



画像・映像の 認識と意味的検索

IPSJ音声研究会, IEICE音声言語情報処理研究会
テーマセッション「音声と情報検索」
2010年7月22日 仙台 秋保温泉 緑水亭

電気通信大学大学院 総合情報学専攻
柳井 啓司

アウトライン

- **画像・映像の「検索」と「認識」**
 - 意味的内容検索の重要性
- **最近の画像認識**
 - 特定物体認識 と 一般物体認識
 - 基本手法: Bag-of-Features (BoF)
 - 動作認識
- **意味的映像検索 と TRECVID**
 - 特徴量 と 特徴統合
- **まとめ**

1. 画像・映像の 「検索」と「認識」

画像検索 (1)

類似画像検索も

キーワード検索

「Web画像検索」が最も身近

ウェブ 画像 動画 地図 ニュース 書籍 Gmail その他 ▼ keiji.yanai@gmail.com | ウェブ履歴 | 設定 ▼ | ログアウト

Google 検索 セーフサーチ: 中 ▼

約 469 件 (0.07 秒) 検索オプション

類似の画像

« ラーメンの結果に戻る

他のキーワード: [ラーメン二郎](#)

 410x308 - 32 k - jpg ramenreview.kitaguni.tv 類似の画像を探す	 375x281 - 77 k - jpg seisansya.blog.ocn.ne.jp 類似の画像を探す	 800x600 - 177 k - jpg okiguru.seesaa.net 類似の画像を探す	 552x455 - 46 k - jpg gourmet.livedoor.com 類似の画像を探す	 1600x1200 - 171 k - jpg chile.sakura.ne.jp 類似の画像を探す	 640x480 - 59 k - jpg aurora.dti.ne.jp 類似の画像を探す	 640x480 - 133 k - jpg blog.livedoor.jp 類似の画像を探す
 1024x768 - 364 k - jpg jirochudoku.blog66... 類似の画像を探す	 500x386 - 72 k - jpg blog.livedoor.jp 類似の画像を探す	 1024x768 - 297 k - jpg jirochudoku.blog66... 類似の画像を探す	 480x360 - 45 k - jpg cho3.blog.fc2.com 類似の画像を探す	 640x480 - 96 k - jpg ujitegami.blog.shinobi.jp 類似の画像を探す	 800x600 - 198 k - jpg okiguru.seesaa.net 類似の画像を探す	 640x480 - 692 k - png kfb.co.jp 類似の画像を探す
 1024x768 - 310 k - jpg jirochudoku.blog66... 類似の画像を探す	 493x370 - 59 k - jpg blog.hokkaido-np.co.jp 類似の画像を探す	 1280x960 - 507 k - jpg chile.sakura.ne.jp 類似の画像を探す	 434x328 - 47 k - jpg 30min.jp 類似の画像を探す	 1024x768 - 126 k - jpg blog.livedoor.jp 類似の画像を探す	 640x480 - 249 k - jpg ramen-jiro.seesaa.net 類似の画像を探す	 1000x750 - 567 k - jpg kk-spc.co.jp 類似の画像を探す

Goooooooooooooogle ▶
1 2 3 4 5 6 7 8 9 10 次へ

検索



2003年

Google
Image
Search
の進歩

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.kathyscreations.com/.../
stuffed%20animals%
20cougar.html



lion.jpg



AliGloves.jpg



image16.gif



lion.GIF

lion animal

2005年

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

表示: [すべてのサイズ](#) - [大](#) - [中](#) - [小](#)



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

Google
Image
Search
の進歩



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: 虫

画像 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)

Endangered animals Amazon.co.jp 日本全国送料無料(1500円以上) コンビニ受取で好きな時受け取り可能 スポンサーリンク

2009年

Google Image Search の進歩



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 312x432 - 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 randomnessandiego.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



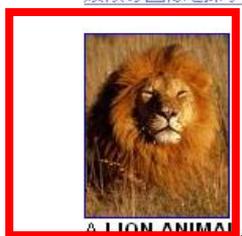
Cub Lion Picture 360x299 - 41k - jpg



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: 虫 ▼

検索結果 約 556 件中 1 - 20 件目 (0.04 秒)

2009年 類似画像検索

Google Image Search の進歩



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](#)



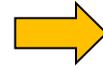
168×293 - 9k
[d230.org](#)
[類似の画像を探す](#)



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

画像検索 (2)

■ 位置で検索



ジオタグ(緯度経度)検索

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

秋保温泉 検索

人々 Recent

Also show photos not selected for Google Earth

地図 航空写真 地図+写真

[Untitled]

Panoramio 投稿: daisingeki

38° 13' 23.44" N 140° 43' 37.70" E

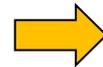
©2010 Google - 画像 ©2010 DigitalGlobe, Cnes/Spot Image, GeoEye - 利用規約

人気写真を Google Earth で見る

◀ 前 次 ▶

動画検索

■ タグ検索



キーワード検索

ニコニコ 動画 | 静画 | 生放送 | チャンネル | その他▼

一般会員 MM さん [プレミアム登録](#) | [マイページ](#) | [動画を投稿](#) | [ランキング](#) | [メニュー](#) ▼ ↑

ニコニコ動画 (9) キーワード タグ マイリスト

PIZZALA [ピザラネット会員に登録して応募すると、富士急ハイランドフリーパスが当たる!](#)

↓ 詳細情報・メニューを表示 ↓

インスタントラーメンを本気で作ってみた
夜中にラーメンを食べたくなるときってあるよね。実際に食べれたのは屋だったけど。(ノ`´*)他の動画→... (続きを読む)

タグ [料理](#) [本格インスタントラーメン](#) [お手軽14時間クッキング](#) [麵「お…お邪魔します…」](#)
[俺の作ってるインスタントラーメンと違う](#) [本気の人](#) [俺の嫁](#) [なるほど、わからん](#) [クッキングバカ](#) [ラーメン](#)

ニコニコ広告で [マイリスト](#) とりあえず [宣伝](#) [登録](#) [一発登録](#)

RSS [【ニコニコニュース】](#)
[【自記】新動画プレイヤーをリリースしました](#)

再生: 651,963 | コメント: 23,870 | マイリスト: 7,797 | 2010/07/21 (水) 20:22:01

コメント

再生時	コメント
00:25	くわせdriftgyふじこlp
01:26	くわせdriftgyふじこlpやめいwww
03:46	おおおおおお
04:57	あれかwwwwww200個卵wwwwww
04:58	Love Tambourinesだよね?曲
00:38	買得商品www
02:16	チェーンのラーメン屋を越えているw
03:39	次の日の深夜に喰うんだらうw
01:13	むちむち
01:53	本気すぎるだろ
03:27	でもこれ一回やっとなげばらくは手軽にいけるね。でも買った方が早い
04:31	500円は軽く出せるw
00:18	くわせdriftgyふじこlp: @:「
01:57	もう日が昇るぞ!

自動でスクロールする

メニュー | コマンド |

身近な画像・映像検索

ウェブ 画像 動画 地図 ニュース 書籍 Gmail その他 ▼

keiji.yanai@gmail.com | ウェブ履歴 | 設定 ▼ | ログアウト



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

Google

検索

セーフサーチ: 中 ▼

約 563 件 (0.08 秒)

検索オプション



50 k - jpg
fukuoka.jp
を探索



105 k - jpg



2048x1536 - 675 k - jpg
yutakaclub.jp
類似の画像を探索



480x424 - 44 k - jpg
www5c.biglobe.ne.jp
類似の画像を探索



925x518 - 146 k - jpg
kuzukiriz.blog111.fc2.com
類似の画像を探索



450x600 - 49 k - jpg
asitahare.exblog.jp
類似の画像を探索



1382x1037 - 256 k - jpg
aspcacommunity.ning.com
類似の画像を探索



400x300 - 34 k - jpg
k-hibino.at.weby.info
類似の画像を探索

alacarteichina.at...
類似の画像を探索

robert1966.blogoz.fc2.com
類似の画像を探索

anney.at.weby.info
類似の画像を探索

marukome.sblo.jp
類似の画像を探索



1280x960 - 84 k - jpg
ikumi-ferret.seesaa.net
類似の画像を探索



448x336 - 121 k - jpg
e-home-get.com
類似の画像を探索



380x285 - 24 k - jpg
starfort...
類似の画像を探索



640x459 - 91 k
blogs.yahoo.co.jp
類似の画像を探索



610x407 - 96 k - jpg
pub.ne.jp
類似の画像を探索



498x330 - 31 k - jpg
blog.goo.ne.jp
類似の画像を探索



340x252 - 27 k - jpg
shop.marche-e.com
類似の画像を探索

Go o o o o o o o o o o g l e

1 2 3 4 5 6 7 8 9 10

次へ



研究レベルの画像・映像検索

■ Keyword-based image / video retrieval

■ キーワードに基づく検索

実用レベル

- ・ テキスト検索と同じ。 研究対象とはならない。
- ・ 事前にキーワードが付与されていることが前提。



■ Content-based image / video retrieval (CBIR / CBVR)

研究レベル

■ 画像・映像特徴量に基づく検索. キーワード不要.

- ・ 画像・映像 ⇒ 画像・映像 の検索 と
キーワード ⇒ (キーワードなし)画像・映像 の2通り.

なぜ「研究レベル」の内容検索は

(20年以上研究されているのに) 普及しないか？

- 「意味」を考慮した画像検索ができない。
 - 見た目の類似性では検索できるが、意味レベルの類似性での検索ができない。



使えない技術？



類似



- すでに十分のキーワード付き画像がある。

- 写真共有サイト **flickr**[®] from **YAHOO!** には、45億枚の画像。大部分がタグ(キーワード)付き画像。
- **Google** 画像検索には、さらに多くの画像。
- **You Tube** の動画もほとんどキーワード付き。



キーワード検索で十分では？

Search

[Photos](#) | [Groups](#) | [People](#)

Everyone's Uploads

さしみ

SEARCH

[Full Text](#) | [Tags Only](#)
[Advanced Search](#)

Sort: [Relevant](#) | [Recent](#) | [Interesting](#)

View: [Small](#) | [Medium](#) | [Detail](#) | [Slideshow](#)



From fake_plastic...



From ryanchirnom...



From chris-lh



From (^_~)...



From nansei



From Dennis Wong



From earthhopper



From earthhopper



From kirschbrunne...



From joart



From earthhopper



From earthhopper



From earthhopper



From earthhopper



From m.fam



From ryanchirnom...



From new flower



From silverlily



From earthhopper



From Jasmin* ** * *



From chris-lh



From earthhopper



From Mmonhsi



From Jasmin* ** * *



From Fuyuhiko



From shimashimane...



From nansei



From Coconutto



From Onildo_Lima ...



From masaaki...



From earthmaratho...



From yuichi.sakur...



From earthhopper



From silverlily



From earthhopper



Groups [more...](#)

- Japanese Food**
5,760 members | 60,802 photos
- I Ate This**
20,915 members | 361,938 photos

Photographers [more...](#)

- localjapantimes - Tokyo, Japan**
See related photos
- earthhopper - Okinawa, Japan**
See related photos

Places [more...](#)

- Tokyo, Tokyo Prefecture**
See related photos
- Karatsu-shi, Saga Prefecture**
See related photos

Google

さしみ

検索

セーフサーチ: 中 ▼

約 269,000 件 (0.12 秒)

検索オプション

すべて
動画
画像
もっと見る

すべてのサイズ
中
大
アイコンサイズ
最小サイズ:
サイズを指定:

すべての種類
顔
写真
クリップアート
線画

すべての色
フルカラー
白黒
色パレット

もしかして: 刺身



さしみ定食.jpg
500×375 - 50 k - jpg
hisuinatosato.com



さしみに醤油
590×472 - 58 k - jpg
yugeta.com



04
400×314 - 63 k - jpg
shodo.co.jp



おさしみ.jpg
420×279 - 86 k - jpg
ie-shinchiku.jp



京都の「さしみの魚紳」
700×378 - 93 k - jpg
ahiru-ie.way-nifty.com



3さしみ.jpg
498×374 - 55 k - jpg
shodo.co.jp



さしみ.jpg
300×225 - 18 k - jpg
hobidas.com



さしみ.
450×337 - 119 k - jpg
photozou.jp



さしみ
300×300 - 29 k - jpg
item.rakuten.co.jp



お徳用さしみ
350×297 - 49 k - jpg
item.rakuten.co.jp



浜めし・さしみ.
450×338 - 51 k - jpg
photozou.jp



これで1200円!
400×300 - 94 k - jpg
nekononjikan.sagafan.jp



大漁屋のさしみ定食
550×355 - 51 k - jpg
hitoritabi.exblog.jp



さしみ
800×535 - 310 k - jpg
f.hatena.ne.jp



こくのある甘口・さしみ
480×360 - 29 k - jpg
item.rakuten.co.jp



A15
500×300 - 44 k - jpg
shodo.co.jp



おさしみ皿
380×380 - 136 k - jpg
edgewire.blog110.fc2.com



こんににさしみこんにゃ
570×427 - 48 k - jpg
zeitaku.jp



酒の肴に粒うに入さしみ
1280×960 - 459 k - jpg
estelle-foods.co.jp



おさしみねぎとろ
640×480 - 176 k - jpg
item.rakuten.co.jp

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



ニコニコ動画 (9)

キーワード タグ マイリスト

検索



ピザーラネット会員に登録して応募すると、

富士急ハイランドフリーパスが当たる!

インスタントラーメンを本気で作ってみた

夜中にラーメンを食べたくなるときってあるよね。実際に食べたのは昼だったけど。(´▽`)他の動画→... (続きを読む)

↓ 詳細情報・メニューを表示 ↓

[#P30 料理](#) [本格インスタントラーメン](#) [お手軽14時間クッキング](#) [麺「お…お邪魔します…」](#)
[登録タグ](#) [俺の作ってるインスタントラーメンと違う](#) [本気の人](#) [俺の嫁](#) [なるほど、わからん](#) [クッキング/パカ](#) [ラーメン](#)

ニコニコで マイリスト とりあえす
 宣伝 登録 一発登録

再生: 651,963
 コメント: 23,876

意味を考慮した複雑な検索が
 可能になる必要がある!
 でも、現時点ではまだ困難。



使い方 | 不適切コメントの通報 | 動画の違反通報 | プレイヤーが表示されない

[待受/Flash](#) [カップラーメン・タイマー](#) 麺々軒 ... 262人が購入/この動画で131人、全体で179人がクリック

[待受/Flash](#) [主食になる待受\(ラーメン\)](#) 定食屋さん ... 37人が購入/この動画で71人、全体で210人がクリック

[寄附](#) [中華料理](#) 山崎まさよし ... この動画で4人、全体で24人がクリック

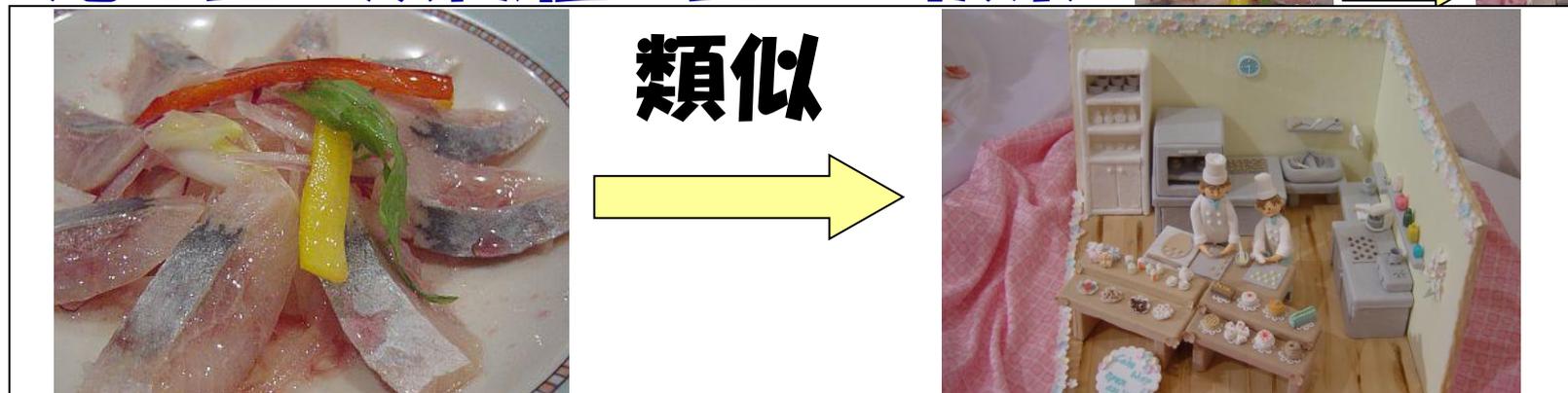
- ・「チャーシューメンを食べているシーン」が見たい時。
- ・「仙台の秋保温泉での露天風呂シーン」が見たい時。

から見たい
 探すには
 検索が不可欠

(話は戻って)

2種類の内容検索(CBIR / CBVR)

■ 見た目の類似性に基づく検索



“semantic gap”

見た目の類似性は、意味の類似性と一致しない！

“semantic gap” が存在！

“semantic gap”の克服のためには 意味の「認識」が必要！

- 「一般物体認識」：物体の「カテゴリー名」を認識
 - 「画像認識」「コンピュータビジョン」研究の一部。
 - ・ IPSJ-CVIM研究会. IEICE-PRMU研究会. MIRU2010.
 - 自動キーワード付与 とみなせる。



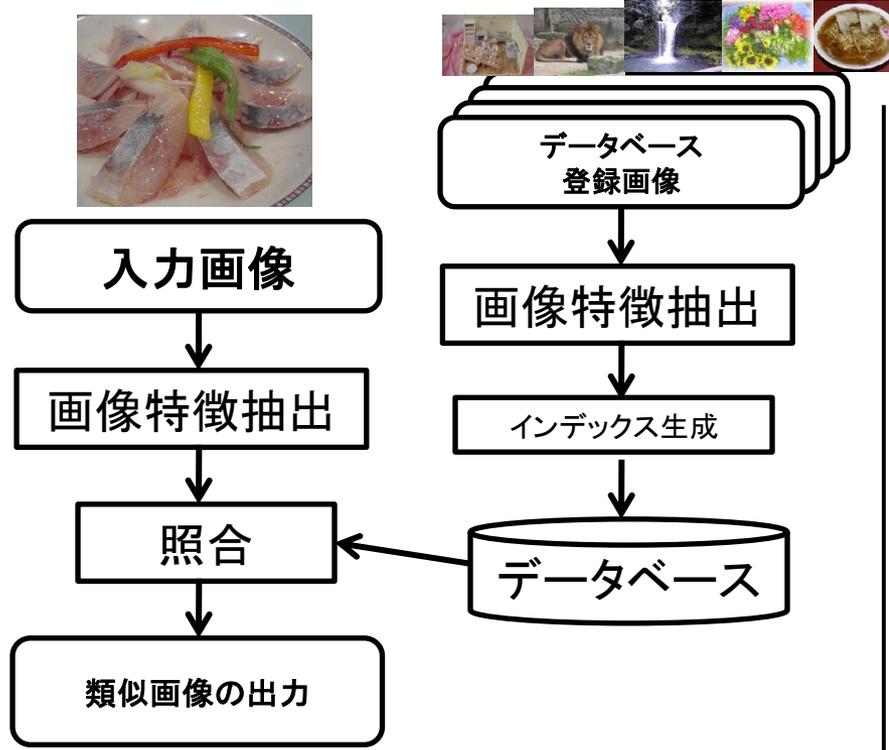
→ 「さしみ」

“semantic gap”の克服は容易ではない。
画像認識研究の究極の目標

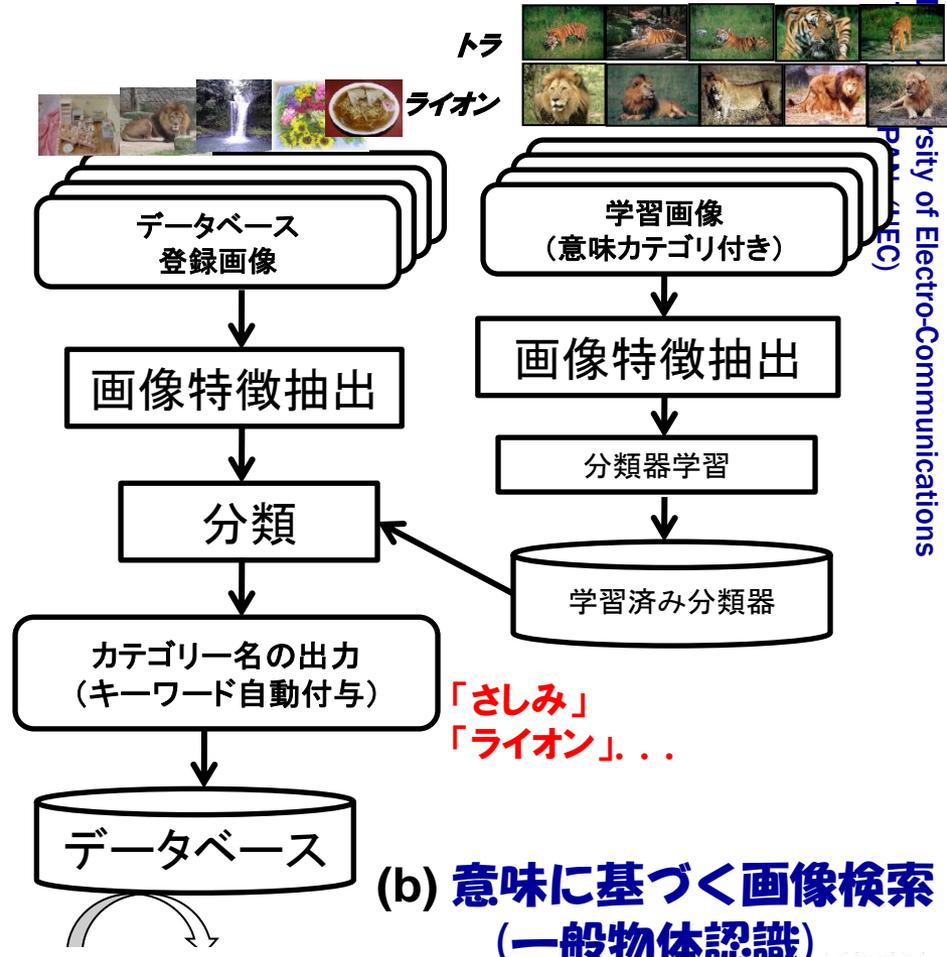
内容検索の手順



University of Electro-Communications



(a) 見た目に基づく画像検索



(b) 意味に基づく画像検索 (一般物体認識)

キーワード検索。
 場合によっては、ユーザのフィードバックに基づいて、見た目に基づく画像検索を実行。



2. 最近の画像認識

【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.

「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

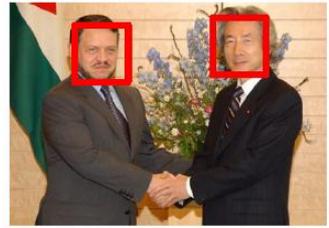
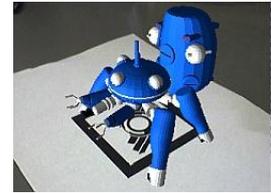
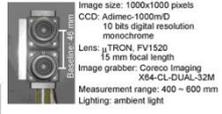
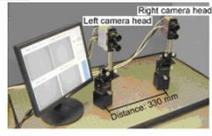
■ 3次元復元, ステレオ復元.

■ 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)

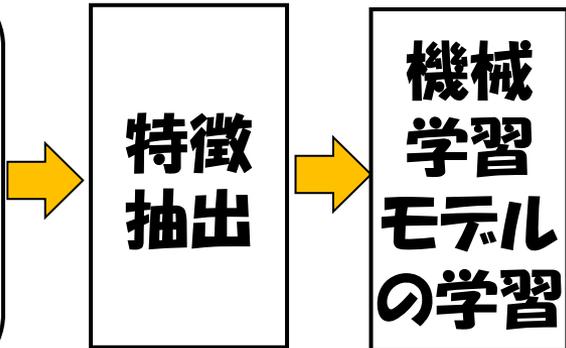


(狭義の)「画像認識」の基本的な処理

学習



ラベル付き学習データ



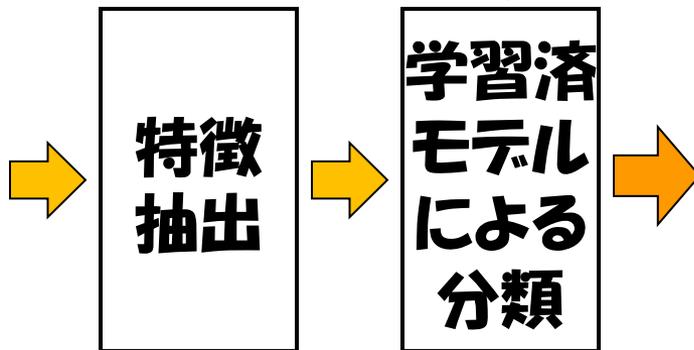
3つの基本要素

- 学習データ
- 特徴抽出
- 機械学習手法

分類



未知の画像



「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 + 機械学習 + 学習データ

- **特徴抽出の決定版の手法の登場 (2004年)**
 - **Bag-of-Features (Bag-of-Keypoints(BoK) 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。**
- **機械学習の進歩**
 - **SVMなどの機械学習手法の進歩**
 - **オープンソースライブラリの整備**
- **Webからの学習画像データの容易化**
- **もちろん、計算機の高速・大容量化も。**

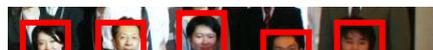
2000年以降、研究が急速に発展！ 今や大ブーム。

「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



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

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

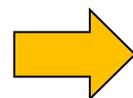


■ 同一物体認識

- 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



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

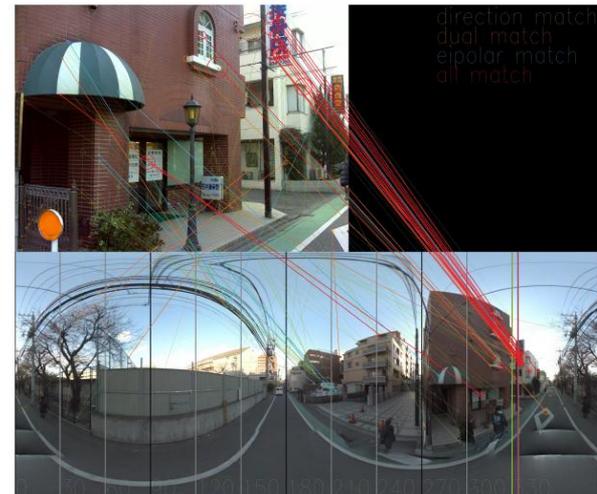
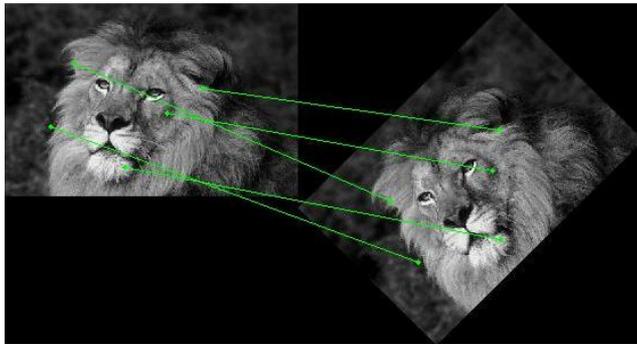


現時点では剛体なら
90%以上認識可能. ただし, 顔や動物などの変形物体は難しい.



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



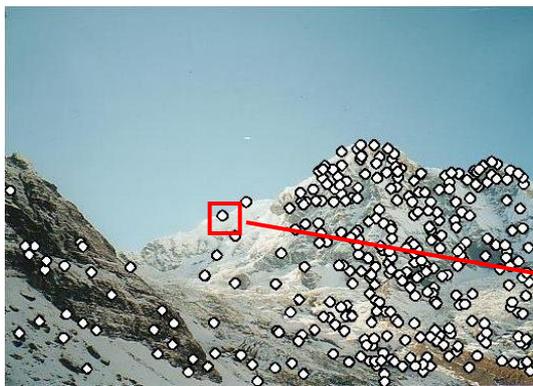
■ 大量の特徴点に対する検索

- 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

認識というよりも
“同一部分検索”

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)



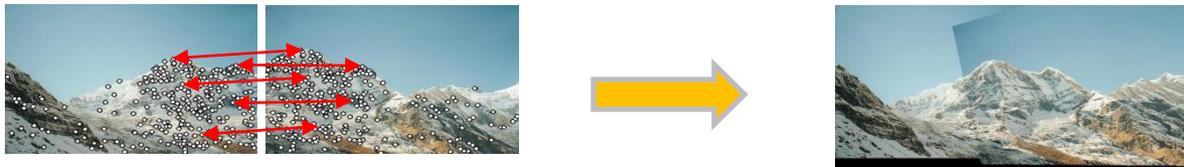
特徴点ベクトル: 回転, 拡大縮小があっても, 同じ見た目のパターンはほぼ同じベクトル値になる



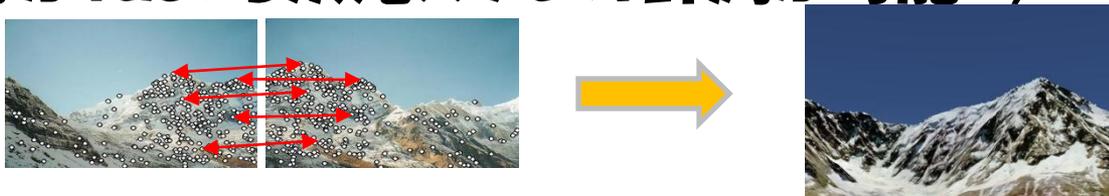
局所特徴量による対応点探索

■ 2枚の画像の点の対応が分かると..

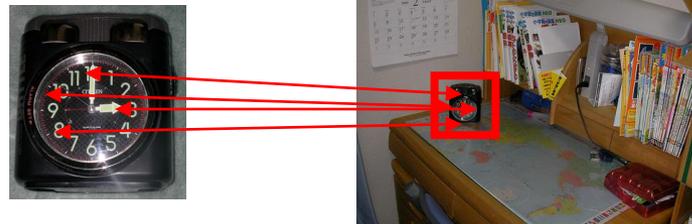
■ パノラマ画像生成. (複数画像からの大画像生成)



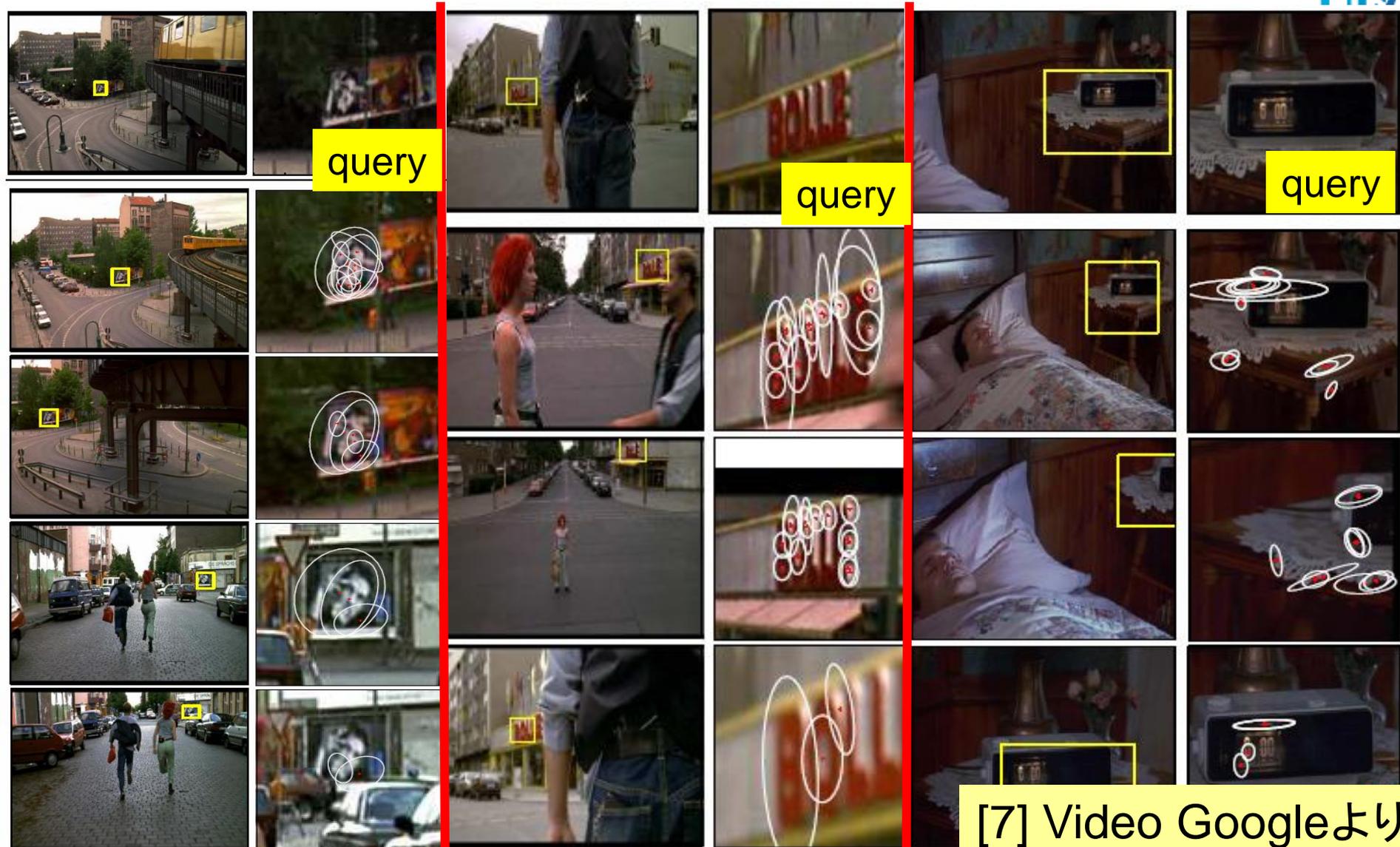
■ 3次元計測. (2枚の対応と相対撮影位置が分かれば, 3次元ステレオ計測が可能.)



■ 特定物体認識. (同一物体の対応付け.)



局所特徴による特定物体検索



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

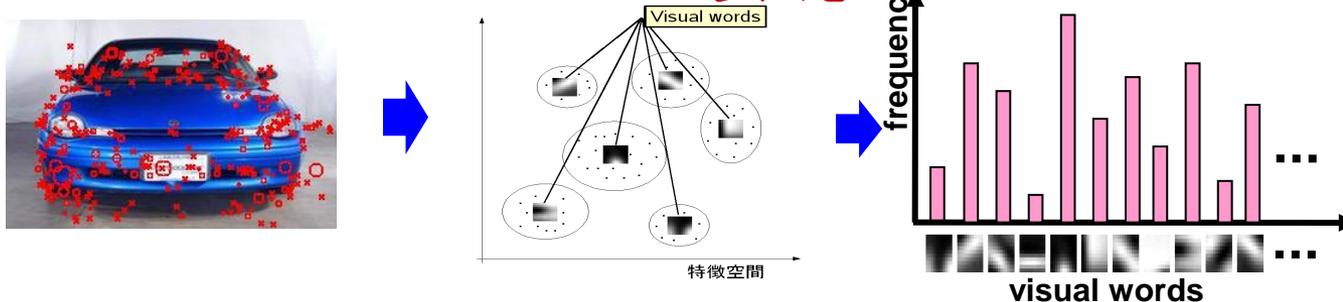
静止画像に対して, その中に含まれる
物体もしくはシーンの一般名称(カテゴリー)を認識

究極的には人間以上にあらゆる画像を認識

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

一般物体認識の分類 (1)

■ 画像全体の 카테고리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



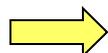
→トラ
草原



→ゾウ
キバ
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

道路

自転車

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

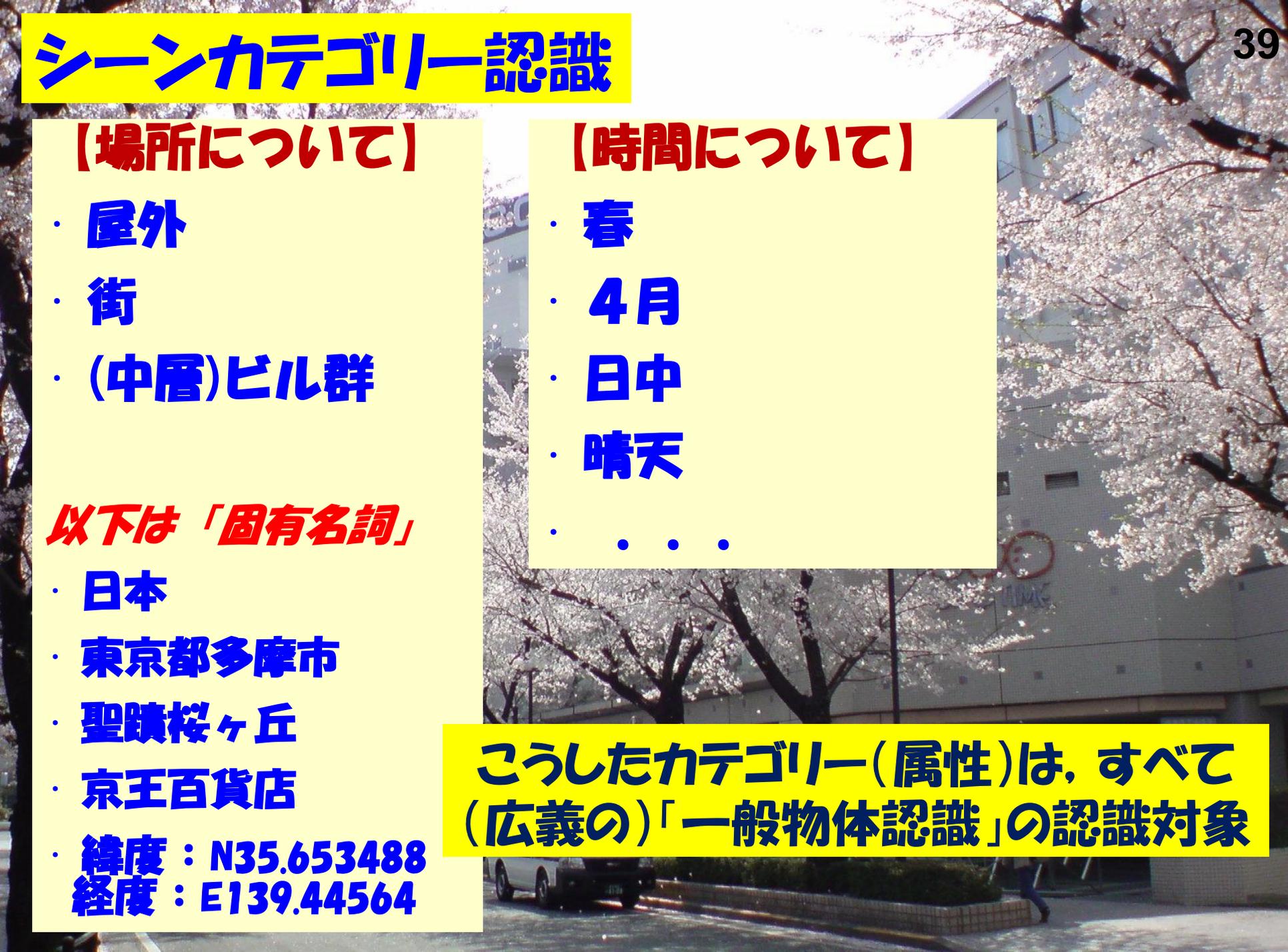
【時間について】

- ・ 春
- ・ 4月
- ・ 日中
- ・ 晴天
- ・ . . .

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

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」

一般画像認識の歴史

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.

- Identificationのみ。形状既知。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identification中心.

- 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。

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

2002年 Word-image translation model

2003年 Video Google (image search by **visual words**)

2004年 **Bag-of-Features**(BoF) + SVM

2004年 Caltech101 (101カテゴリーのデータセット)登場

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

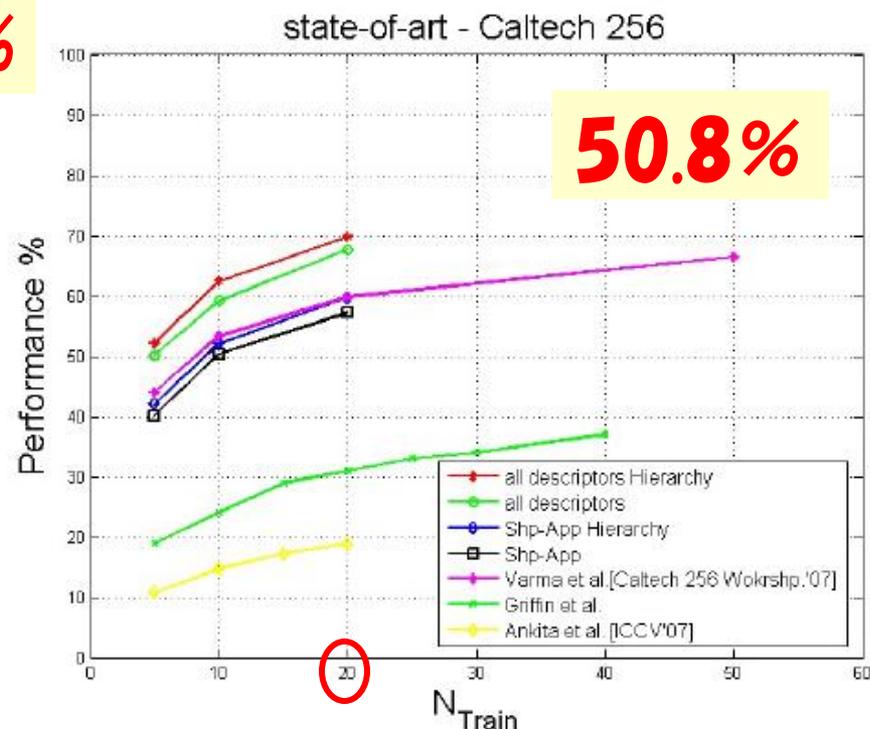
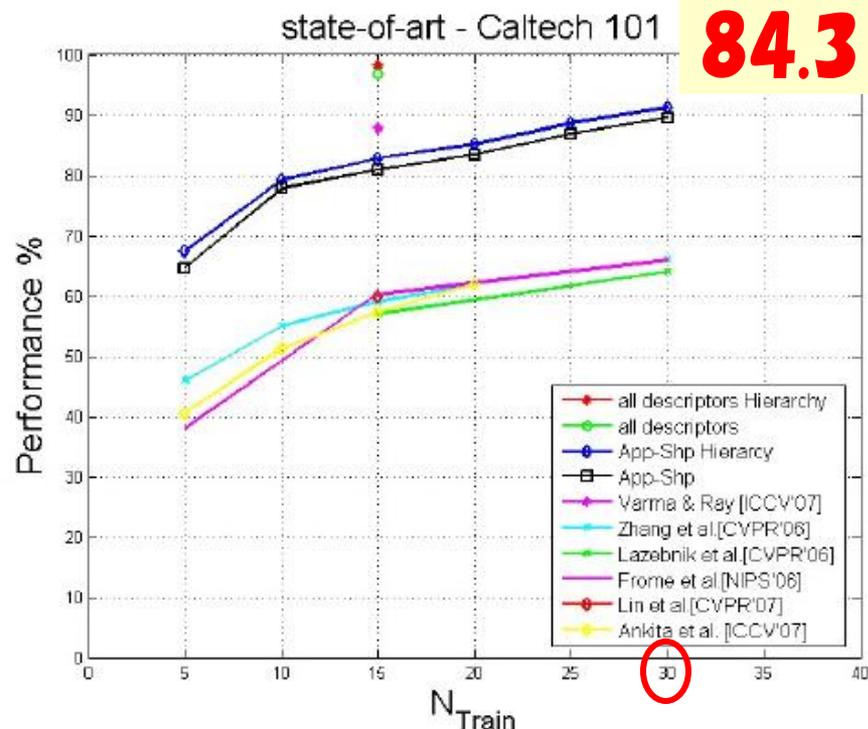
BoF + SVM with modified kernel

BoF + CRF for semantic region segmentation

2007年 Caltech256 (256カテゴリーのデータセット)登場

最新の一般物体認識の分類精度

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



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 → 新しいフレームワークが必要！

3. 基本的な手法:

Bag-of-features

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
 - 問題点: 物体の隠れや変形に弱い.



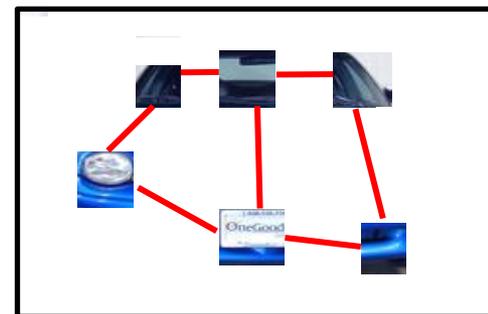
原画像

Constellation model

- Part-based手法の登場:

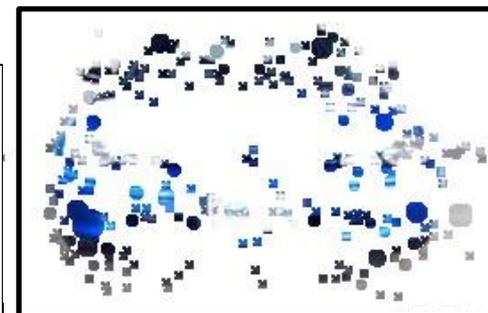
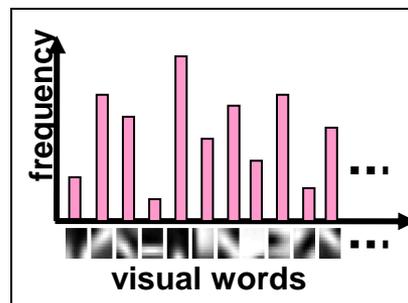
複数の部分の組み合わせで認識

位置関係を考慮すると視点変化に弱くなる.



星座モデル

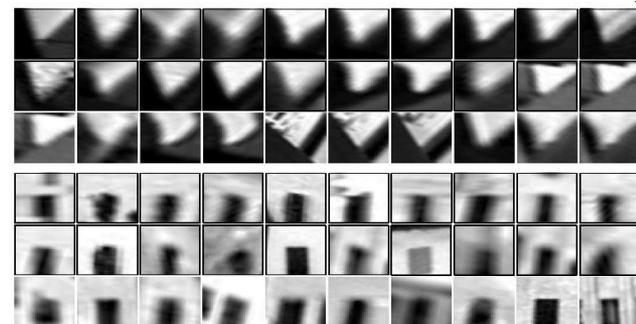
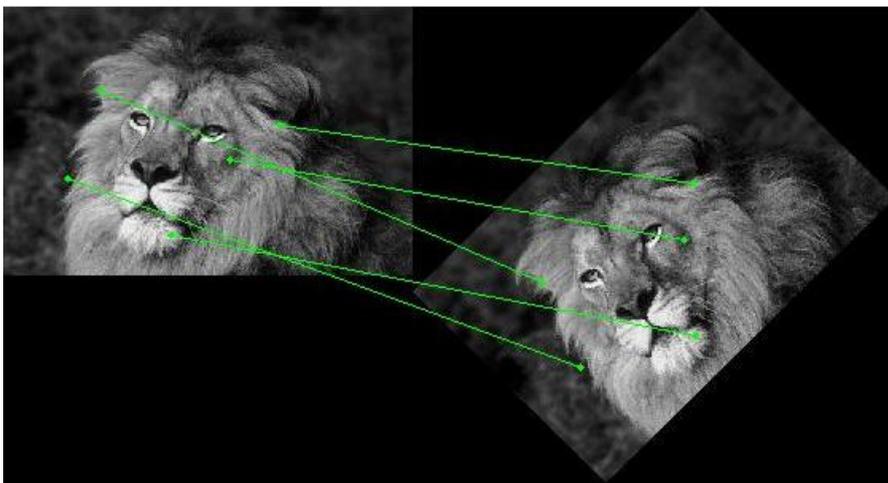
- 局所パターン↓の分布に基づく認識 (bag-of-features)



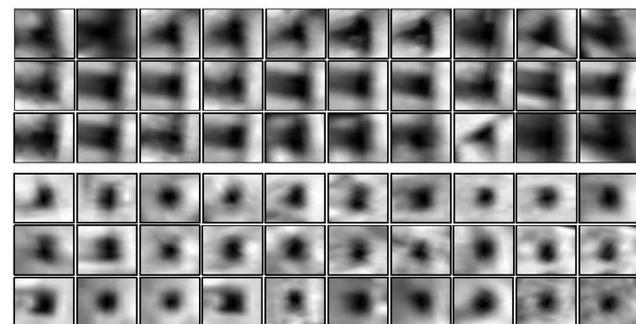
Bag-of-featuresモデル

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
 - ・ フリーソフトを利用可能.



(a)



(b)

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



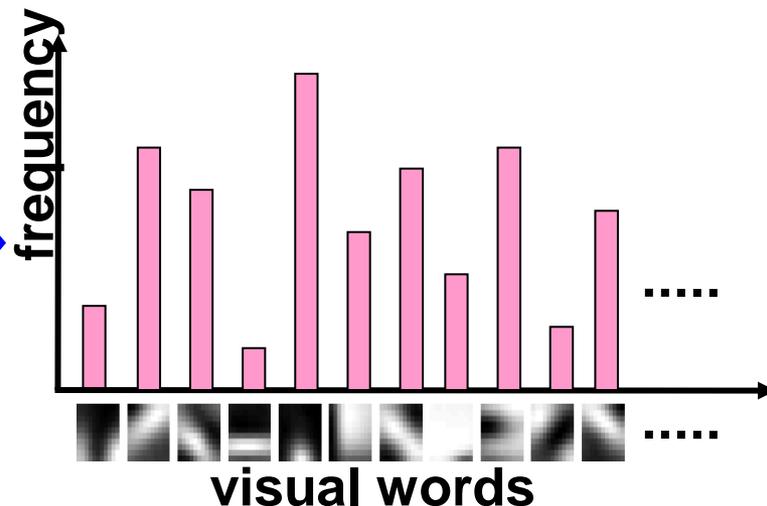
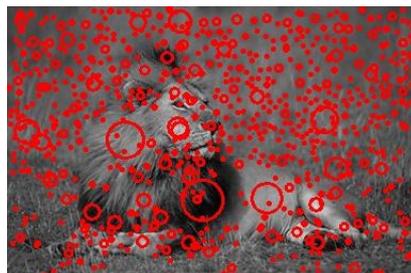
- Visual words の集合として画像を表現
 - Visual words のヒストグラムを画像特徴とする
 - 単語出現頻度によりテキストを表現する方法の bag-of-words の考え方を画像に応用. 語順を無視するのと同様に, 位置を無視.
- Bag-of-features によって表現された特徴ベクトルを Naive Bayes, SVM などの機械学習手法で分類. **テキスト分類と同じ!**
- Bag-of-visual-words (BoVW), **Bag-of-keypoints (BoK)** とも言うことがある.

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

■ 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)



特徴点のサンプリングの方法

■ 主な3つの方法

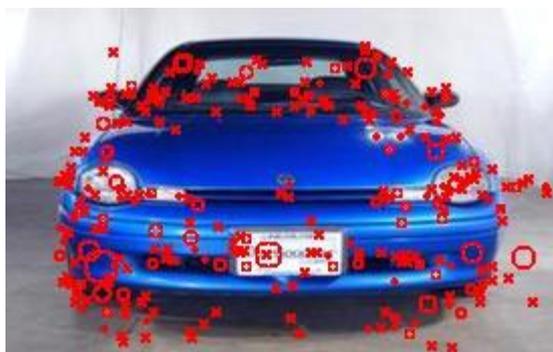
- Difference of Gaussian (DoG)
- Random sampling
- Grid sampling

SIFT法の方法

こちらが主流

sparse sampling と呼ぶ
1画像 数百点

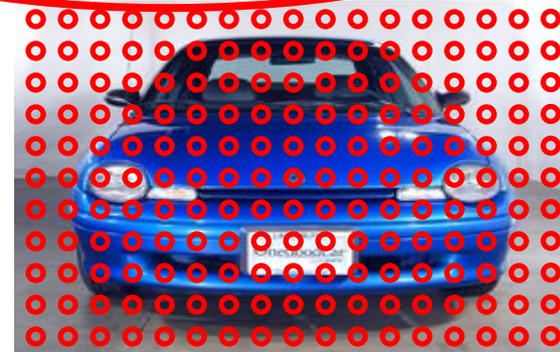
dense sampling と呼ぶ
1画像 数千点



DoG (sparse)



random(dense)



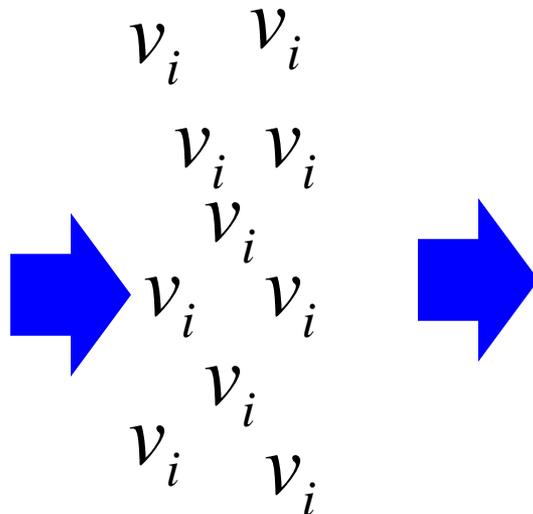
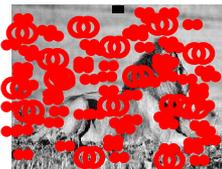
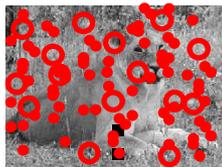
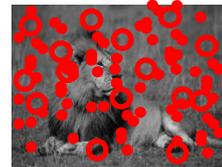
grid(dense)

カテゴリー分類 (classification) においては、
パターンのない部分の情報も重要。

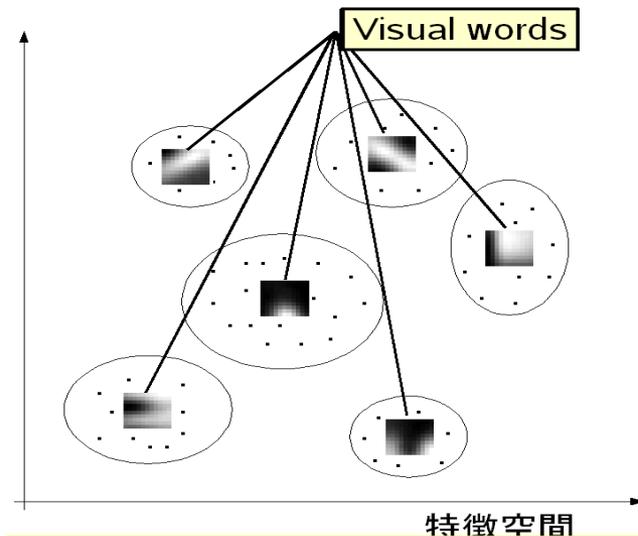
Visual words の求め方

- 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”



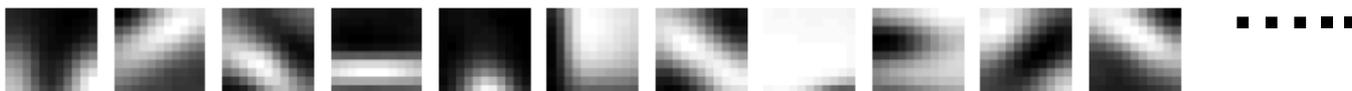
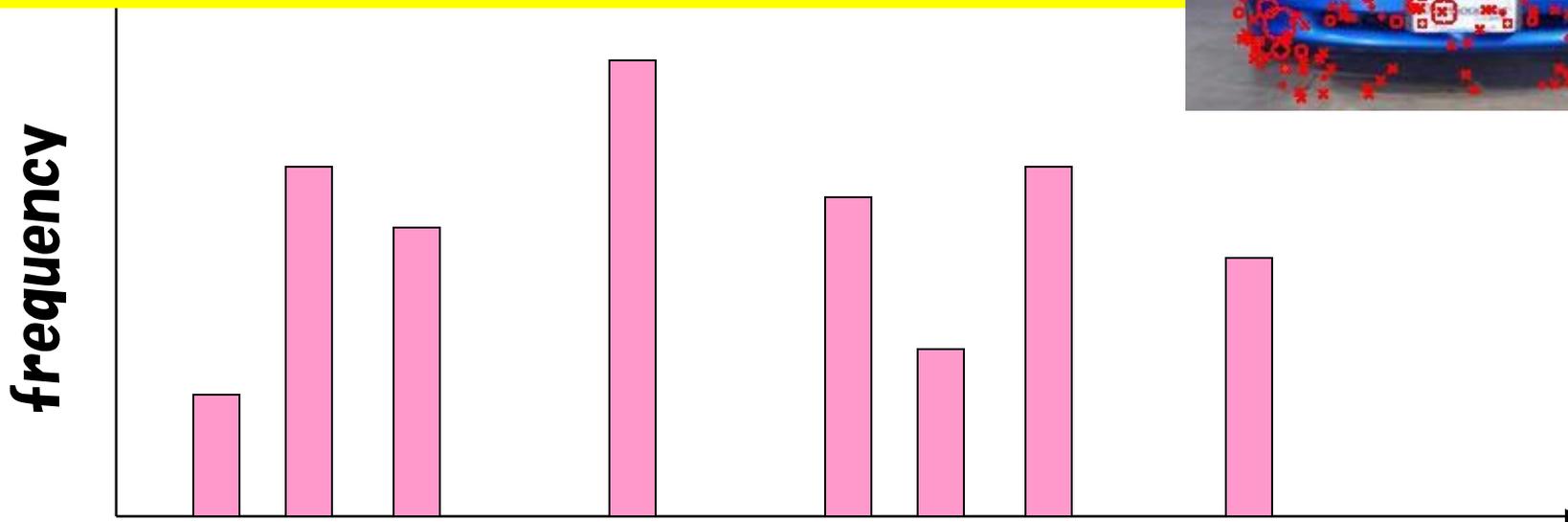
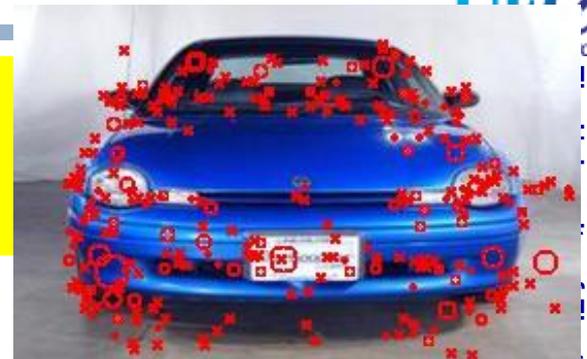
SIFT vectors



“Visual words” は, 代表的な局所パターンに相当する.

Bag-of-features 表現

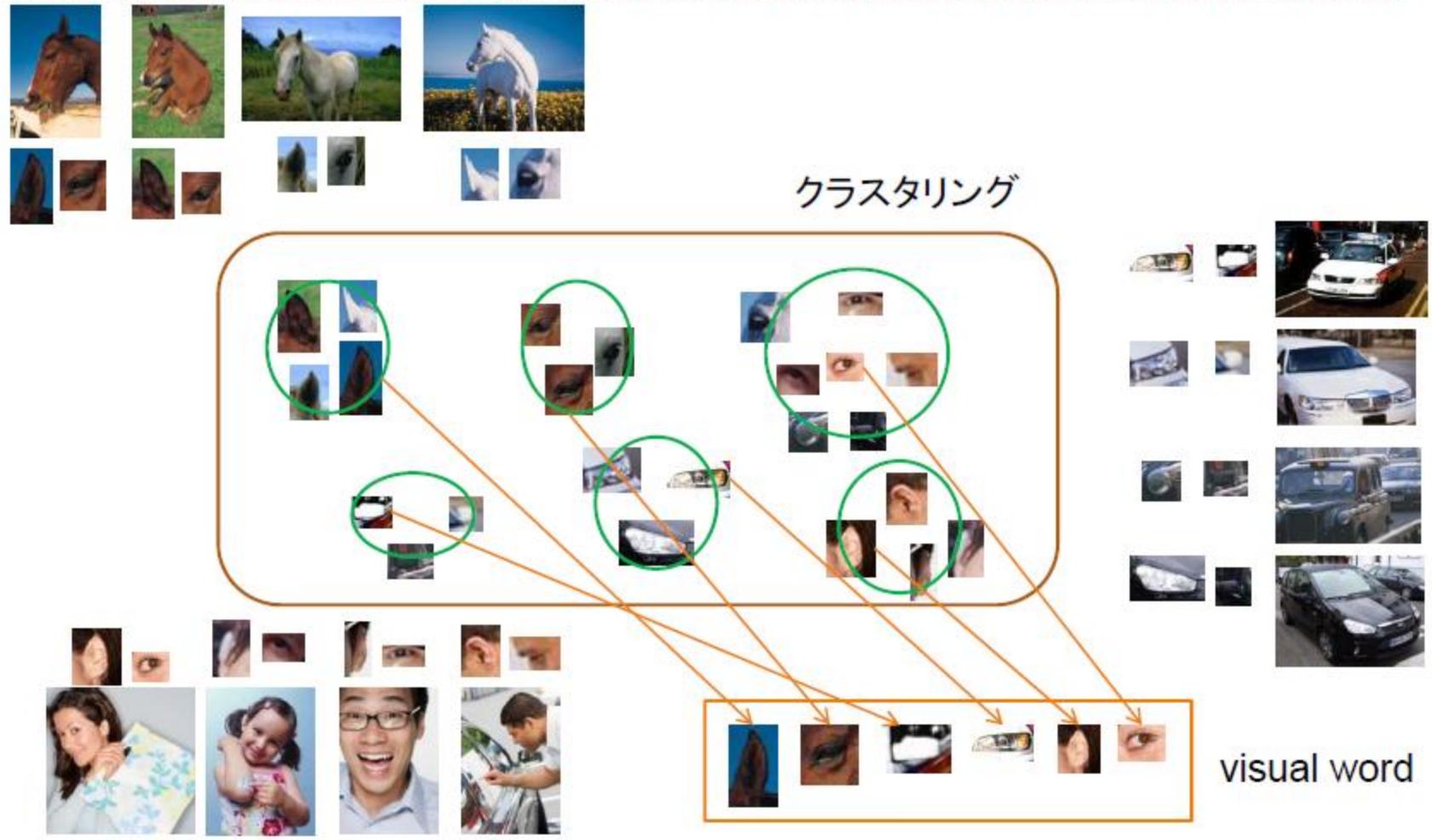
Visual wordsの出現頻度(ヒストグラム)によって画像を表現.



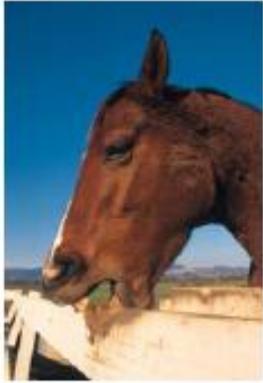
Visual words
(数百～数千個)

次元は、数百～数千次元。
スパースなベクトルになる。

Bag of Features



個々の画像の表現



Bag-of-features 表現: 代表パターンの集合による表現

Object



Bag of 'words'

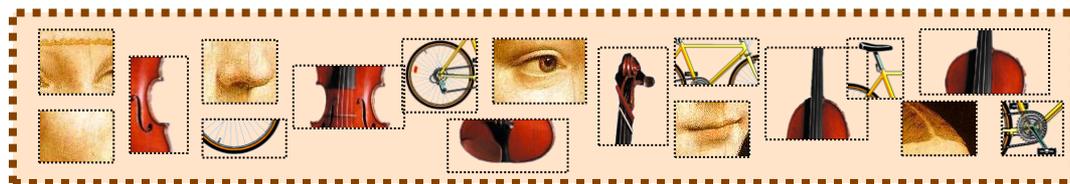
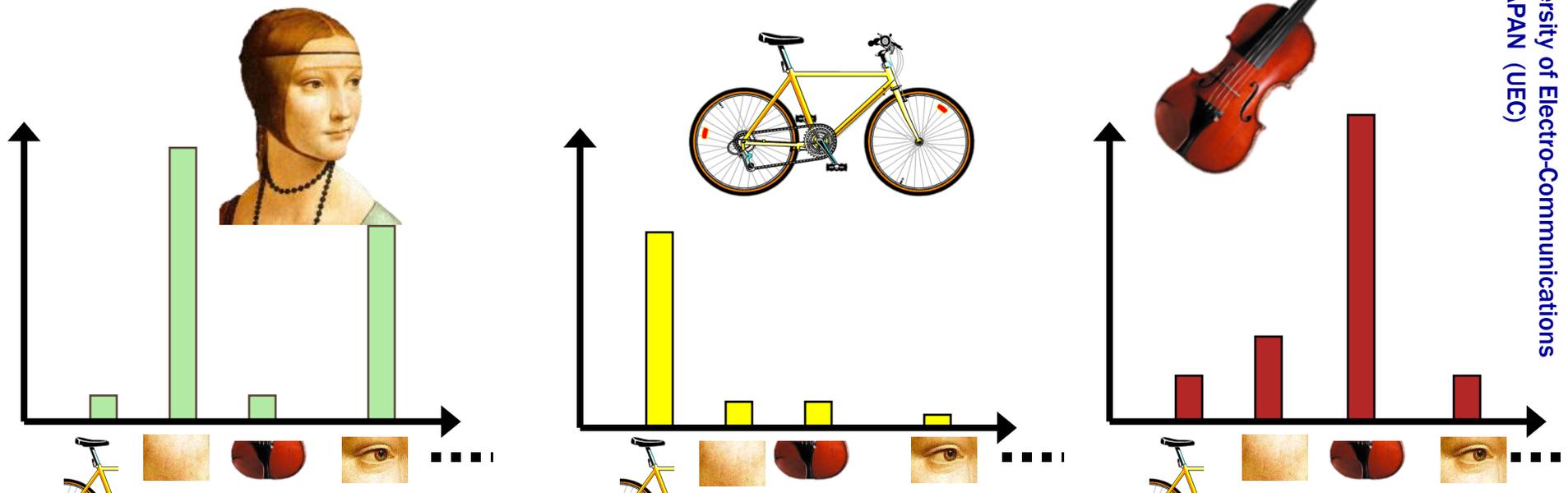


[ICCV 2009 tutorial スライドより]



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



Bag-of-features表現を用いた 画像認識

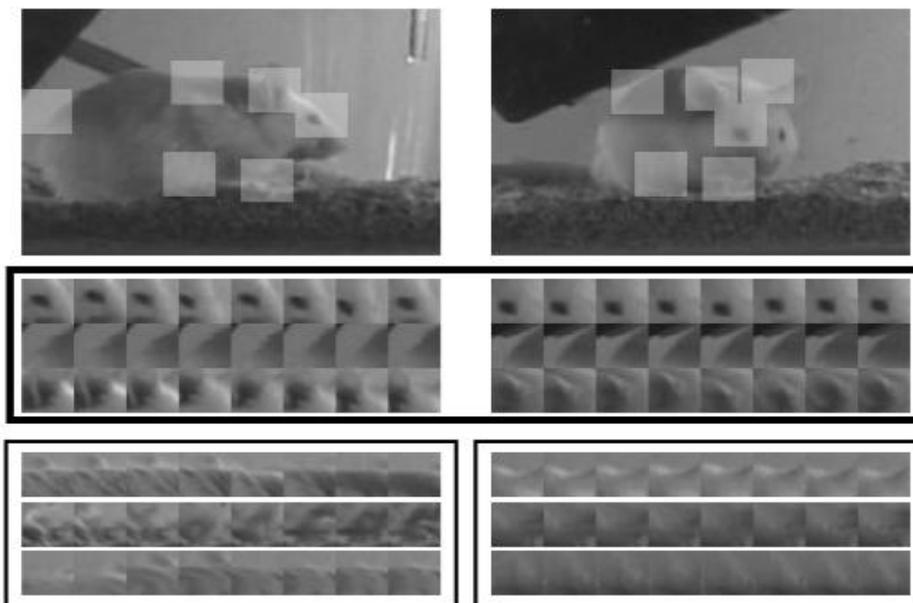
- **あとは, 多次元ベクトルの分類問題**
 - **最初の論文[Csu04]では, 以下の2つの手法で実験**
 - ・ SVM (support vector machine)
 - ・ Naive Bayes
 - ➡ **従来手法の結果を大きく改善**
従来の特徴量と異なり, 局所特徴の集合は
画像の「意味」を反映しているとみなせる。
- **例: Web画像の分類**
 - **10種類のキーワードについて, 平均適合率**
 - ・ 従来手法(領域分割 + GMM) : 73.5%
 - ・ BoK + SVM : 82.4%

[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.)

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

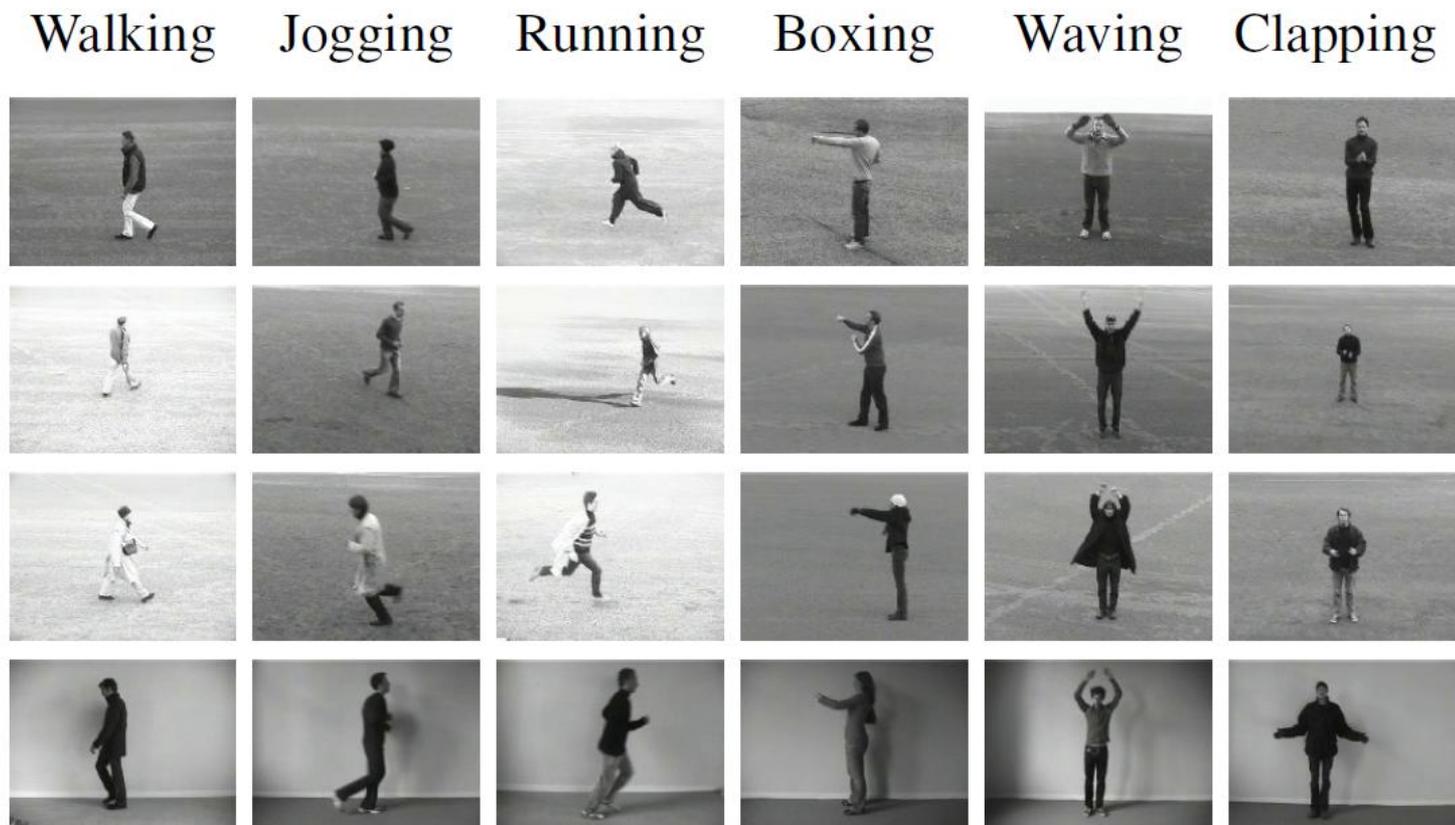


Figure 8. Sample frames from the KTH actions sequences. All six classes (columns) and scenarios (rows) are presented.

KTHデータセットの最新結果

Table 5. Compare to state-of-art for KTH dataset

Dállor et al. [3]	81.2%
Liu et al. [2]	91.8%
Gilbert et al. [9]	96.7%
Kim et al. [11]	95.33%
Uemura et al. [10]	93.7%
ours	94.5%

- 我々が実験中の結果は, **94.5%**
- 今までの最高は, **96.7%**

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.

教師信号あいの大量Web動画ショット ト分類

大量のWebショット

Runningの学習セット

分類器

1 位



2 位

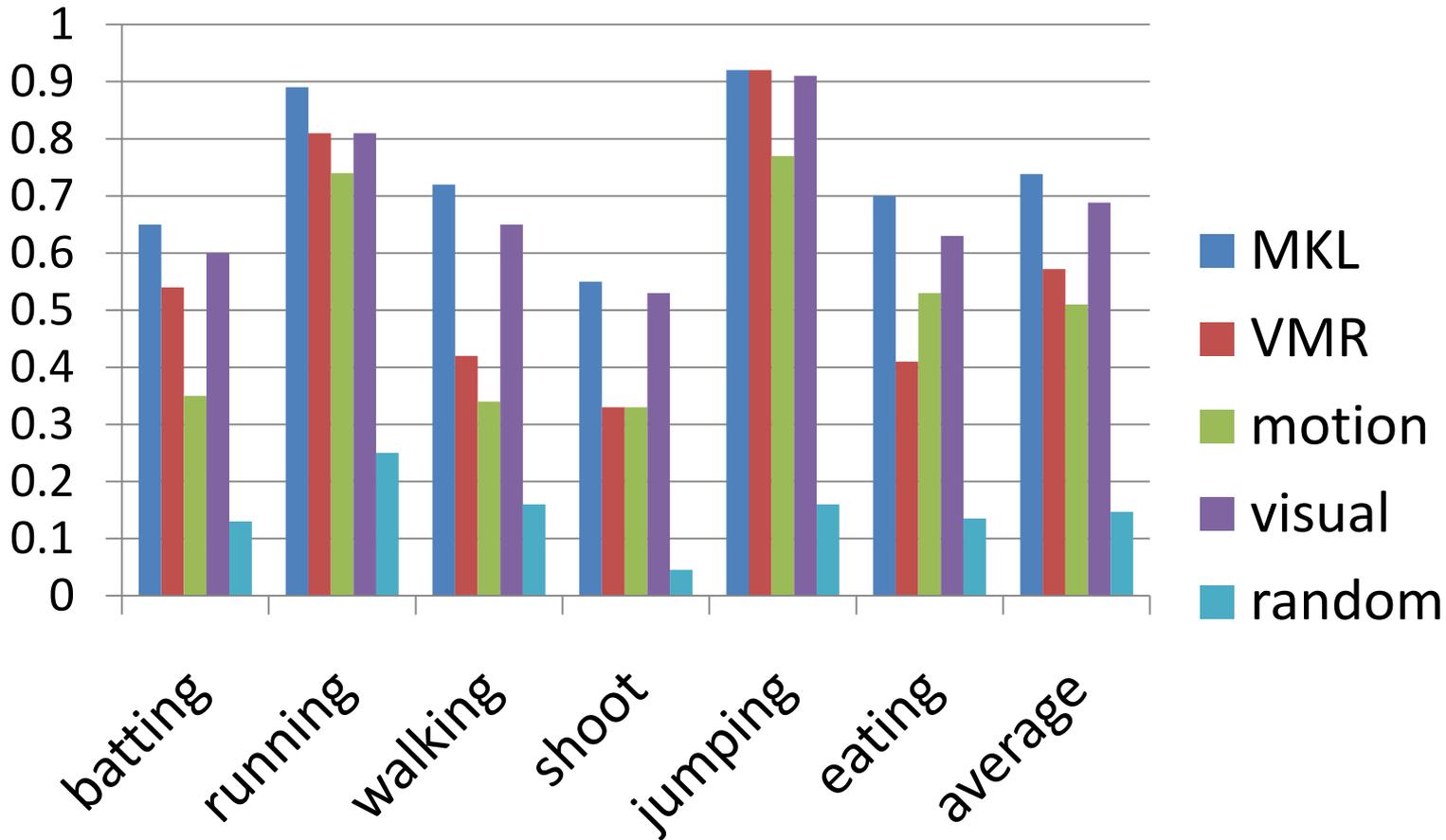


3 位



実験結果

■ 教師信号あいランキング付け



Web動画分類

- 教師信号あいランキング付け

結果のデモ

<http://img.cs.uec.ac.jp/noguchi-a/reject/index.html>

物体認識, 動作認識の組み合わせ

- 物体認識は「名詞」を認識
 - 物体名, シーン名
- 動作認識は「動詞」を認識
 - 主に 人間の動作



- 組み合わせれば, 「ラーメンを食べている人」
「運動会で走っている人」などが認識可能.
 - でも, 現時点では, 組み合わせの研究は
始まったばかりで, まだ多くの研究が必要.



3. 意味的映像検索, TRECVIDとは？

意味的映像検索とは？

大量の映像データに対して、言葉で表現されたシーンの検索を行うこと

- 「走行中の車のフロントガラス越しに見た路上」
- 「多数の人がいる屋外シーン」
- 「最上階まで見えているビル」
- 「電話で話している人」

TRECVIDという映像認識
ベンチマークでのタスク



TRECVID の 概要

- **ビデオ検索の研究を促進するための競争型国際ワークショップ。**
 - **米国の国立標準技術研究所(NIST)が主催**
 - **テキスト検索ワークショップTREC(Text Tetrieval Conference)から派生したもの**
 - **映像データがNISTより参加者に提供される**
 - **参加は無料。ただし、映像データを入手したら、結果を提出する必要がある。**
 - **4つのタスクから1つ以上を選んで参加。**
 - **参加者な結果を提出し、NISTが評価。全結果は、名前入りでWS及びWebで公表される。**

TRECVID 2009 の4つのタスク

1. **高次特徴抽出** (High-level feature extraction)
 - **一般物体認識, 動作認識のタスク**
2. **検索** (Search)
 - **最初のスライドの意味的映像検索**
3. **Event detection**
 - **監視カメラ映像に対して, eventを検出**
PersonRuns, CellToEars, Pointing, OpposingFlow,...
4. **Copy detection**
 - **同一映像が再利用されている箇所を検出**

本講演では, 1を中心に, 1, 2について扱う.

2009年の状況

	高次特徴	サーチ	Copy detecti on	Event detecti on
登録チー ム	71	48	51	45
結果提出 チーム	44	9/9 締切	20	9/23 締 切
	9/17 結 果発表			10/1結果 発表

合計で96チームが登録

高次特徴抽出は、日本からは NII, NHK, 旭化成, UEC.



TRECVID2009の実験データ

- オランダ語映像(ニュース, ドキュメンタリー, 教育番組など様々なジャンルを含む)
 - 学習データ : 106時間分 64GB
 - テストデータ: 280時間分 179GB
- 付随データ
 - 各ショット分割境界情報, ショットID
 - ・ 学習データ : 36,262 ショット
 - ・ テストデータ: 61,382 ショット
 - 各ショットと対応する音声認識テキスト
 - ・ オランダ語テキスト, 機械翻訳された英語テキスト
 - ・ 無音の映像もある。

それなりの計算設備が必要

クラスタ計算機
高速大容量ファイルサーバ
が必須



高次特徴抽出タスク

- 意味的検索のための要素となる「高次特徴」の抽出を行うタスク
 - 2009年は、20種類の特徴。
- 「特徴」と名が付いているが、一般物体認識、動作認識そのもの
- ショット映像、そのフレーム画像、動き、声、音声認識テキスト が利用可能。
- Generic method を用いること。対象毎に専用detectorを作ることには奨励されない。
(ブラウザで例を示します)

2009年 高次特徴(1)

001 Classroom

002 Chair

003 Infant

004 Traffic-intersection

005 Doorway

006 Airplane_flying

007 Person-playing-a-musical-instrument

008 Bus

009 Person-playing-soccer

010 Cityscape

2009年 高次特徴(2)

- 011 Person-riding-a-bicycle**
- 012 Telephone**
- 013 Person-eating**
- 014 Demonstration_Or_Protest**
- 015 Hand**
- 016 People-dancing**
- 017 Nighttime**
- 018 Boat_Ship**
- 019 Female-human-face-closeup**
- 020 Singing**

20種類の認識対象 (1)



Classroom : 学校の教室
(一人以上の学生が必ず現れる)

Airplane_flying : 飛行中の飛行機(ヘリコプター、ロケットを含まない)



Bridge : 橋

Two people : 二人
(人込みの一部を除く)



Emergency Vehicle : 緊急車両(軍用車両を除く)

Bus : バス
(バンとSUVを除く)



Dog : 犬(狼を除く)

Driver : 運転手



Kitchen : キッチン

Cityscape : 大都市の眺め(町レベルを除く)



20種類の認識対象 (2)



Harbor : 港

Mountain : 山



Telephone : 電話

Nighttime : 夜
(光の下で行われたスポーツイベントを除く)



Street : 街路(高速道路, 特殊道路を除く)

Boat Ship : 船



Demonstration Or Protest : デモ抗議活動

Flower : 花



Hand : 手(手はショットの主な部分でなければならぬ)

Singing : 一人以上が歌っているシーン



高次特徴抽出タスク(2)



- **使用データ: オランダ語ドキュメンタリー映像**
 - **学習データ : 106時間分 36,262 ショット**
 - **テストデータ: 280時間分 61,382 ショット**
- **学習データは参加者が共同でWeb上でアノテーションして作成.**
- **過去のTRECVIDの学習データ(2003~)も利用可能. LSCOM-439も利用可能.**
- **学習データに独自画像映像も利用可能. Web画像(Flickr)を用いるチームもある. (A:共通GT, B: +独自アノテーション, C: +独自映像)**

検索(search) タスク

■ 意味的映像検索そのもの.

■ 3つの方法がある.

FULLY AUTOMATIC:

完全自動



System takes query as input and produces result without any human intervention

MANUALLY-ASSISTED:

人手で質問解釈

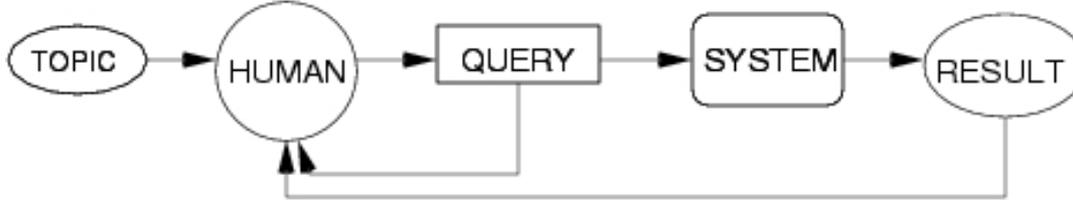


Human formulates query based on topic and query interface, not on knowledge of collection or search results

System takes query as input and produces result without further human intervention

INTERACTIVE:

インタラクティブ



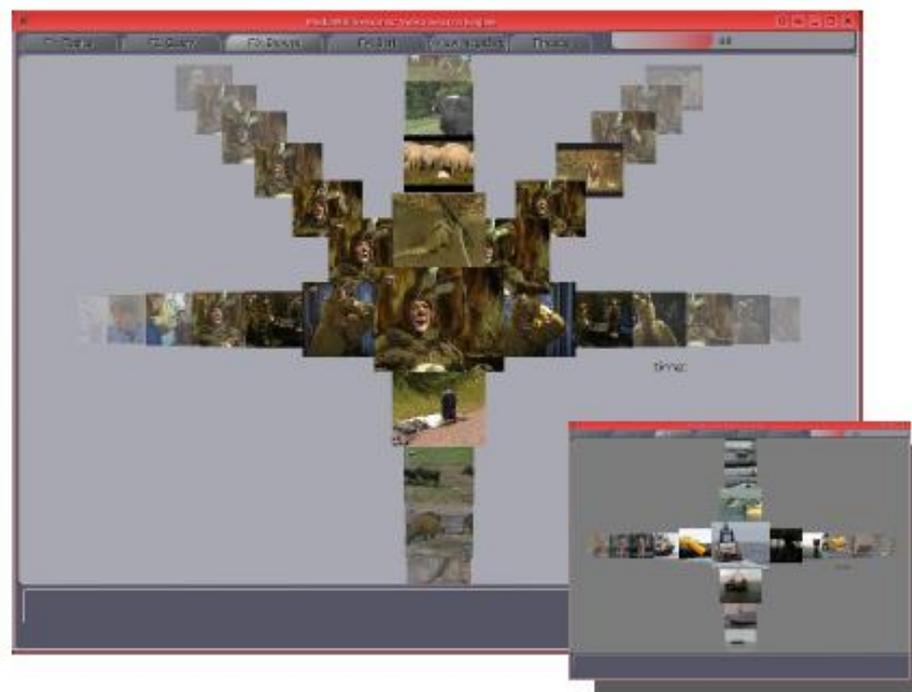
Human (re)formulates query based on topic, query interface, and/or results

System takes query as input and produces result without further human intervention on this invocation

■ 高次特徴の結果を利用.

検索(search) タスク (2)

- インタラクティブサーチは、サーチ時間も評価対象。
 - ユーザインタフェースの工夫も重要。



2009年のサーチトピック(全24)

- Find shots of a road taken from a moving vehicle through the front window
- Find shots of a crowd of people, outdoors, filling more than half of the frame area
- Find shots with a view of one or more tall buildings (more than 4 stories) and the top story visible
- Find shots of a person talking on a telephone
- Find shots of a closeup of a hand, writing, drawing, coloring, or painting

2009年のサーチトピック(全24)

- **Find shots of exactly two people sitting at a table**
- **Find shots of one or more people, each walking up one or more steps**
- **Find shots of one or more dogs, walking, running, or jumping**
- **Find shots of a person talking behind a microphone**

2009年のサーチトピック(全24)

- Find shots of a *building entrance*
- Find shots of *people shaking hands*
- Find shots of a *microscope*
- Find shots of *two more people, each singing and/or playing a musical instrument*
- Find shots of a *person pointing*
- Find shots of a *person playing a piano*

2009年のサーチトピック(全24)

- **Find shots of a street scene at night**
- **Find shots of printed, typed, or handwritten text, filling more than half of the frame area**
- **Find shots of something burning with flames visible**
- **Find shots of one or more people, each at a table or desk with a computer visible**

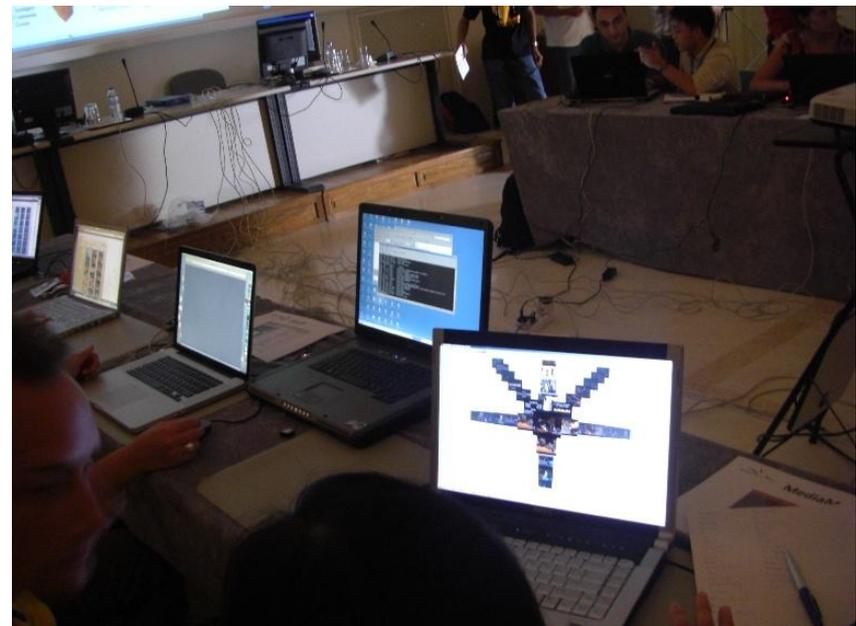
2009年のサーチトピック(全24)

- **Find shots of an airplane or helicopter on the ground, seen from outside**
- **Find shots of one or more people, each sitting in a chair, talking**
- **Find shots of one or more ships or boats, in the water**
- **Find shots of a train in motion, seen from outside**
- **Find shots with the camera zooming in on a person's face**



Video Olympic

- インタラクティブサーチタスクを参加者が1か所に集まって行うイベント
 - ACM Conf. on Image and Video Retrieval (CIVR) で開催. 今年は, Santorini, Greece



4. TRECVIDの方法:

多種類特徴抽出と

特徴統合手法

TRECVIDで用いられる主な方法

■ 高次特徴抽出

■ 多種類の特徴の抽出

■ Early fusion もしくは Late fusion で統合

- ・ **Early fusion** : すべての特徴をベクトル結合もしくはカーネルで統合して, SVMで学習
- ・ **Late fusion** : それぞれの特徴を単独でSVMで学習してから, SVM出力値を統合

■ 検索(サーチ)

- 検索トピックを抽出済みの高次特徴に対応付けし, 重み付きで高次特徴のランキングを統合.
自動の場合は, 辞書やWebで検索語拡張.

高次特徴抽出のための特徴量

- Bag-of-features: 数通りのサンプリング法
- Color histogram, Color moment
- Gabor特徴
- 顔特徴 (face detectorで検出された顔の数)
- 動き特徴 (optical flow histogram), 時空間特徴
- 音声テキスト特徴 (bag-of-words), 音声特徴(MFCC)

画像特徴はショットの代表フレームから抽出
1ショットから多数のフレーム画像を抽出する方が高精度であることが2008に示された。

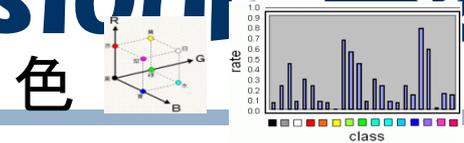
チームによって異なるが主な特徴は以上。

画像特徴のみ利用のチームも多い。上位チームでも。

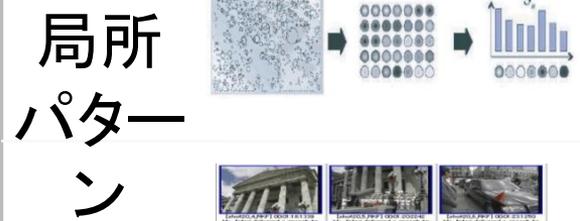
Mountain, buildingなど背景物体が多いので、

テキスト特徴や音声特徴は、役立たないことが多い。

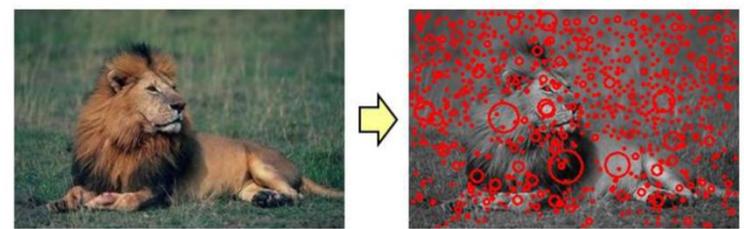
Late fusionによる特徴統合



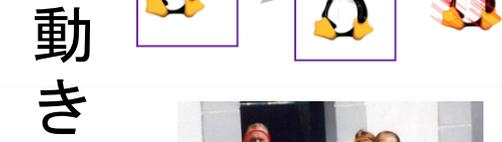
1. RGB各軸を4分割した64次元の正規化したカラーヒストグラム



SIFT特徴のBag-of-Keypoints表現



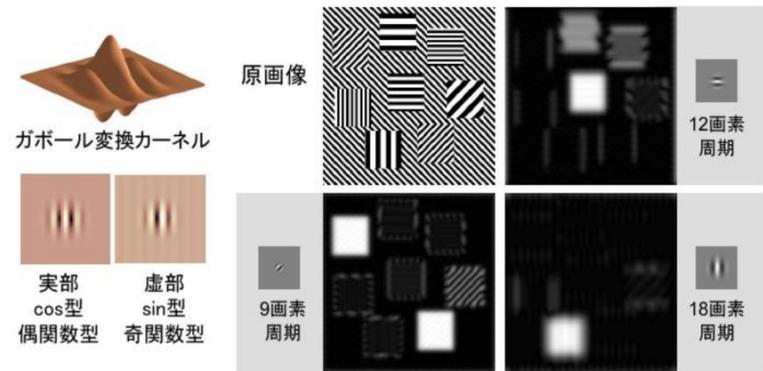
テキスト



顔



ヒストグラム



特徴抽出

モ

Late fusion による統合

- **AdaBoost**
 - ラウンドごとに重み付きデータでSVMを再学習し、誤分類されたデータに大きい重みをつける
- **AP-weighted fusion**
 - 各特徴単独のcross-validationで求めた平均適合率 (AP)を特徴の重みとして使用する手法
- **Max : 最大値をとる**
- **Average : 平均値で統合**



Early fusionによる特徴統合



Early fusionによる統合

- 単純なベクトル結合 → SVMで学習
- カーネル結合
 - 1つの特徴を1つのカーネルに割り当て、均一に統合して1つの統合カーネルを構築
 - ・ 和を取る方法, 積を取る方法がある.
- Multiple Kernel Learning
 - 複数のSVMのカーネルを統合するときに最適な重みを学習する手法

$$K_{combined}(\mathbf{x}, \mathbf{x}') = \sum_{j=1}^K \beta_j k_j(\mathbf{x}, \mathbf{x}') \quad \text{with } \beta_j \geq 0, \quad \sum_{j=1}^K \beta_j = 1.$$



重要な細かいテクニック

- SVM のカーネルはヒストグラム系特徴は
カイ2乗RBFカーネル

$$K(x, y) = \exp\left(-\gamma \sum_i \frac{(x_i - y_i)^2}{x_i + y_i}\right)$$

- パラメータ γ は、すべての学習データ間の
カイ2乗距離の平均の逆数がよいことが多い: **スケールの正規化**
 - SVMパラメータ(C , 場合によっては γ も)は,
cross-validation で **grid探索** して, **チューニング** する
- SVMの最適パラメータ探索が重要**

評価指標

■ 適合率(*precision*)

$$precision = \frac{\text{検索された適合画像の数}}{\text{検索結果の画像の数}}$$

■ 平均適合率(*AP*)

$$\text{平均適合率} = \frac{1}{N} \sum_{k=1}^N pre_k \quad N = 2000$$

■ 推定平均適合率(*infAP*)

- 平均適合率の推定値
- ランダムサンプリングで一部のテストデータのみを評価

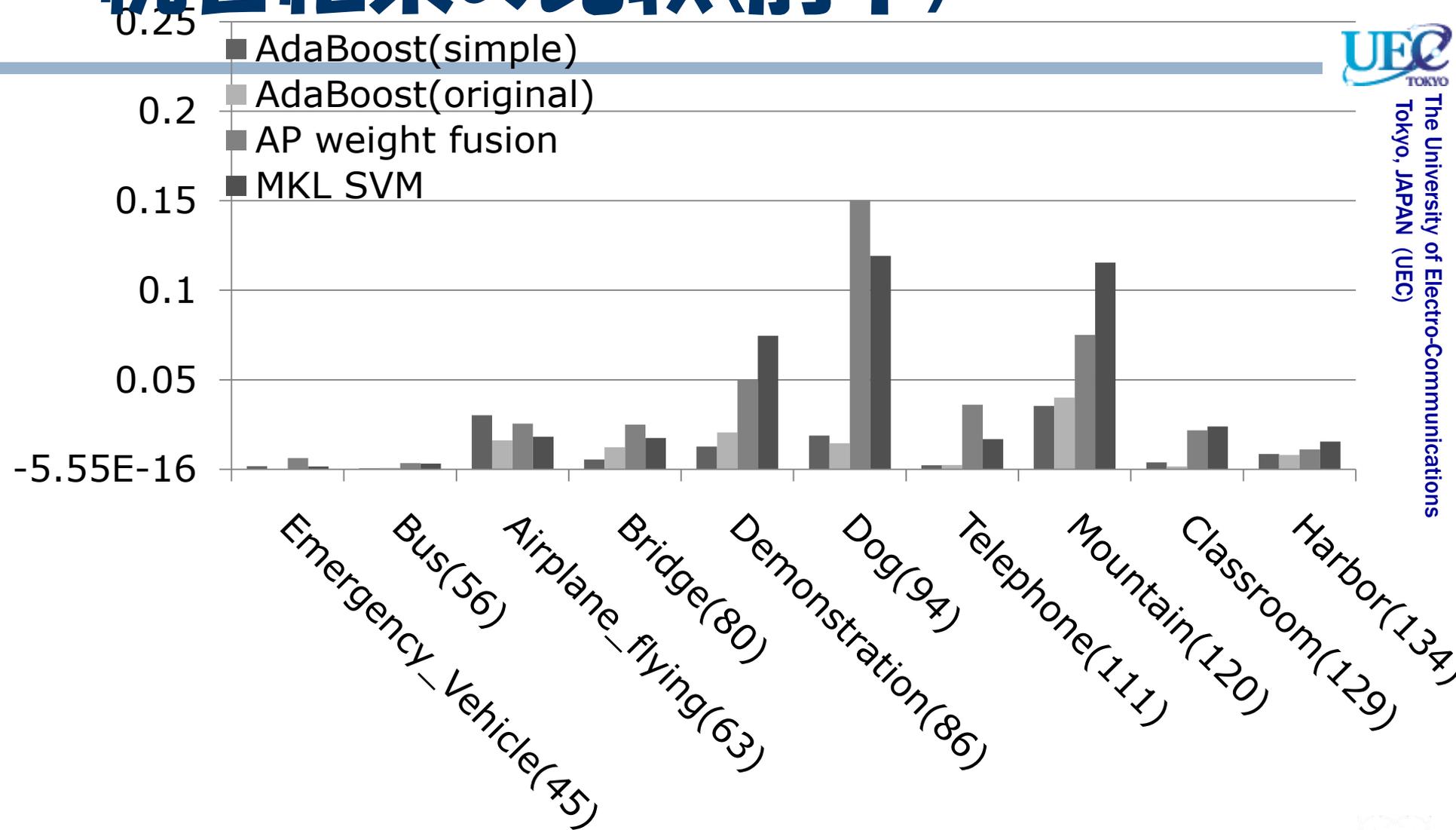
単独特徴の認識結果

特徴	infAP	特徴	infAP
カラーヒストグラム(全体)	0.0160	BoK_D_D	0.0157
カラーヒストグラム(分割)	0.0220	BoK_D_r	0.0129
カラーモーメント(全体)	0.0114	BoK_g_g	0.0125
カラーモーメント(分割)	0.0275	BoK_r_D	0.0108
face	0.0029	BoK_r_r	0.0097
motion	0.0009	gabor(全体)	0.0154
text	0.0018	gabor (分割)	0.0342

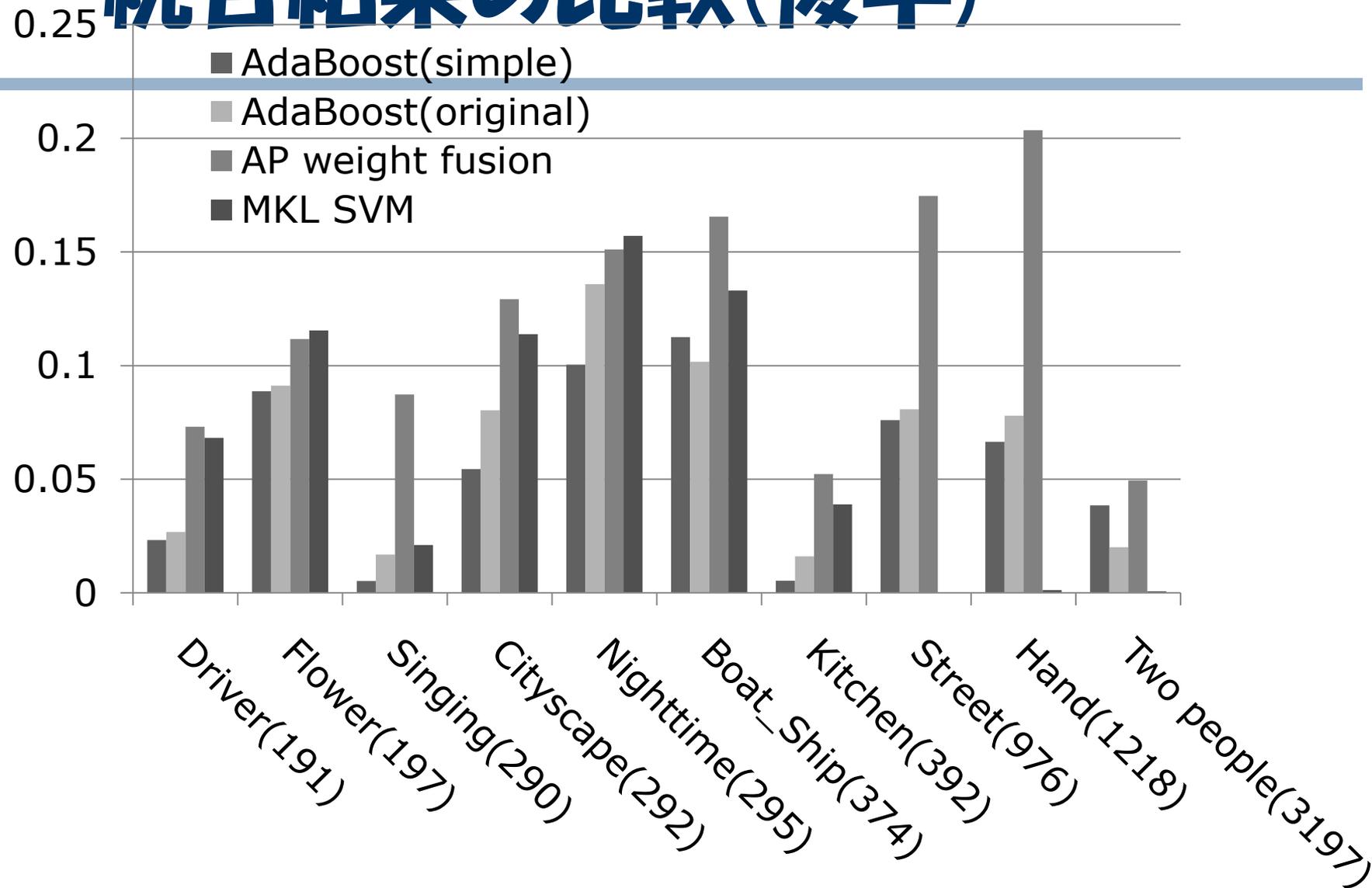
統合結果の比較

統合手法	4種類	10種類	14種類
AdaBoost			0.0319
AdaBoost	カラーヒストグラム(全体)	BoK_D_D	
AdaBoost	カラーヒストグラム(分割)	BoK_D_r	0.0500
AdaBoost	カラーモーメント(全体)	BoK_g_g	
AdaBoost	カラーモーメント(分割)	BoK_r_D	0.0801
AdaBoost	face	BoK_r_r	
AdaBoost	motion	gabor(全体)	-----
AdaBoost	text	gabor(分割)	

統合結果の比較(前半)



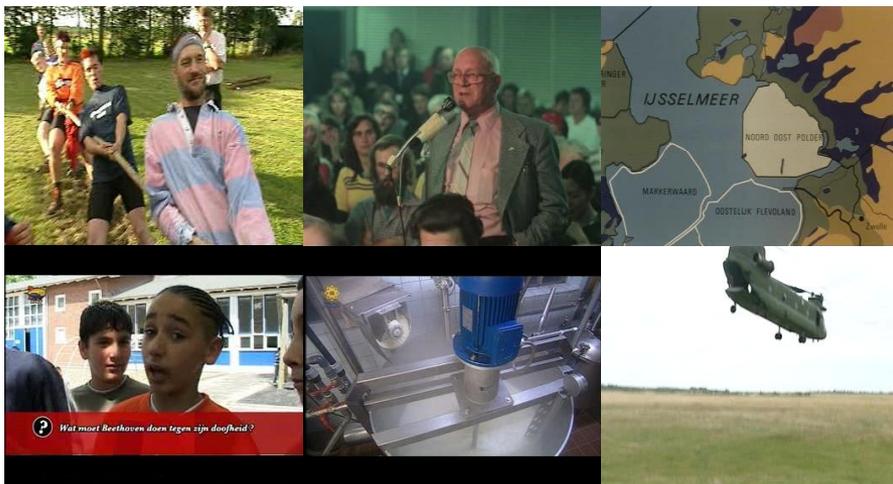
統合結果の比較(後半)



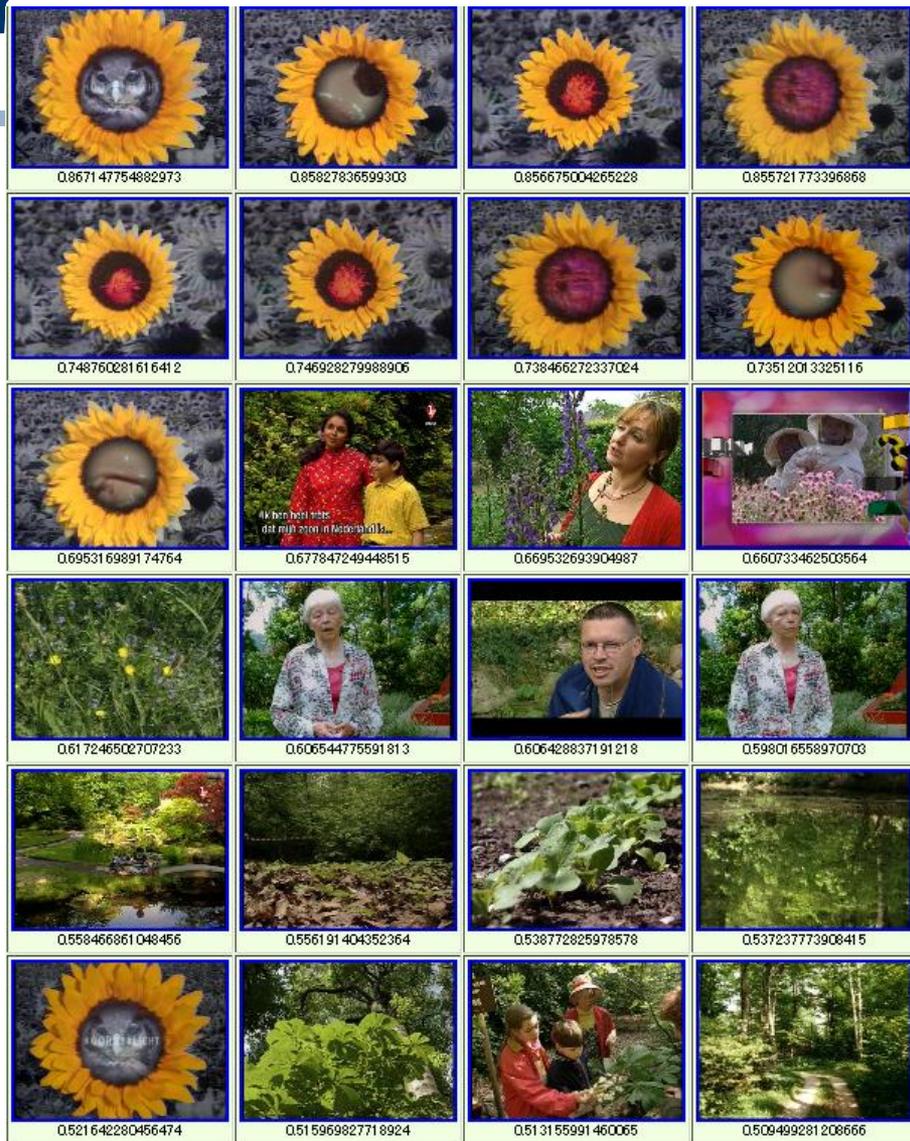
結果の例: Flower



Positive画像



Negative画像



結果

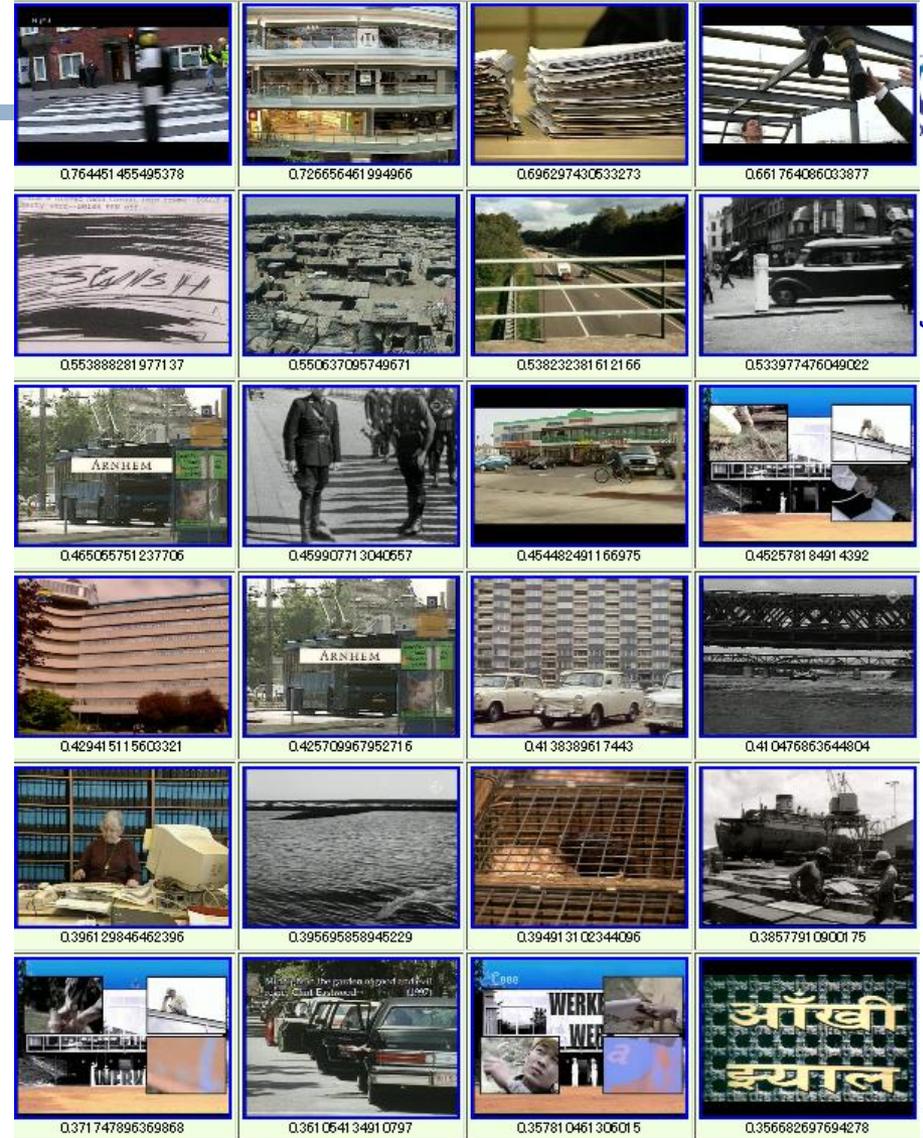
結果の例: Bus



Positive画像



Negative画像



結果

重み選択の例: Two people

fusion	カラーヒストグラム(全体)	カラーヒストグラム(分割)	face	BoK	infAP
AdaBoost (simple)	0	0.5240	0.4332	0.0428	0.0164
AdaBoost (original)	0	0	1.0000	0	0.0201
AP weighted	0.2504	0.2677	0.2542	0.2276	0.0289
MKL	1.0000	0	0	0	0.0007

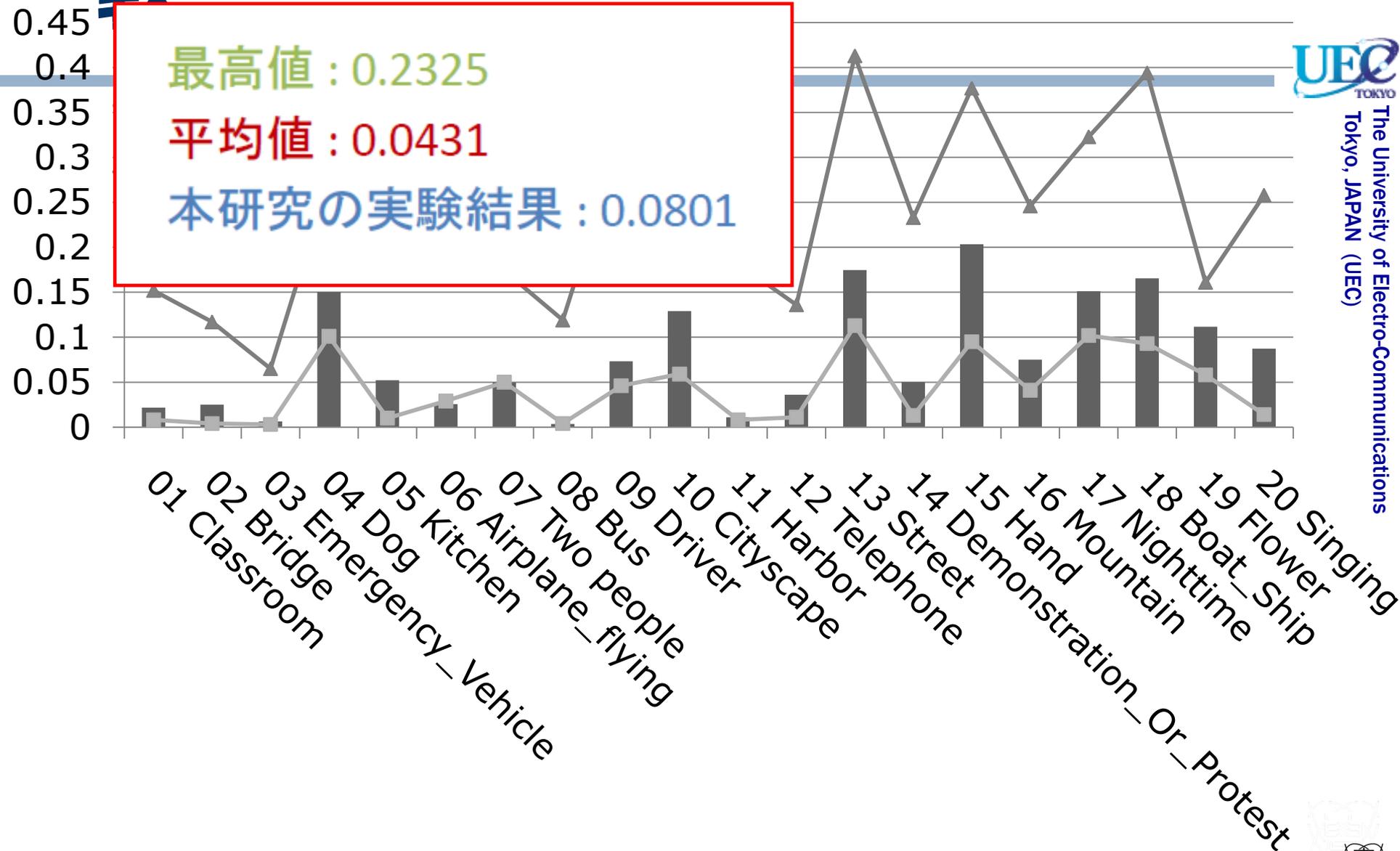


重み選択の例: Nighttime

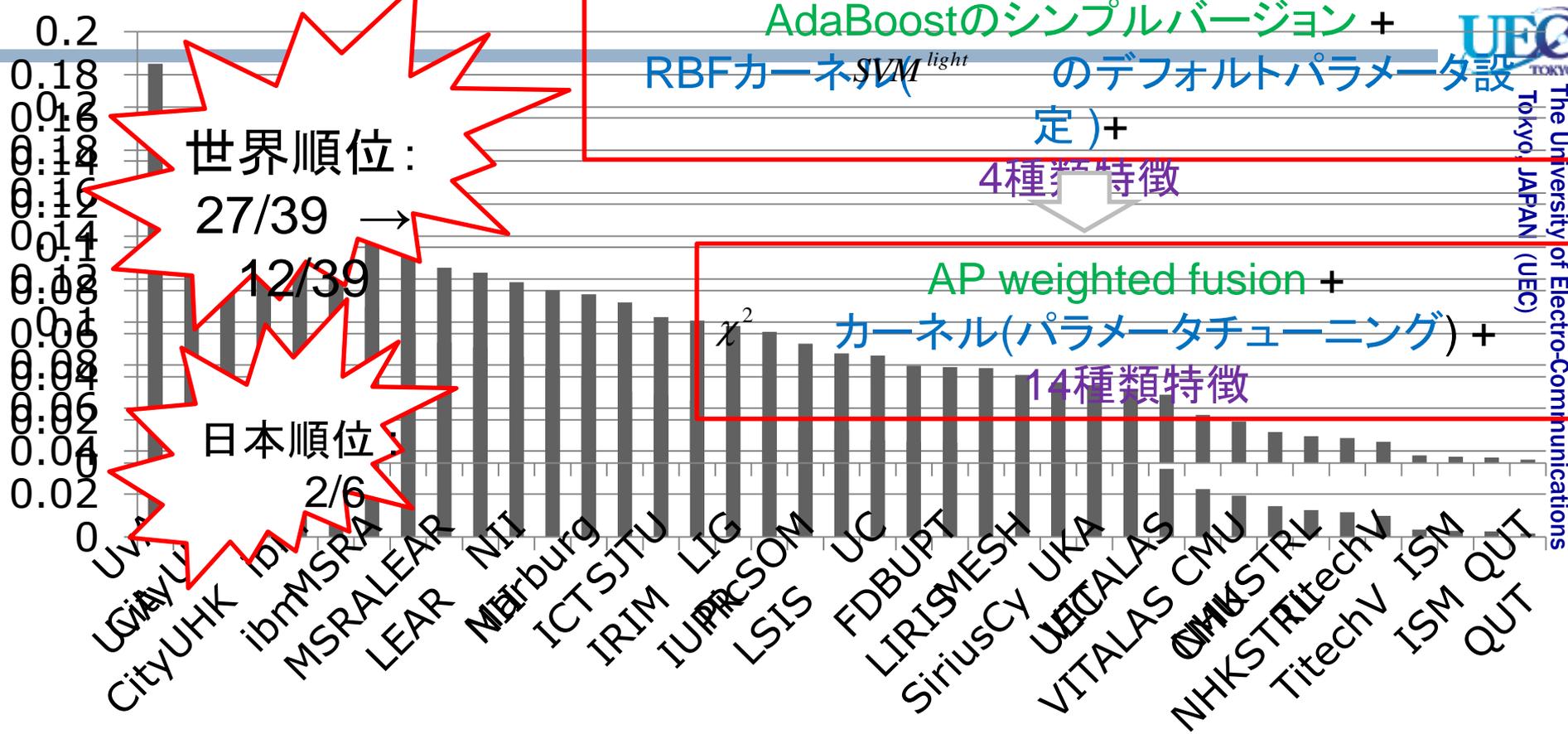
fusion	カラーヒストグラム(全体)	カラーヒストグラム(分割)	face	BoK	infAP
AdaBoost (simple)	0	0.9864	0	0.0146	0.1264
AdaBoost (original)	0	0.4658	0	0.5342	0.1316
AP weighted	0.2694	0.2699	0.2005	0.2603	0.1349
MKL	0	0.6854	0	0.3146	0.1184



参加チーム平均値と最高値との比較



改良による順位の上昇



順位

世界

1	アムステルダム大学
2	コロンビア大学
3	香港シティ大学
	...
1	電気通信大学
2	
	↑
2	電気通信大学
7	

日本

1	国立情報学研究所
2	電気通信大学
3	旭化成
4	NHK放送技術研究所
5	東京工業大学
6	統計数理研究所

2008年 高次・サーチ 上位10



高次特徴抽出 top 10

Run name (mean infAP)

CU_2_face_base_2 (0.167)
CU_4_fuse_baseline_4 (0.165)
CU_5_local_global_5 (0.162)
CU_6_local_only_6 (0.157)
CityUHK2_2 (0.156)
CityUHK1_1 (0.155)
REGIM1_1 (0.140)
PKU-ICST-HLFE-2_2 (0.138)
PKU-ICST-HLFE-4_4 (0.137)
PKU-ICST-HLFE-1_1 (0.134)

サーチ(自動) top 10

Run name (mean infAP)

A_2_MCG-ICT-CAS_1	0.067
B_2_UvA-MM-3_3	0.054
B_2_UvA-MM-4_4	0.053
A_2_MCG-ICT-CAS_2	0.053
B_1_UvA-MM-5_5	0.044
A_2_CityUHK2_2	0.042
A_2_MSRA.TV8_2	0.041
A_2_MSRA.TV8_1	0.041
A_2_CityUHK1_1	0.041
A_2_FD_IMI_SZC_1	0.040



2008年サーチタスク結果 上位10

サーチ(自動) top 10

Run name	(mean infAP)
A_2_MCG-ICT-CAS_1	0.067
B_2_UvA-MM-3_3	0.054
B_2_UvA-MM-4_4	0.053
A_2_MCG-ICT-CAS_2	0.053
B_1_UvA-MM-5_5	0.044
A_2_CityUHK2_2	0.042
A_2_MSRA.TV8_2	0.041
A_2_MSRA.TV8_1	0.041
A_2_CityUHK1_1	0.041
A_2_FD_IMI_SZC_1	0.040

サーチ(インタラクティブ) 10分以内 top 10

Run name	(mean infAP)
B 2 UvA-MM-1 1	0.194
B 2 UvA-MM-2 2	0.181
C 2 OXVGG_I_1 1	0.158
A 2 FX-CoHoMm 1	0.148
A 2 FX-CoxxSv 2	0.147
A 1 REGIM1 1	0.125
A 2 FX-AIHoMm 3	0.112
A 2 FX-AIHoSv 4	0.109
A 2 FX-AISeMm 5	0.100
C 2 thucrc.AIIMG 5	0.099
A 2 FX-AIxxMm 6	0.076
A 2 KSpace-E1 1	0.068



2009年 高次特徴結果

ブラウザを参照



5. まとめ

まとめ

■ 意味的映像検索

- TRECVIDのサーチタスクがまさにそれに相当
- 完全自動はまだ難しい。
- インタラクティブサーチはある程度使える。

■ 高次特徴抽出（一般物体認識，動作認識）

- Bag-of-features表現の提案，機械学習手法の進歩などにより，急速に発展
- BoFに加えて，音声テキスト，動き，従来からの color histogram をすべて統合して学習
- パラメータチューニングが重要。



おわり