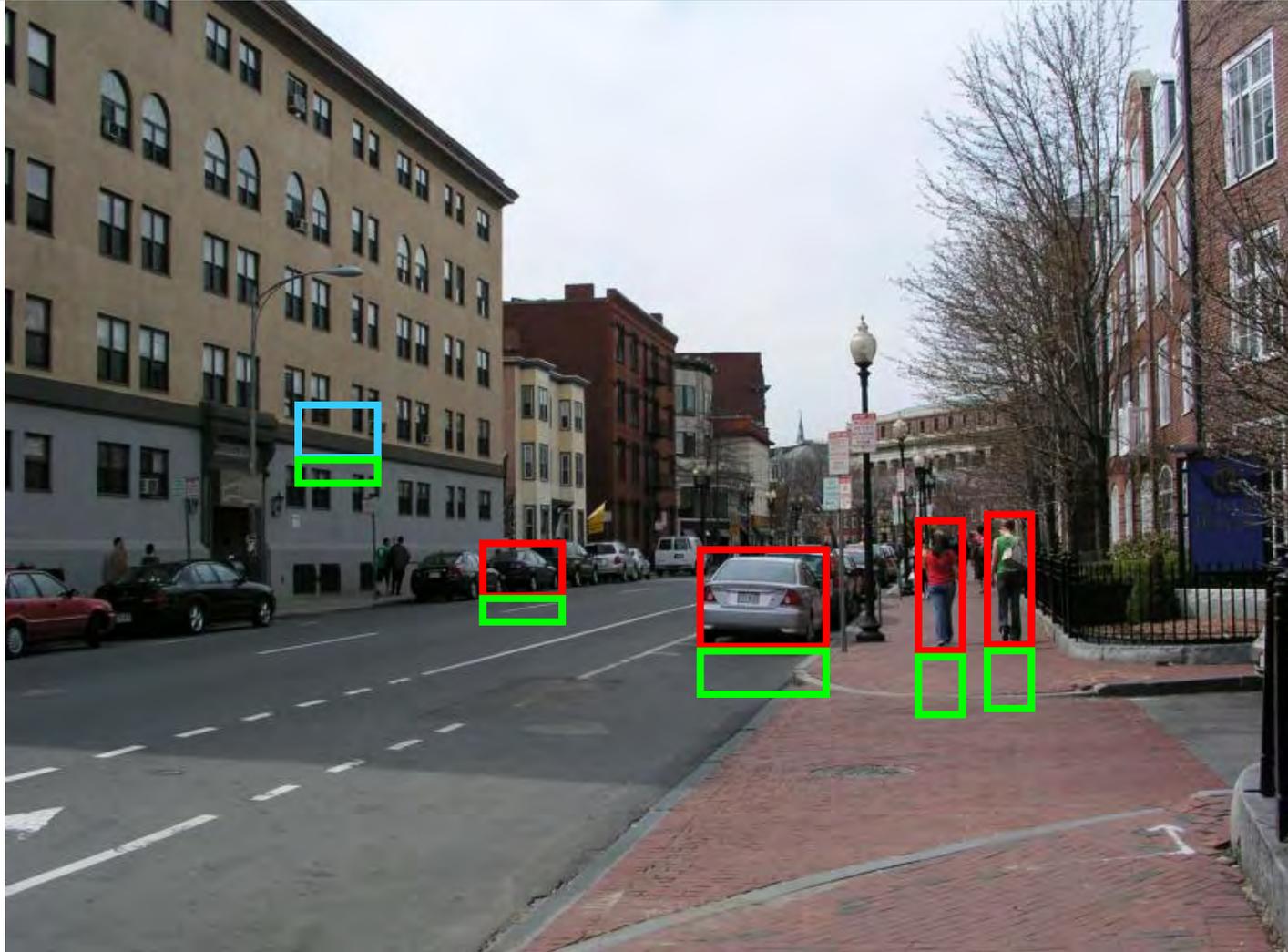


[Hoiem, Efros, Hebert ICCV 2005]





Object Support





言語の階層的分類(タクソノミー)との関係

物体

member-of 関係
instance-of 関係

動物

植物

無生物

脊椎動物

自然物

人工物

ほ乳類

鳥類

どのレベルで認識するか？

バク

イノシシ

ライチョウ

カメラ



5. データセット

【参考文献】

[Tor08] Antonio Torralba, Rob Fergus and William T. Freeman: 80 million tiny images: a large dataset for non-parametric object and scene recognition, IEEE trans. PAMI, pp.1958-1970, (2008).

[Den09] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, Li Fei-Fei : ImageNet: A Large-Scale Hierarchical Image Database, CVPR, (2009).

World Wide Webの

情報爆発!

■ あらゆる情報が存在。

- テキスト情報に加えて、**画像**、**動画**の宝庫。
- Web2.0 の浸透によって、誰でも手軽にWeb上に画像、動画をアップすることが可能になった
 - ・ 携帯からも、スマートフォンからも、デジカメからもアップ可能






写真動画共有サイト



不特定多数の人々によって撮影された画像・映像がWeb上には、大量に存在する。 → **新しい価値**



2003年

 Searched images for **lion animal**. (BETA)

 Results **1 - 20** of about **853**. Search took **0.05** seconds.

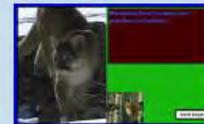

animal-lion.jpg
748 x 519 pixels - 88k
[members.tripodasia.com.tw/
Kasha/](http://members.tripodasia.com.tw/Kasha/)



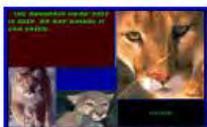
lion.gif
164 x 158 pixels - 26k
www.akronzoo.com/animal.asp



lion watercolor.GIF
640 x 480 pixels - 254k
[biology.ecsu.ctstateu.edu/
HighSchool/animal.htm](http://biology.ecsu.ctstateu.edu/HighSchool/animal.htm)



mtnlion1.gif
518 x 348 pixels - 45k
[www.beavton.k12.or.us/~leahy/
00-01/animals/mtnlion.htm](http://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](http://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/
sliders/online/lion.html](http://www.billybear4kids.com/~sliders/sliders/online/lion.html)



lion-girl.jpg
400 x 400 pixels - 48k
[www.billybear4kids.com/~sliders/
online/lion-girl.html](http://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.ragsdalefinals.com/
animalheads.htm](http://www.ragsdalefinals.com/animalheads.htm)



lion.jpg
288 x 229 pixels - 29k
[www.uiuc.edu/unit/ATAM/
conservation/iran.html](http://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](http://www.kathyskreations.com/~stuffed%20animals%20cougar.html)



lion.jpg
150 x 140 pixels - 7k
gateway.library.uiuc.edu/



AliGloves.jpg
200 x 200 pixels - 7k
auction.causelink.com/



image16.gif
468 x 648 pixels - 37k
cyberfair.gsn.org/cybergang/



lion.GIF
145 x 216 pixels - 24k
www.oxygengroup.com/animals2/

Google 画像検索

lion animal

画像検索

画像検索オプション
表示設定

セーフサーチ: 虫 ▼

2009年

検索ツールを表示

検索結果 約 8,320,000 件中 1 - 20 件目 (0.30 秒)



Lion - Animal Backgrounds
800x600 - 89k - jpg
raidmyspace.com



Pictures of African Animals
448x299 - 30k - jpg
pictures-of-african-animals.info



If you were an animal,
470x324 - 31k - jpg
sodahead.com



lion animal headcover
312x432 - 14k - jpg
shop-safely.com



9 million members making a
792x768 - 99k - jpg
care2.com



Lion, Lion Animal
600x450 - 26k - jpg
dibosdownload.blogspot.com



tags: cat lion animal
599x642 - 101k - jpg
pixdaus.com



lion animal photo
800x600 - 71k - jpg
graphicshunt.com



Area ライオン、トラ、ヒョウ
484x530 - 47k - jpg
j-area2.com



Lion
625x450 - 105k - jpg
animal.discovery.com



lion animal headcover
348x432 - 15k - jpg
shop-safely.com



Cub Lion Picture
360x299 - 41k - jpg
lion-cats.com



Clipart - lion, animal
300x208 - 18k - jpg
fotosearch.com



African Animal Image Gallery
400x300 - 26k - jpg
animals.howstuffworks.com



アフリカライオン
470x324 - 30k - jpg
nationalgeographic.co.jp



Wild Animal Park Lion and
450x338 - 56k - jpg
randomsandiego.com



Male Lion - Animal - Wild
468x351 - 45k
article.wn.com



It's a lion riding
468x366 - 26k - jpg
thermocaster.com



The lion... Animal Planet had
246x330 - 134k - jpg
sheknows.com



Lion on Horseback
468x393 - 39k - jpg
morecoolpictures.blogspot..



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!

100515Chokai0097

Share This

ADD NOTE SEND TO GROUP ADD TO SET BLOG THIS ALL SIZES ORDER PRINTS ROTATE EDIT PHOTO DELETE



At the top of Mt. Chokai just before skiing down

Add your comment

How do I format my comment?

PREVIEW

OR

POST COMMENT

Uploaded on May 17, 2010 by [keiji.y](#)

keiji.y's photostream

761 uploads

[browse](#)

People in this photo

[Add a person](#)

タグ

Tags

- Mountain ×
- snow ×
- backcountry ×
- ski ×
- cloud ×
- sky ×

[Add a tag](#)

Additional Information

- All rights reserved ([edit](#))
- Anyone can see this photo ([edit](#))
- [Add to your map](#)
- Taken with a Sony DSC-HX5V. [More properties](#)
- Taken on May 15, 2010 ([edit](#))
- [Photo stats](#)
- Viewed 26 times (Not including you)
- [Edit](#) title, description, and tags.
- [Replace](#) this photo.

[Flag your photo](#)

Web2.0 サイト例

Flickr 写真共有 サイト



例2: Panoramio 位置情報画像共有サイト

Panoramio

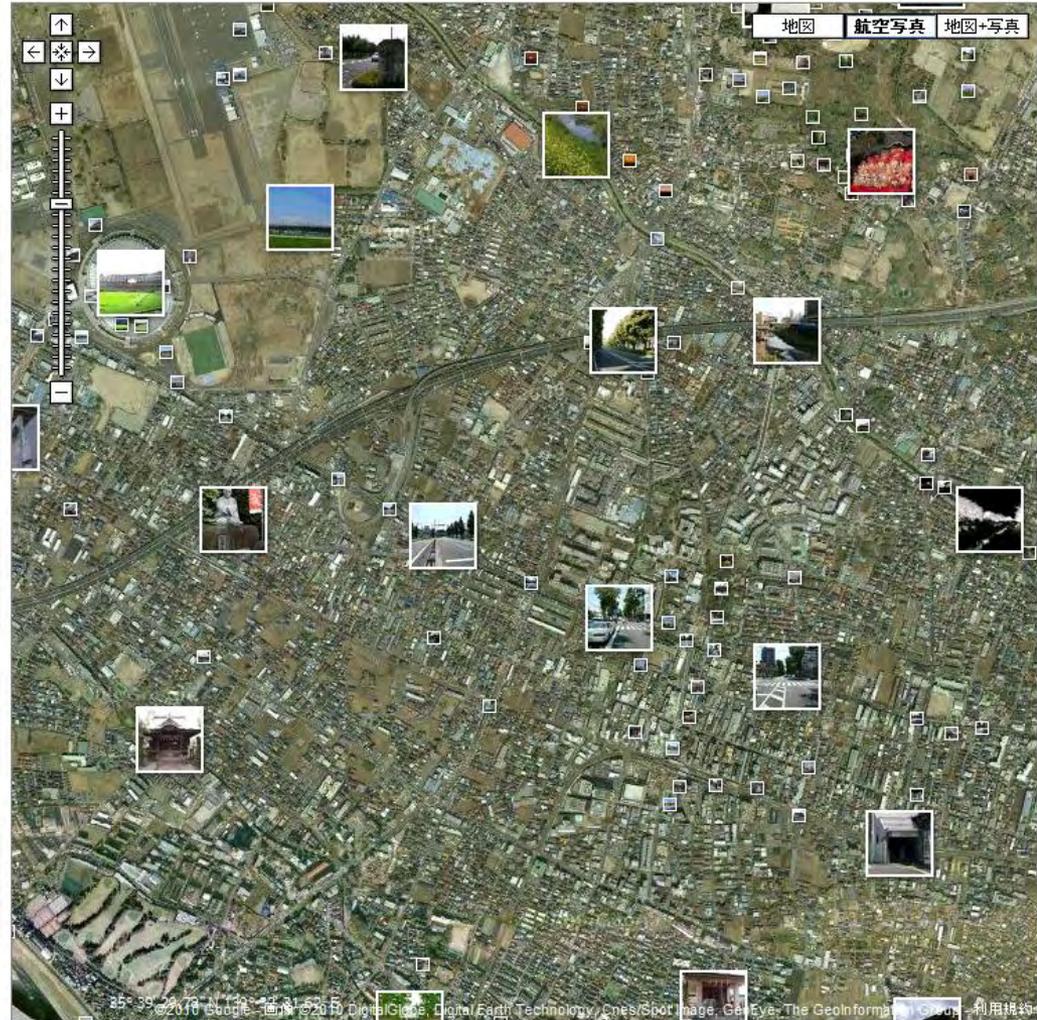
サインアップ | アップロード | 世界旅行 | タグ

[フォーラム](#) [ブログ](#) [サインイン](#) [ヘルプ](#)

検索

人気 Recent

Also show photos not selected for Google Earth



例3: Youtube



morenci arizona

Search

Browse

Upload

tucson2ochikawa ▾

Si

Edit Video

Edit Video

Annotations

AudioSwap

Captions and Subtitles

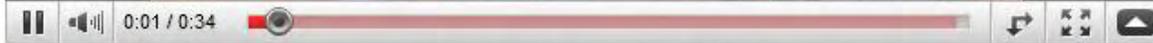
Insight

Mine in Morenci, Arizona

tucson2ochikawa

2 videos ▾

Subscribe



tucson2ochikawa — November 29, 2008 — This movie was recorded near Morenci, AZ on the Clifton-Alipne Highway.

556 views



Save to ▾



テキスト

<Embed>



Respond to this video...



Mine in Morenci, Arizona
5,293 views
exlamb



Morenci Mine
2,471 views
askurka



Morenci Thanksgiving Trip Video 2 (tunnels)
307 views
dbacksterry



Morenci Thanksgiving Trip Video 4 (sheep #2)
195 views
dbacksterry



Copper ore coming off conveyor
3,132 views
EdwardEscobar



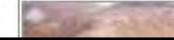
Wild Animals in Morenci
2,607 views
metalmulishafuck



Family
548 views
AnnaQuintino22



Morenci Thanksgiving Trip Video 7 (driving the ...)
112 views
dbacksterry



Morenci Thanksgiving Trip Video 5 (Mine Overlook)





Webデータ収集の容易化

- 2002年4月 Google Web API が公開
 - Googleの検索DBが, 自作のシステムから無料で誰でも利用可能. 画期的!!
 - それまでは, HTTP GET を 80ポートに送信. 戻ってきたHTMLから結果抽出(wrapper).

■ 現在は,

Google
画像検索

YAHOO!

bing
Beta

You Tube

ニコニコ動画(ββ)
NICO NICO DOUGA

flickr
LOVES YOU

Picasa™

amazon.com

価格.com
「買ってよかった」をすべてのひとに。

楽R天
ICHIBA

Panoramio

ほとんどの検索エンジン, 共有サイト, ECサイト
がWebAPIを無償公開している. →

データ収集が容易



Web画像・動画像の特徴

- **無料！** 研究に利用しやすい。
- **多種多様な画像が、大量にある**（数百億？）
- **一般の人が撮った写真が多い**(Consumer-generated) 特に、写真・動画共有サイト
 - 「**一般**物体認識」研究のためのデータに向いている
- **メタデータが付いている。**
 - **テキスト情報**(周辺テキスト, タグ): **ノイズあり**
 - ・ **テキスト情報を利用して、特定種類の画像を収集可能**
 - **ジオタグ**(位置情報), 撮影日時, 使用カメラ, 焦点距離

こうした特性を考慮して利用すべき

WebAPIによる収集例： ラーメン





WebAPIによる収集例：寿司





8千万枚のWeb画像による認識

[Torralba et al. PAMI 2008]

- **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($32 \times 32 \times 3$), 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

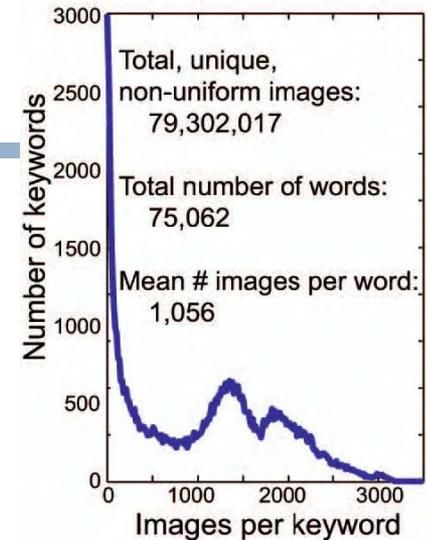
- Altavista, Ask, Flickr, Cydral, Google, Picsearch, Webshots

■ 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



(a)

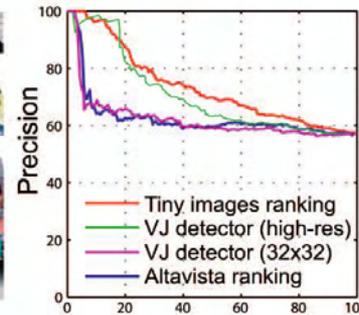
「人」の認識



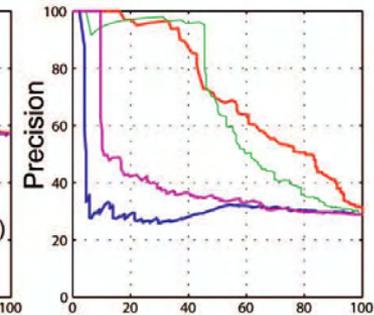
(a)



(b)



(c)



(d)

- **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)



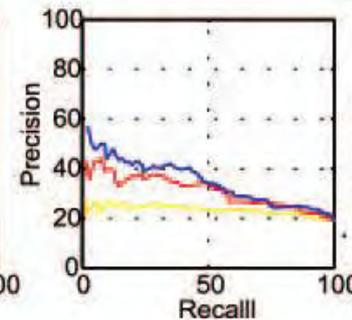
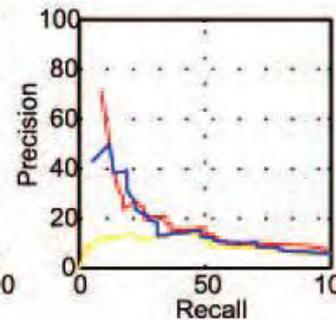
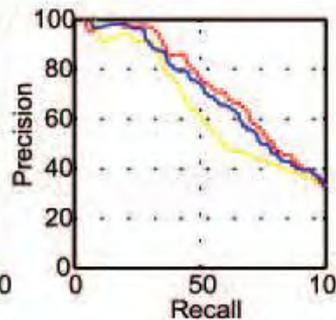
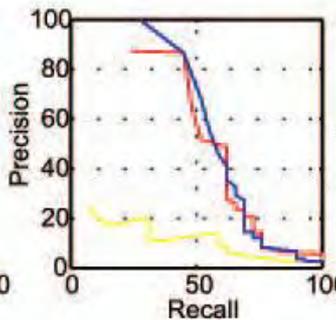
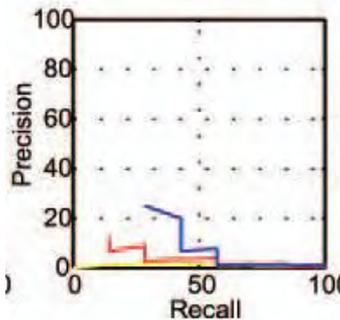
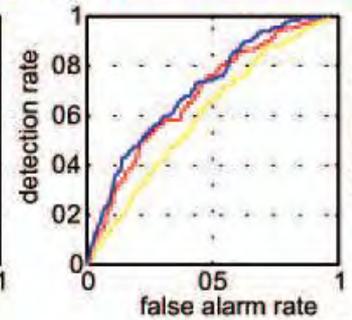
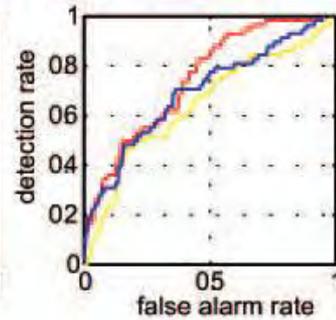
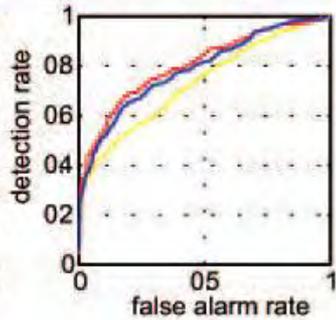
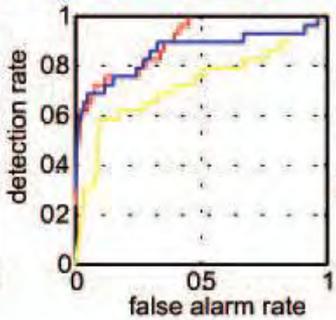
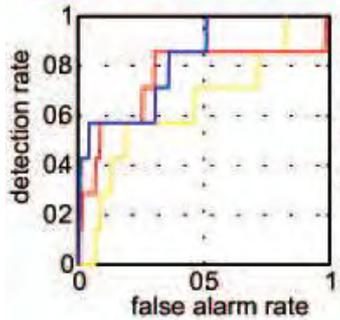
Plant life
(335)



Flower
(58)



Artifact
(187)





8千万枚のWeb画像による認識

- 画像が多いクラスは特定クラス検出器に匹敵
- 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
 - ・ 多くの研究ではモデルの研究
 - ・ この研究はデータ側からアプローチしている。

Xin-Jing Wang, Lei Zhang, Ming Liu, Yi Li, Wei-Ying Ma:
ARISTA - Image Search to Annotation on Billions of Web Photos,
CVPR, (2010). では, 20億枚で同様の実験を実施. さらによい結果が得られた!

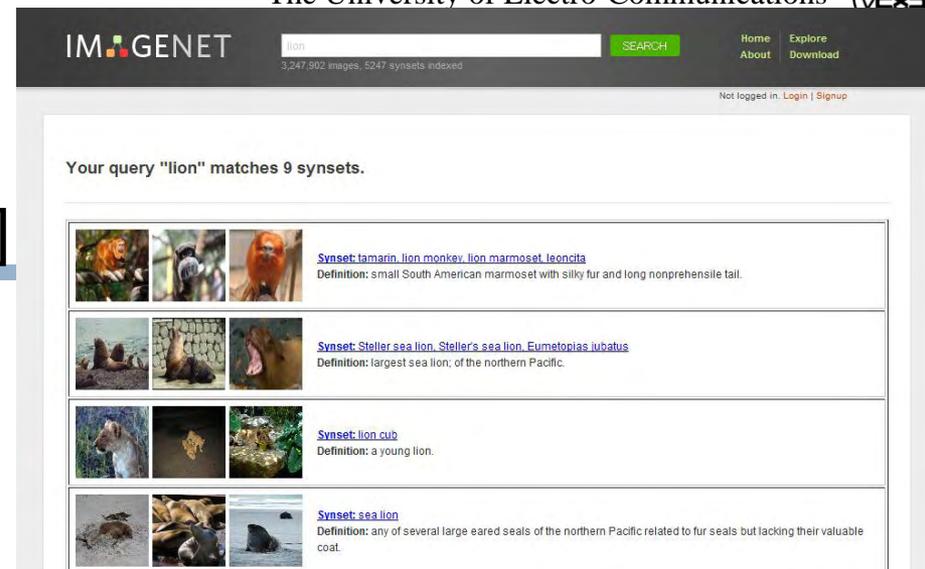




ImageNet

[Deng et al. CVPR2009]

- 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15,589語について1,123万枚収集
- Amazon Mechanical Turk を利用して作成。
(不特定多数の人の知識の利用。Crowd-sourcing)
 - 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



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



Amazon Mechanical Turk

- 大量の“人力”の利用 : crowd-sourcing
- 学習画像データ作成, 認識結果評価に利用可能.

The screenshot shows the top navigation bar of the Amazon Mechanical Turk website. It includes the logo, navigation links for 'Your Account', 'HITS', and 'Qualifications', and a sign-in prompt for workers and requesters. Below the navigation bar is a yellow banner with the text: 'Mechanical Turk is a marketplace for work. We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient. 104,620 HITs available. View them now.'

Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work



or [learn more about being a Worker](#)

Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Register Now](#)

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results



or [learn more about being a Requester](#)



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80,000 synset (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

■ 構築状況

- 5247 synsets (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



Large Scale Visual Recognition Challenge 2010



- ImageNet のデータを使った,
1000種類画像認識Challenge (1000クラス分類)
 - PASCAL VOC (visual object challenge)の一部
 - 1000クラス120万枚のラベル付き学習データ
 - 1000クラス15万枚のラベルなしテストデータ
- 1000クラスマルチクラス分類による
分類率などで結果を評価.
- **8月23日締め切り.**

6. 今後の方向



今後の方向 (1)

- **80million, 2billion 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. 同一(near-duplication)インスタンス検索によるクラス分類.**
 - **でも, MS / Google以外には手に負えない可能性. .**
 - ・ **80million収集に1年掛かるとすると. . .**
 - ・ **2billion⇒25年. . . もはや大学などでは不可能.**



2003年

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.ragsdalefinals.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

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

2005年



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...





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

2009年

検索結果 約 7,950,000 件中 1 - 20 件目 (0.04 秒)

危険な動物 [dangered animals](#) スポンサーリンク
amazon.co.jp 日本全国送料無料で(1500円以上) コンビニ受取で好きな時受け取り可能



Lion -
800x600 - 89k - jpg
raidmyspace.com
[類似の画像を探す](#)



If you were an
470x324 - 31k - jpg
sodahead.com
[類似の画像を探す](#)



Lion Bait
792x768 - 99k - jpg
whohastimeforthis...
[類似の画像を探す](#)



Animal picture -
601x480 - 204k
animalpicturegallery.net
[類似の画像を探す](#)



Pictures of African
448x299 - 30k - jpg
pictures-of-african-anim...
[類似の画像を探す](#)



lion animal
800x600 - 14k - jpg
shop-safely.com



lion animal photo
800x600 - 71k - jpg
graphicshunt.com



African Animal
400x300 - 26k - jpg
animals.howstuffworks.com



Lion
625x450 - 105k - jpg
animal.discovery.com
[類似の画像を探す](#)



Area ライオン、トラ、ヒョウ
484x530 - 47k - jpg
area2.sakura.ne.jp
[類似の画像を探す](#)



Wild Animal Park
450x338 - 56k - jpg
randomsandiego.com
[類似の画像を探す](#)



australian animals
417x297 - 31k - jpg
wildanimalsplanet.com



Starring horse-riding
468x366 - 26k - jpg
letstalkug.net
[類似の画像を探す](#)



Wild animals, as
479x450 - 62k - jpg
adamthinks.com
[類似の画像を探す](#)



Clipart - lion,
300x208 - 18k - jpg
fotosearch.com



Cub Lion Picture
360x299 - 41k - jpg



Lion Face
600x465 - 75k - jpg



Nature Animal Lion
350x460 - 210k - jpg



A LION ANIMAL
350x504 - 28k - jpg



Lion on Horseback
468x393 - 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 ▾

検索結果 約 556 件中 1 - 20 件目 (0.04 秒)

2009年 類似画像検索



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](#)



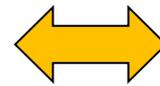
781×1000 - 143k
[brianhamptonphotogra...](#)
[類似の画像を探す](#)



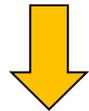
201?年

Google

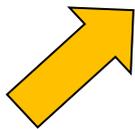
画像認識



80 billion image DB

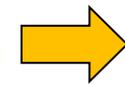


NN search (特定物体サーチ)



 448*299 - 30k aucklandzoo.co.nz 類似の画像を探す	 600*491 - 67k - imgcache silva.com 類似の画像を探す	 210*217 - 43k - jpg tuffdog.com 類似の画像を探す	 500*332 - 47k - jpg worldofstock.com 類似の画像を探す	 300*275 - 56k eln7abn.com 類似の画像を探す
 840*563 - 114k - jpg abc.net.au 類似の画像を探す	 600*600 - 154k - jpg pt.treknature.com 類似の画像を探す	 600*594 - 156k commons.wikimedia.org 類似の画像を探す	 500*375 - 50k - jpg mlazm.net 類似の画像を探す	 518*333 - 63k blog.uncovering.org 類似の画像を探す
 1600*1200 - 298k - jpg 217-219-193-18 類似の画像を探す	 384*268 - 42k plaza.rakuten.co.jp 類似の画像を探す	 400*257 - 107k politics.sforums.com 類似の画像を探す	 613*439 - 90k - jpg pixdaus.com 類似の画像を探す	 620*615 - 68k retiredaussies.com 類似の画像を探す

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)

- **重要になる機械学習・データマイニング技術**
 - **大量データに対する高速類似データ検索**
 - ・ Near-duplication detection
 - ・ 高速クラスタリング
 - **学習データが少ない場合(特殊な認識対象, 新しい認識対象)への対処**
 - ・ 半教師あり学習(semi-supervised learning)
 - ・ 転移学習(transfer learning)
 - **機械学習による認識対象毎の最適な特徴抽出.**
 - **K-NNにおけるメトリックの工夫**
 - ・ Metric learning, 距離関数の学習



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？

A dense grid of approximately 80 small images, each showing a different bowl of ramen. The bowls contain various types of noodles, soups, and toppings like meat, vegetables, and eggs. In the center of the grid is a large, rounded yellow rectangle containing the Japanese text 'おわり' (The End) in blue characters.

おわり