

e-サイエンスに向けた革新的アルゴリズム基盤

# アルゴリズムの応用例： 物体の認識と画像検索

徳山 豪（東北大学）

# 近年考えること: 科学としてのアルゴリズム

科学の凄さ: 原理の追及が驚くべき説明を生む

人間生活を大きく変革する長期的影響(化学、医学、物理など)

工学と違い、「役に立つ」かどうかは後世が決める

アルゴリズムの凄さはなんだろう

- 現代の情報社会を支える「凄く『役に立つ』やつ」
  - ネット、携帯、カーナビ、暗号通信
  - ロボティクス、医療、天気予報、グラフィクス
- 人間を超えるインテリジェンスの創出(素人目に『凄い』)
  - インターネット検索、ゲノム解析 → 情報システムの原理
  - チェス、将棋、ワトソンシステム → 知能の原理
- 「凄いもの」を生まない科学は認知されない
  - NP完全性が『凄い』と『役に立つ』を分けるのなら、我々はどうする？
  - 『アルゴリズムの科学としての認知』に成功しているか？

# アルゴリズム研究の2つのアプローチ

従来手法では解決不可能な大規模かつ複雑な諸問題を数理解析に基づく**革新的なアルゴリズム**によって解決する

## 1: インパクトのある重要問題の解決要求からの出発

- 「いかにも難しい」問題の数理的な定式化自体がメイン
  - 計算機で扱えること自体がサプライズになる
  - 実用的に動くアルゴリズム（時には理論保証なしでも）

## 2: 計算理論からの出発

- 理論的アルゴリズムで解ける範囲をとことん追求
  - 「解けそうもない」ところを「解きほぐす」研究
  - 「解けない」限界を立証する研究
  - 一般人に『凄さ』を感じさせる（なかなか難しい）

# 実用問題の解決要求から出発する例

## 画像検索のための自動アノテーション

**Cam-Tu Nguyen, Natsuda Kaothanthong,  
Xuan Hieu Phan, Takeshi Tokuyama:**  
**A feature-word-topic model for image annotation.**  
CIKM 2010 (20<sup>th</sup> ACM Conf. on Information and  
Knowledge Management): 1481-1484



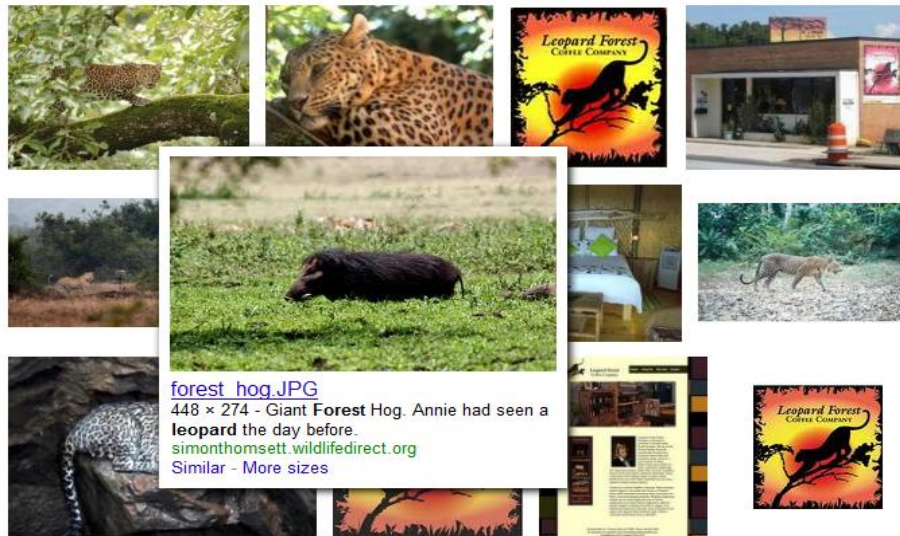
# 画像検索 (Image retrieval)

数億の人が毎日行っている日常的なIT応用

- ユーザは視覚的な概念を検索語にする
- システムは検索語を受け取って画像を検索する
- 概念の抽象化と汎化 → 意味的ギャップの解消が難しい

現状は、画像に(主に人手により)添付された文書を用いて検索する

Search for  
“leopard  
forest” with  
Google  
Image



# 画像注釈問題

画像に対して、自動的に注釈(キーワード)を与える。

- 人手を介せず、画像を解説する説明を作成
- 画像検索の全自動化に向けた技術
  - 本質的に学習なしでは不可能

Image  
Annotation to  
map an image to  
textual space.



Coral, ocean, people



Building, market,  
skyline, street



Beach, palm, people,  
tree



Costume, people,  
street, village

# 画像注釈問題と画像認識の違い

画像認識: あるドメインの単独物体を識別するシステム

画像注釈: 混然とした画像を特徴づける複数の語句を与える

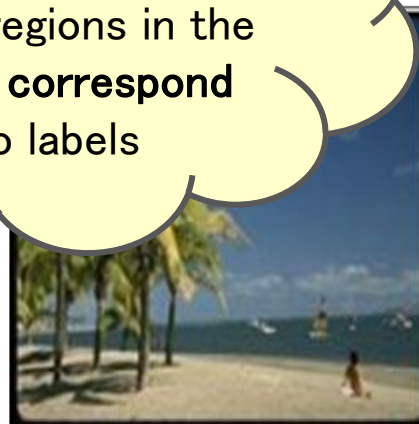
- 画像のどこにオブジェクトがあるかは指示されない
- 学習データも注釈付きの画像(複数の概念の混合)



Coral, ocean, people



Building, market,  
skyline, street



Beach, palm, people,  
tree



Costume, people,  
street, village

No indication of  
which regions in the  
image **correspond**  
to labels

# 人間の画像理解



# 人間の脳の認識と情報の汎化

左の画像には何がありましたか？ → 熊です（画像認識）

右の画像には何がありましたか？ → ？？

そんなことは考えない(記憶しない)

複雑な画像で、人間がとらえるのは、画像のトピック（物語）である

右の画像は海中の画像で、きれいな画像でも、探索時のキーワードは、「海、魚、イソギンチャク」のようなもの  
「花」でなく「イソギンチャク」と判断するには？



# Topic Model : テキストマイニングでは

|            | Topics |   |   |
|------------|--------|---|---|
|            | 1      | 2 | 3 |
| BANK       |        |   |   |
| COMMERCIAL |        |   |   |
| CRUDE      |        |   |   |
| DEEP       |        |   |   |
| DEPOSITS   |        |   |   |
| DRILL      |        |   |   |
| FEDERAL    |        |   |   |
| FIELD      |        |   |   |
| GASOLINE   |        |   |   |
| LOANS      |        |   |   |
| MEADOW     |        |   |   |
| MONEY      |        |   |   |
| OIL        |        |   |   |
| PETROLEUM  |        |   |   |
| RESERVE    |        |   |   |
| RIVER      |        |   |   |
| STREAM     |        |   |   |
| WOODS      |        |   |   |

(a) Probability Table

| TOPIC 1    | TOPIC 2    | TOPIC 3    |
|------------|------------|------------|
| MONEY      | STREAM     | OIL        |
| BANK       | RIVER      | PETROLEUM  |
| FEDERAL    | BANK       | GASOLINE   |
| RESERVE    | DEEP       | CRUDE      |
| LOANS      | WOODS      | COMMERCIAL |
| DEPOSITS   | FIELD      | DEPOSITS   |
| COMMERCIAL | MONEY      | RIVER      |
| DEEP       | MEADOW     | DRILL      |
| MEADOW     | OIL        | MONEY      |
| OIL        | FEDERAL    | DEEP       |
| RIVER      | DEPOSITS   | FIELD      |
| CRUDE      | LOANS      | STREAM     |
| DRILL      | COMMERCIAL | RESERVE    |
| FIELD      | CRUDE      | BANK       |
| GASOLINE   | DRILL      | LOANS      |
| PETROLEUM  | GASOLINE   | FEDERAL    |
| STREAM     | PETROLEUM  | MEADOW     |
| WOODS      | RESERVE    | WOODS      |

Words sorted by word-topic probabilities

(b) Words sorted according to Topics

# 画像注釈でのトピックモデルの構築

やりたいこと: 「画像のストーリー」を示すトピックモデルを作り、高品質の画像注釈を行う

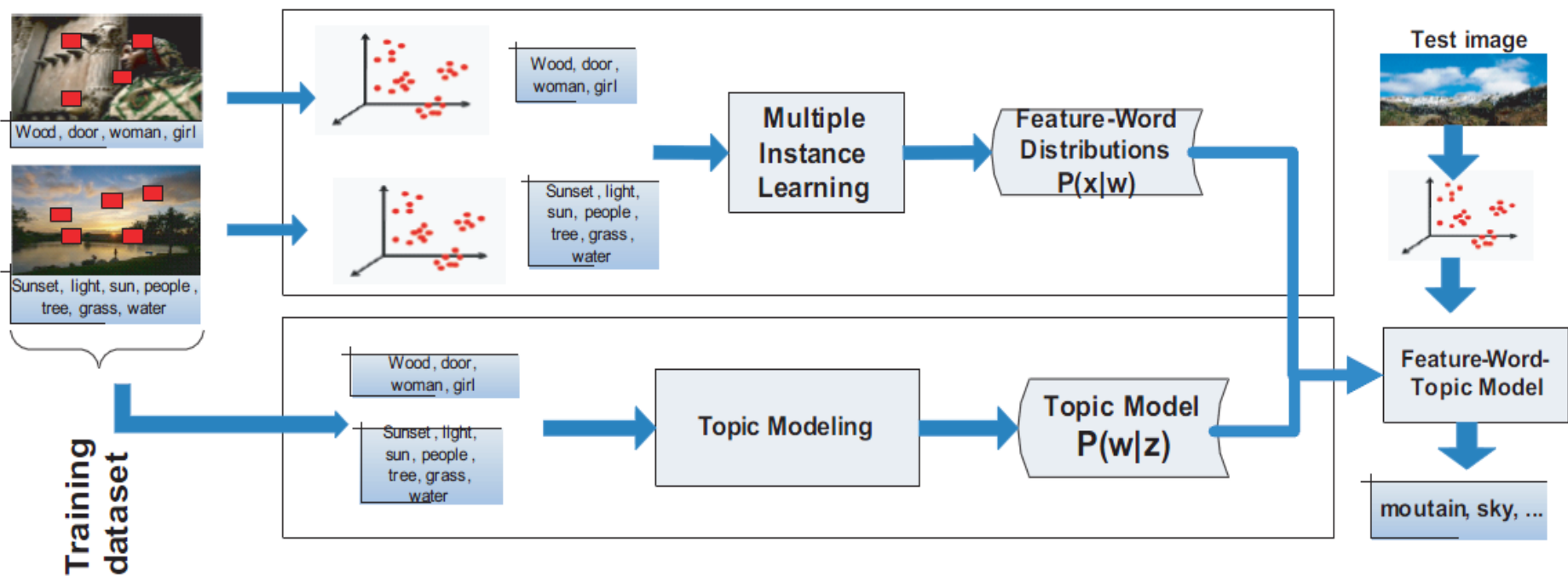
困難性: 画像の空間は文書の空間に比べて曖昧で取り扱いが難しい

- 文書のトピック空間は「語(おもに名詞)」の作る有限次元空間
- 画像の特徴ベクトルの空間の有効な基底は？

解決策:

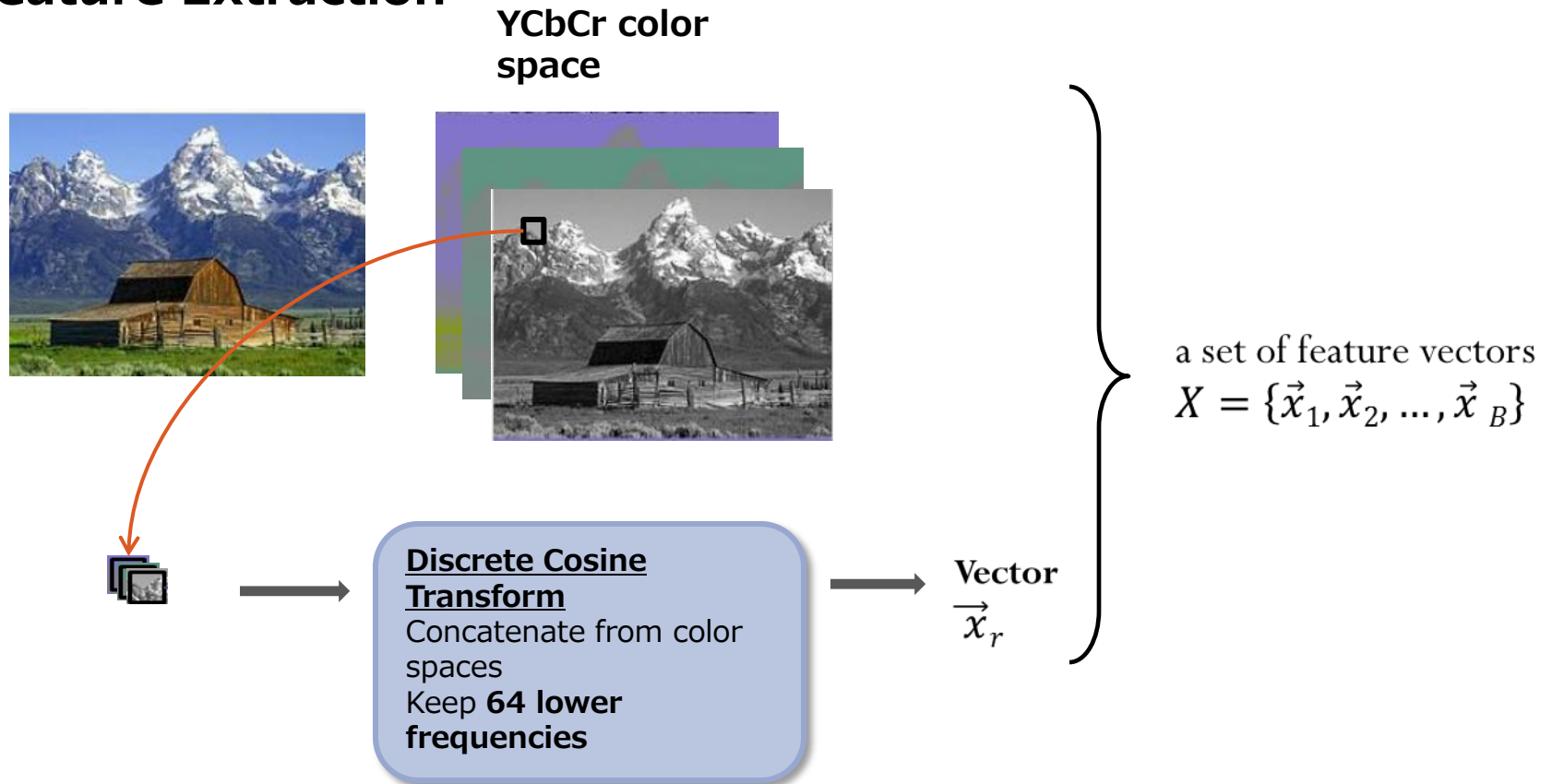
- 特徴ベクトル空間を確率分布の空間としてとらえる(既存研究)
  - オブジェクト＝特徴ベクトルのガウス分布
  - 画像＝ ガウス分布の集合和で近似
- 学習画像に付与されたキーワードの持つトピック空間と、上記の確率分布空間を対応付ける

# Feature-Word-Topic Model



# Estimation of Feature-Word Distribution

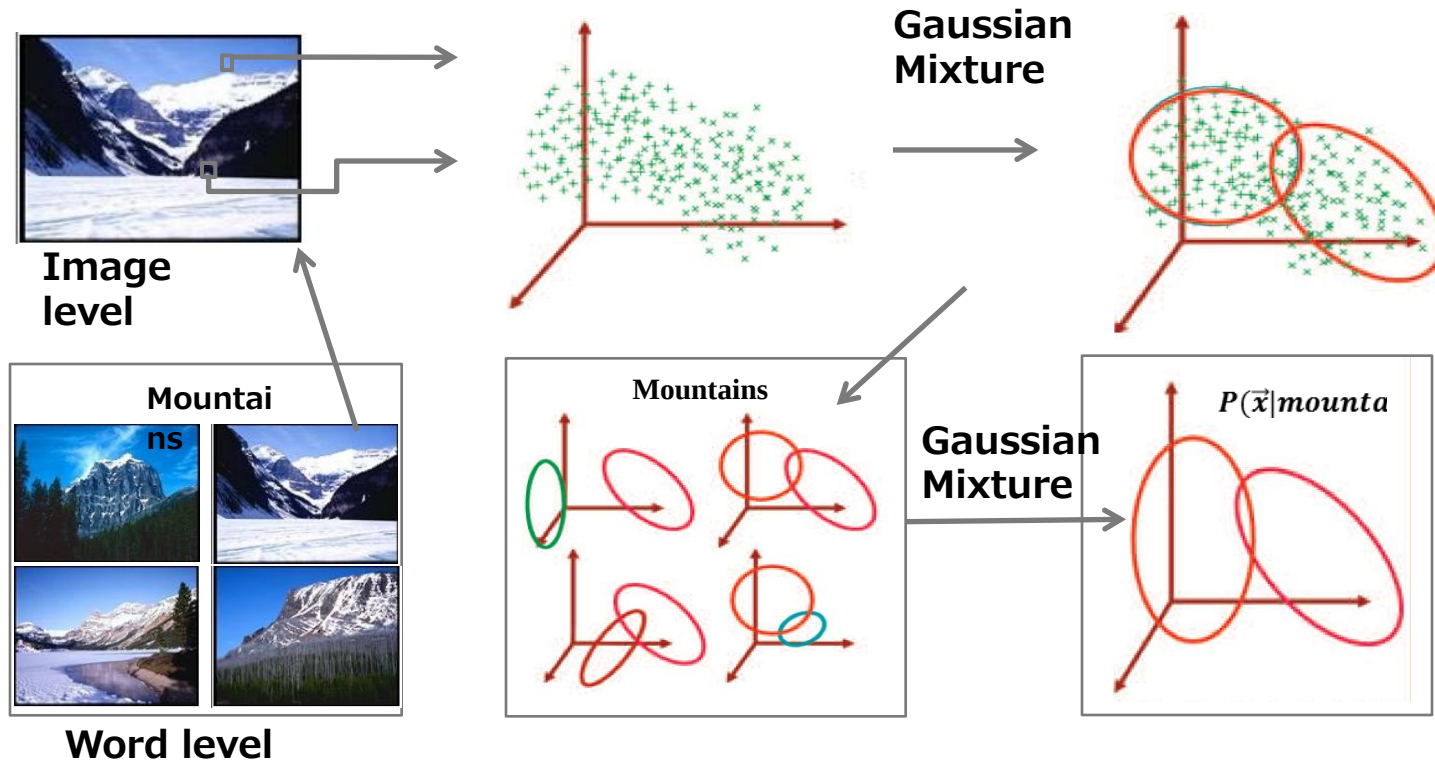
## Feature Extraction



# Estimation of Feature-Word Distribution

## Mixture Hierarchies and Multiple Instance Learning

Carneiro et al., Supervised Learning of Semantic Classes for Image Annotation and Retrieval, IEEE Trans. on PAMI, Vol. 29, No. 3, 2007



# Estimation of Word-Topic Distribution

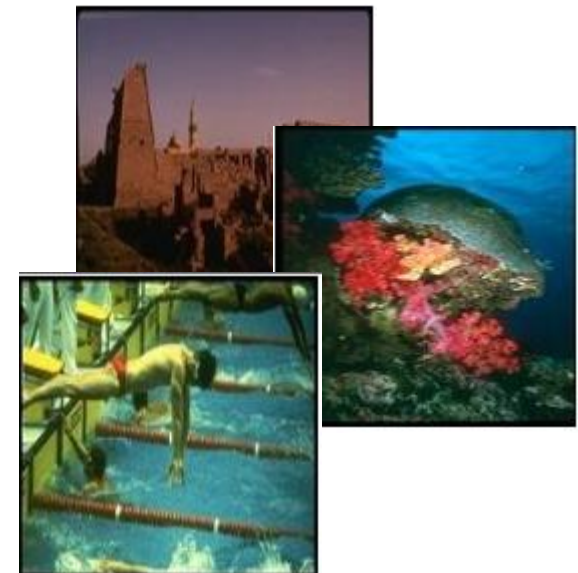
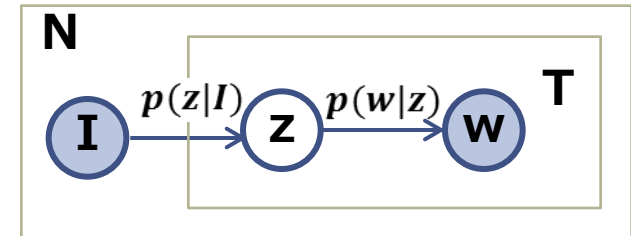
## Probabilistic Latent Semantic Analysis (pLSA)

*Hofmann et al., 1999*

- Consider captions of one image as a short document
- Given  $K$ , an **EM algorithm** is used to estimate  $p(w|z)$  from the caption set of an annotated image collection.

| Topic 165 | Topic 164 | Topic 123 | Topic 198 | Topic 224 |
|-----------|-----------|-----------|-----------|-----------|
| Pool      | Pyramid   | Cat       | Sponges   | Shop      |
| People    | Stone     | Tiger     | Coral     | Street    |
| Canoe     | Tomb      | Forest    | Ocean     | City      |
| Race      | Ruins     | Bengal    | Sea       | People    |
| Swimmer   | Sun       | Sun       | Anemone   | Crafts    |
| Water     | Sand      | Grass     | Basket    | Sign      |

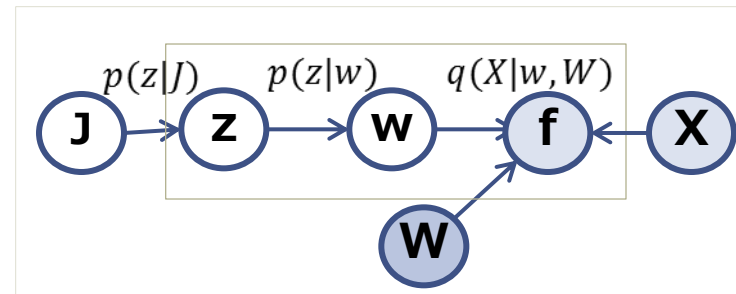
$w$ : labels  
 $z$ : topic assignment



# Feature-Word-Topic Model

From the two-stage training, we have

- Feature-word distribution  $p(\vec{x}|w)$
- Word-topic distribution  $p(z|w)$



**Topic Inference**

- Candidate words are *weighted* using (some type of scaling)

$$q(X|w, W) \propto p(X|w) - \min\{p(X|w_m)\} \text{ where } w_m \in \mathbf{W}$$

Topic Inference on candidates  $\rightarrow$  *Topic distribution*  $p(z|J)$

**Re-ranking for Annotation**

$$p(w|J, X, W) = \underbrace{q(X|w, W)}_{\text{Depends on feature}} \sum_k \underbrace{p(z_k|J) \times p(w|z_k)}_{\text{Depends on topics}}$$

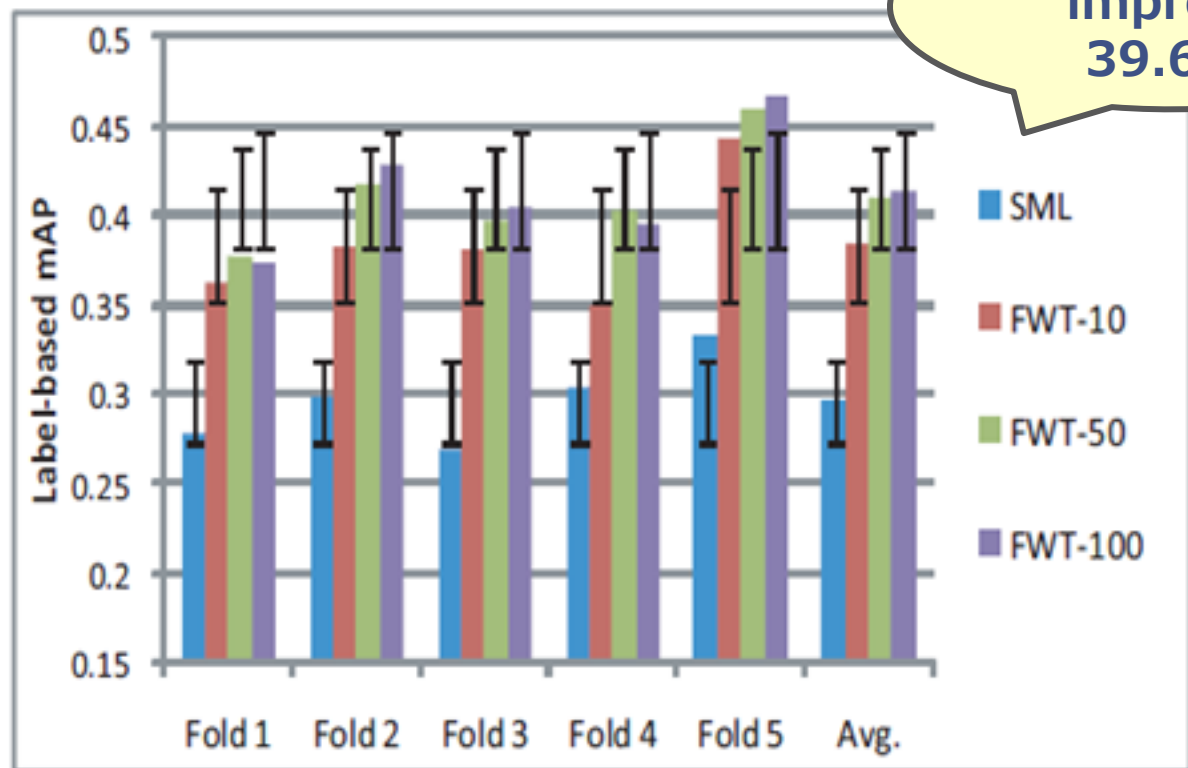
# Feature-Word-Topic Model: Demonstration

|                                                                                                                                                                                                     |                                                                                                                                                 |                                                                                                                                                                                                     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                                     |                                                               |                                                                                                                  |
| <p><b>Manual:</b> trees, sky, water, beach, cloudy, sailboat, boat, small, mast</p> <p><b>SML:</b> ferryboat, mast, many, cloudy, dock</p> <p><b>FWT</b> water, boat, mast, ferryboat, sailboat</p> | <p><b>Manual:</b> snow, polar, bear</p> <p><b>SML:</b> rope, ferryboat, polar, bear, geyser</p> <p><b>FWT</b> bear, polar, snow, beach, sky</p> | <p><b>Manual:</b> grass, sidewalk, sky, tree, clear, people, building</p> <p><b>SML:</b> engineering, bush, red, building, leaves</p> <p><b>FWT</b> building, sidewalk, bush, tree, engineering</p> |

# Experimental Results

## University of Washington Database (UWDB)

- 1109 images of size 300x200; vocabulary of size 292 words;
- For retrieval evaluation, we obtain top 20 words for indexing **Retrieval Evaluation**



**FWT-100  
improve  
39.6%**

# 考察

## 画像検索という実用問題からスタート

- 画像注釈：一昔前の感覚だと、とても出来そうもない難問
  - 画像認識の延長と思うと、難しそう
- Feature-Word-Topic モデル
  - なんか素人目には『凄そう』に見える
  - 実態は、既存技法の組み合わせとちょっとひねったアイデアで進歩
    - 「自然科学的な前進(積み重ね)」

## 計算理論の見地から

- 困難なアルゴリズム問題をたくさん含む
  - 高次元の混合クラスタリング
  - データ分布のガウス分布への分解
    - STOC等でたくさん近似アルゴリズムの論文があるが「使えない」
  - 学習効果の理論保証は？
- 評価は実験で示さざるを得ない

# 多項式アルゴリズムの追及から アプローチする例

多項式時間画像切り出しアルゴリズム

# 研究の経歴

画像切り出し： 通常はヒューリスティクス

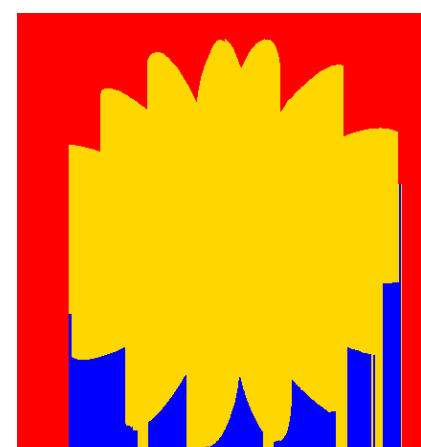
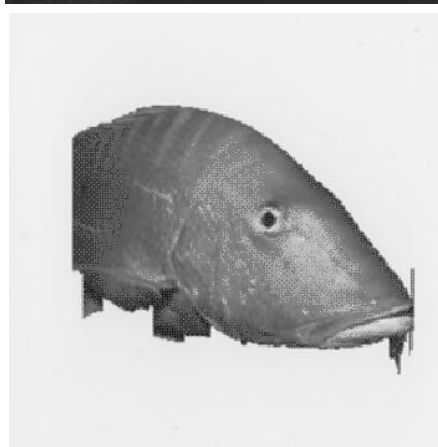
- Snakes (D. Mumford: 1974 Fields medal, 2010 National Medal of Science)



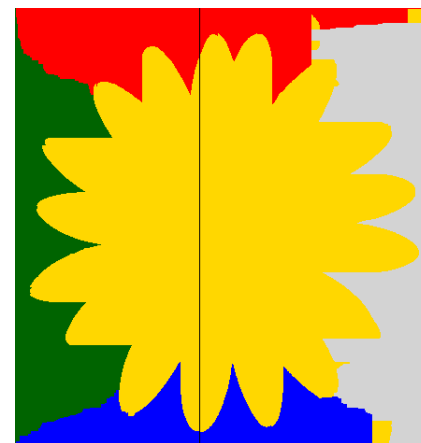
The rubber-band shrinks to minimize an energy function

**Asano-Chen-Kato-Tokuyama (1996 SODA)**

- Polynomial time solutions to image segmentation
- X単調もしくはY単調な連結な領域を切り取る
  - 最適性： 元の画像との輝度関数のL2距離を最小化



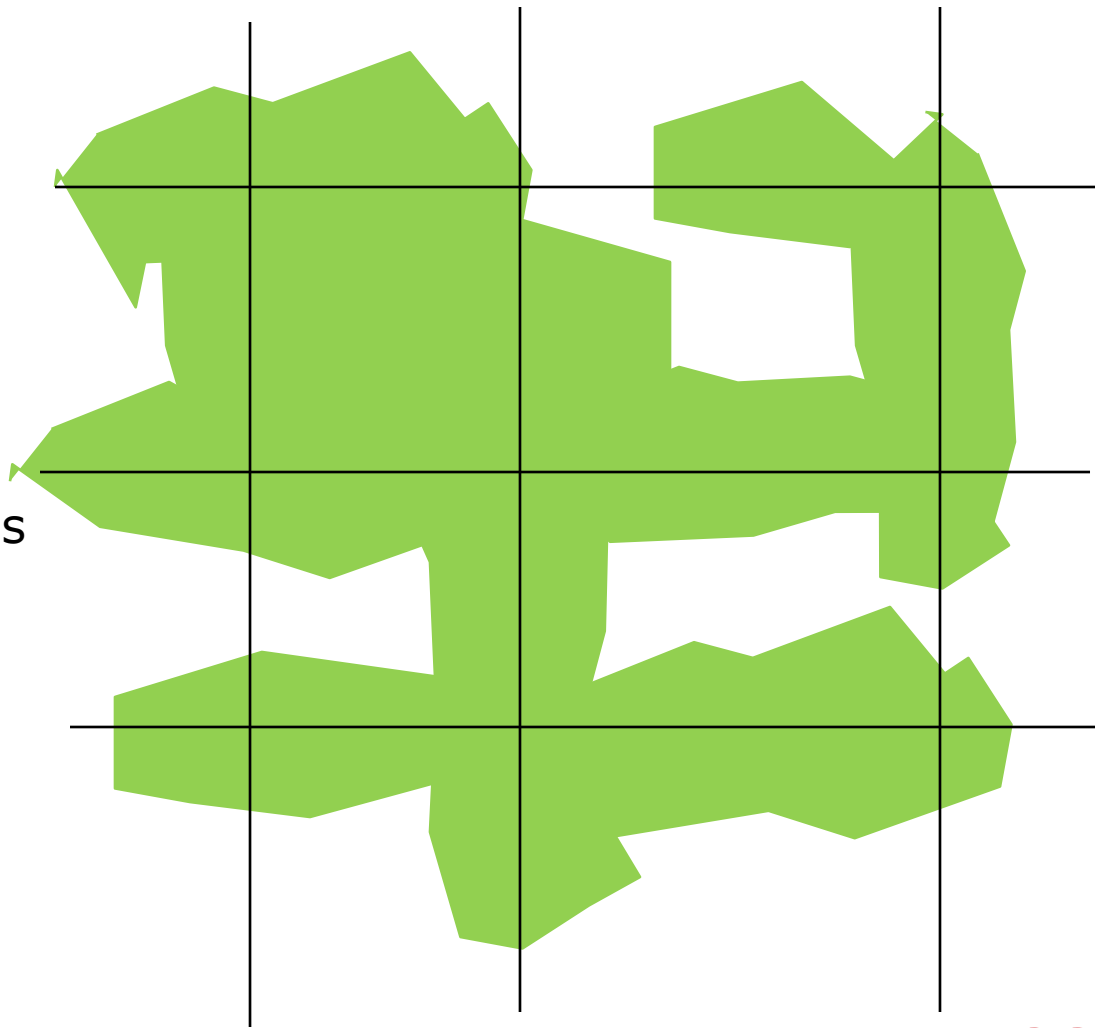
欠点：  
複雑な図形や複数図形の切り出しは無理  
問題：  
多項式時間でどこまで追求できるか



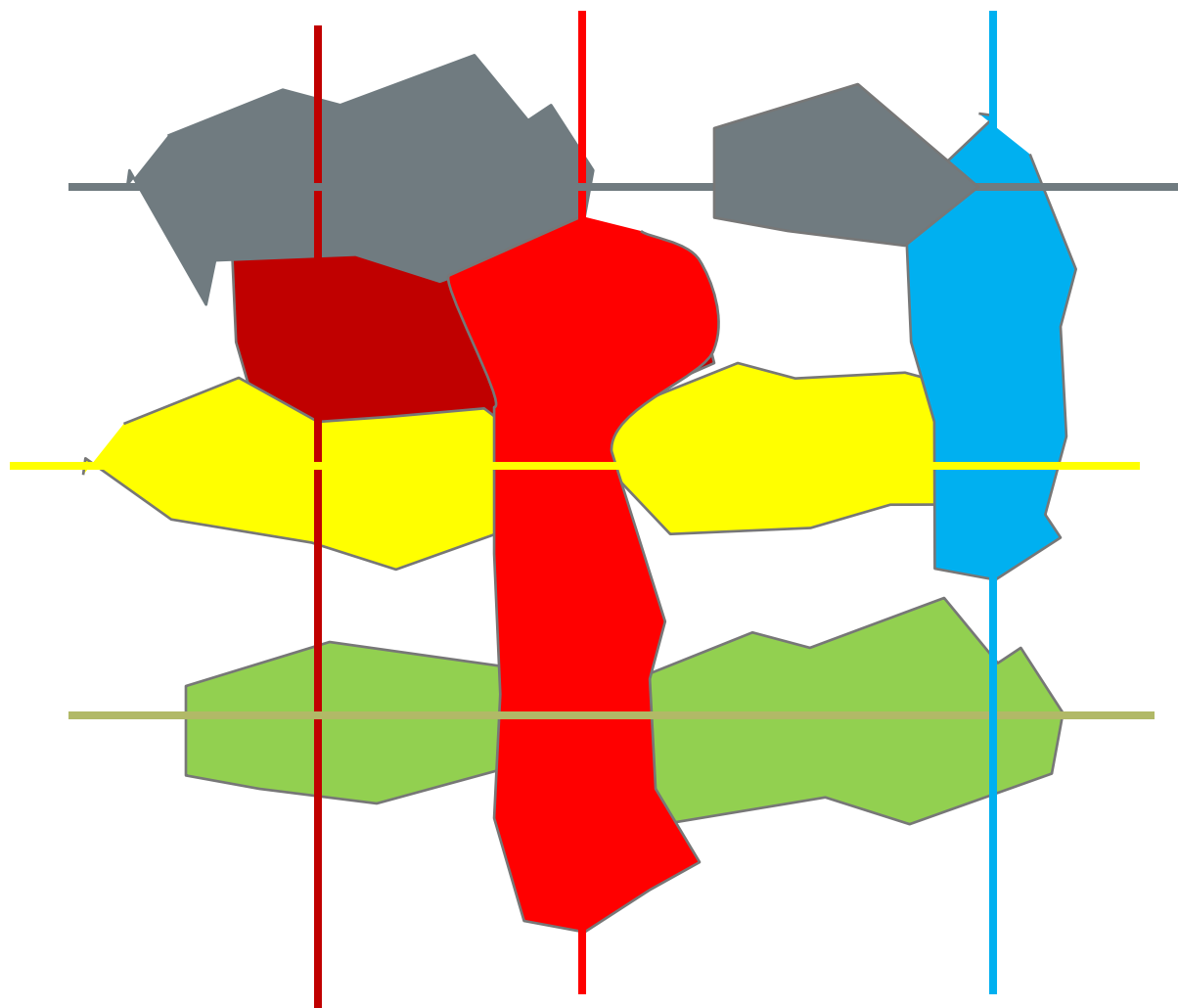
# ベースラインからの単調領域の非交差和は 多項式時間で切り出せる

右図の物体: 6本のベースラインから  
単調に広がる図形の非交差和

Jinhee Chun, Ryosei Kasai, Matias  
Korman, Takeshi Tokuyama:  
Algorithms for Computing the  
Maximum Weight Region  
Decomposable into Elementary  
Shapes. **ISAAC 2009**: 1166-  
1174

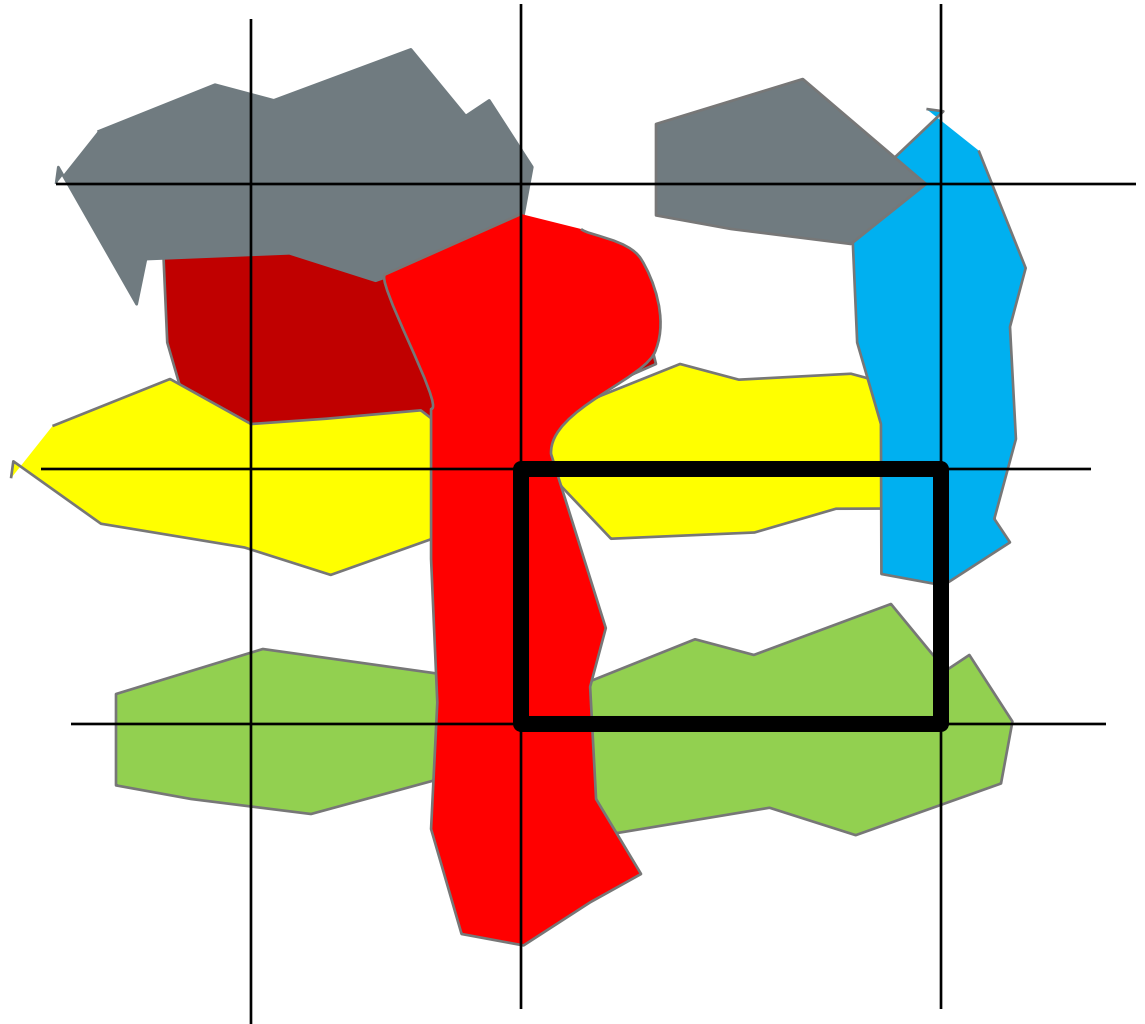


# ベースライン単調領域への分割



# The algorithm

- 部分ごとに解く
- それを結合



# Room problem:

|         |         |    |    |    |    |         |
|---------|---------|----|----|----|----|---------|
|         | $(n,1)$ |    |    |    |    | $(n,n)$ |
|         | 3       | 4  | 1  | 5  | 2  | -1      |
|         | 2       | -4 | -1 | -5 | -1 | 4       |
|         | 1       | -5 | 9  | 3  | -2 | 1       |
|         | 4       | 1  | -8 | 2  | 7  | 8       |
|         | 5       | -4 | 2  | 4  | -3 | 2       |
|         | -2      | 6  | 1  | -1 | 4  | 3       |
| $(1,1)$ |         |    |    |    |    | $(1,n)$ |

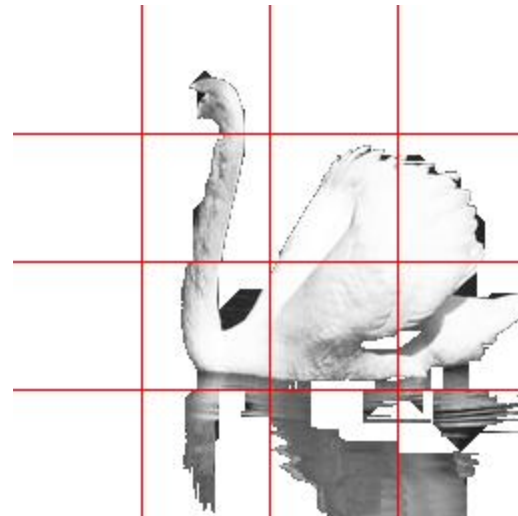
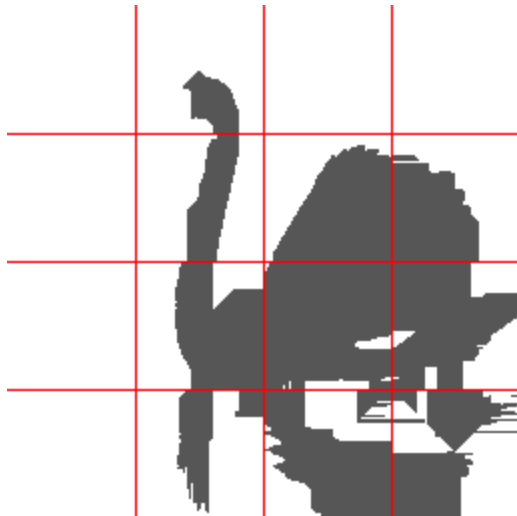


# 実験してみると

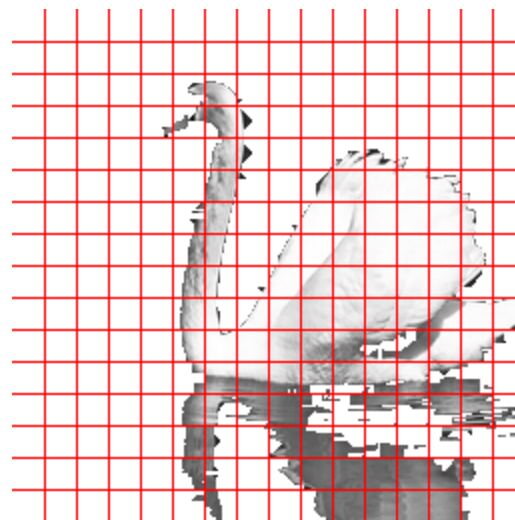
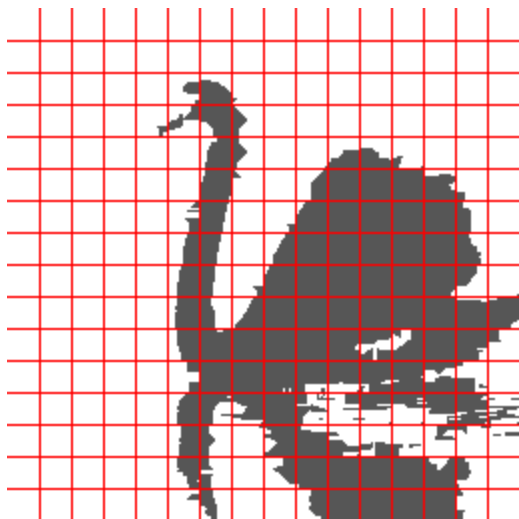
白鳥： ガイドラインなし (Room problem)



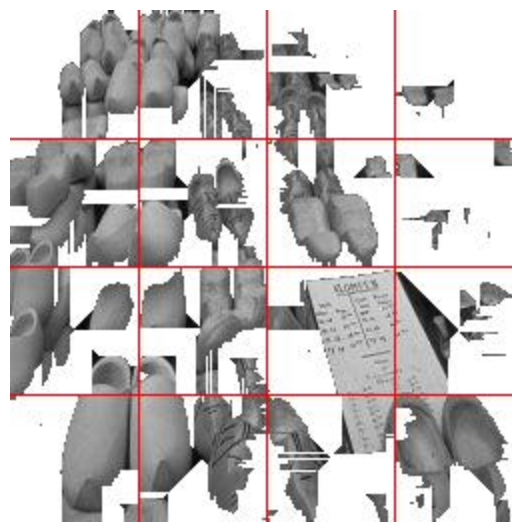
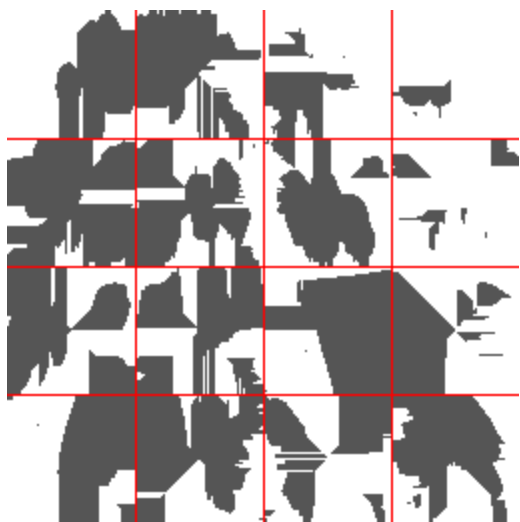
白鳥： 6本のガイドラインを均等に引いた場合

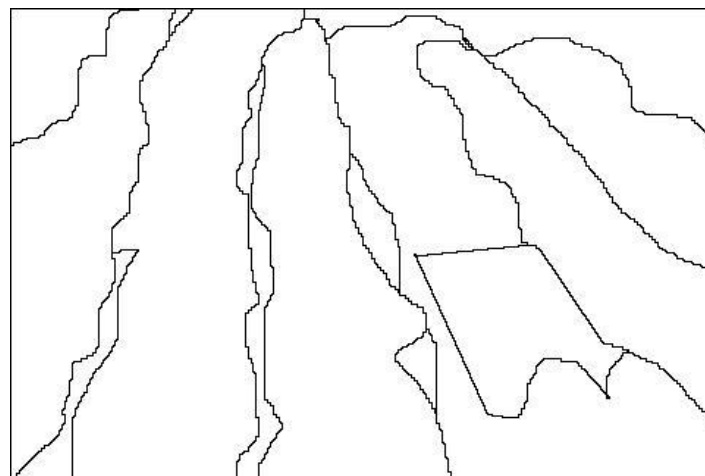
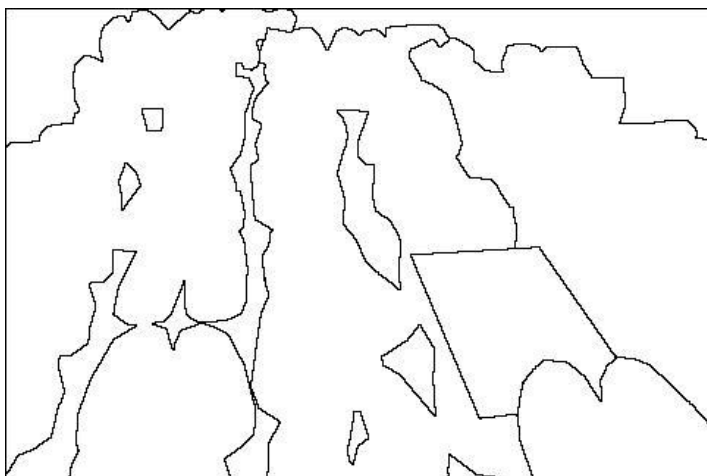
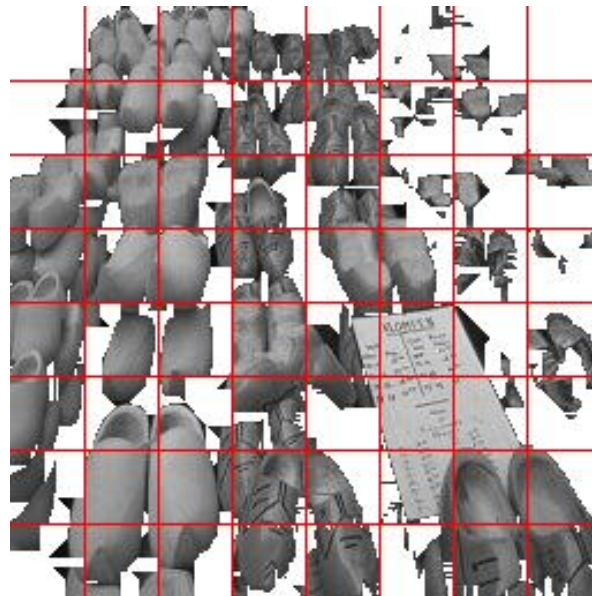


白鳥： ガイドライン30本の場合



## 靴屋さん： ガイドライン6本

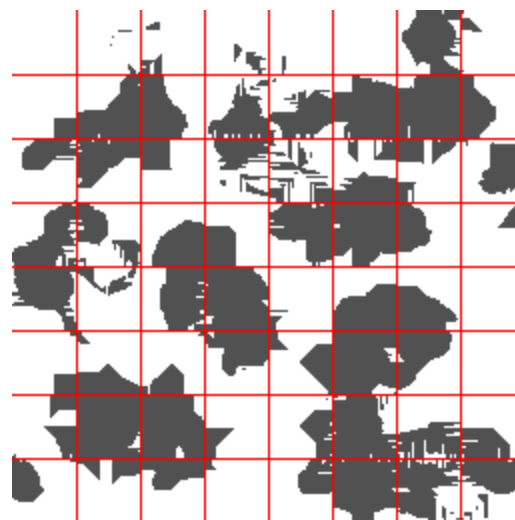
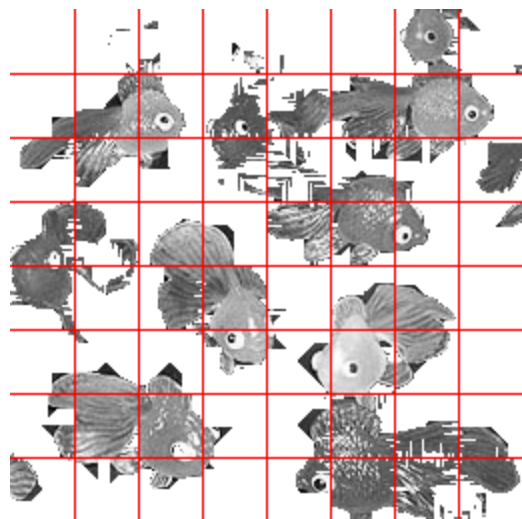
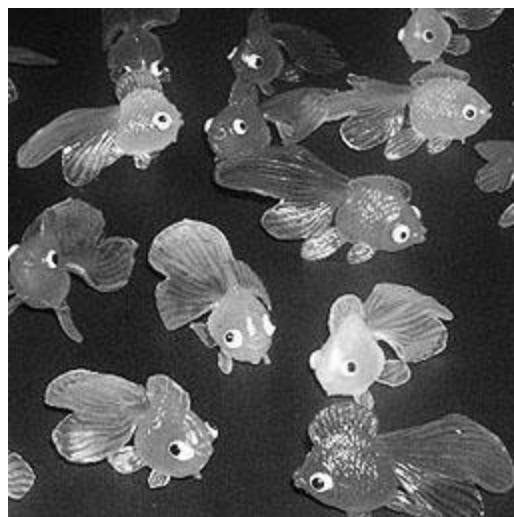




## Segmentation result by human

(<http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/BSDS300/html/dataset/images/gray/140075.html>)

# 金魚： ガイドライン14本



# 考察

## 画像切り出しという実用問題で

- 多項式時間でどこまでできるかというチャレンジ
- 単純な図形の「非交差和」を切り出せる理論結果
- 実装可能なアルゴリズム

## 実用の見地から

- 自由度が多すぎる??
- 滑らかさや連結性の制御はない

素人目の「凄み」には欠ける

# 終わりに: 数学と計算

## 最近の計算理論の風潮 : 数学的に高度化

- 「数学的に面白く、新規性と重要性が高い」  
「純粋数学」との違いは「重要性」の感性
- ○○の○○に興味があります (数学者)
  - なぜ? と聞く人はほとんどいない (数学的な美と歴史の重み)
- 「**なぜ興味があるか**」を情報社会の基盤として常に意識しているのが計算理論の研究者

## 世界の理論トップ研究者は応用分野へ発展

- 計算手法の進歩はE-サイエンスを先導する
- データベース、バイオ情報、グラフィクス、人工知能など
- それぞれの分野のトレンドを作る