

# 大規模画像認識のための局所特徴量の性能比較

本道 貴行<sup>†</sup> 黄瀬 浩一<sup>†</sup>

<sup>†</sup> 大阪府立大学大学院 工学研究科 〒 599-8531 大阪府堺市中区学園町 1-1

E-mail: †hondo@m.cs.osakafu-u.ac.jp, †kise@cs.osakafu-u.ac.jp

あらまし 画像を認識するための、様々な特徴点/領域の抽出手法や、局所記述子が提案されている。しかし、物体インスタンス(特定物体)を対象とした大規模画像認識において、どのような特徴点/領域を抽出し、どのような局所記述子を用いて、特徴量を記述することが最適なのは十分明確であるとは言い難い。そこで本稿では、1万画像データベースを用いた大規模画像認識において、複数の特徴点/領域抽出手法、局所記述子を比較した結果を報告する。我々の実験では、特徴点/領域単独では、Hessian Affine 領域、局所記述子としては GLOH や Shape Context、また、特徴点と局所記述子を一括で扱うもの場合には PCA-SIFT を用いることにより、最もよい結果が得られることが分かった。さらに、同一の記述子を用いても、特徴点/領域抽出の手法が異なれば、認識結果の誤りの傾向も異なり、特徴点/領域を組み合わせることにより精度が向上することが分かった。

キーワード 画像認識, 特徴点/領域, 局所記述子, Hessian-Affine 領域, GLOH, 大規模データベース

## Performance Evaluation of Local Features for Large-scale Image Recognition

Takayuki HONDO<sup>†</sup> and Koichi KISE<sup>†</sup>

<sup>†</sup> Graduate School of Engineering, Osaka Prefecture University

1-1 Gakuen-cho, Naka, Sakai, Osaka, 599-8531 Japan

E-mail: †hondo@m.cs.osakafu-u.ac.jp, †kise@cs.osakafu-u.ac.jp

**Abstract** Various interest points/regions detectors and local descriptors have been proposed for image recognition. However, it is still unclear which interest points/regions detector or local descriptor gives the best performance for large-scale image recognition of object instances(specific objects). In this report, we present the results of comparison about performance of image recognition with several interest points/regions detectors and local descriptors using a 10,000 image-database. In our experiment, we confirmed that it was the best to employ Hessian-Affine Region as the interest points/regions detector and GLOH/Shape Context as the local descriptor. And PCA-SIFT, providing the functions of both the detector and the descriptor, gives the best performance as the combinational method. In addition, we confirmed that the tendency of misrecognition is different even if the same local descriptor is employed, in case of employing interest regions computed with another detector, furthermore, combination of detectors gives higher accuracy.

**Key words** Image recognition, Interest point/region, local descriptor, Hessian-Affine Region, GLOH, Large-scale database

### 1. はじめに

近年、画像認識技術が脚光を浴びている。応用例としては、カメラ付き携帯電話で撮影された平面画像を認識することにより、その画像に関連するサービスへ誘導するもの<sup>(注1)</sup>や、車載

カメラ画像から道路交通標識を認識する研究 [1]、街路の店頭看板などを認識する研究 [2] などが行われている。

画像認識における実現課題の1つとして、どのような撮影条件下においても、高速、高精度で物体を認識できるようにすることが挙げられる。例えば、入力画像が歪んでいたりと、照

(注1): クレメンテックの M4 サービス [http://www.clementec.com/service\\_](http://www.clementec.com/service_)

明条件が変化していても、認識を可能にすることが望まれる。このような画像認識の課題を解決する手法の一つとして、局所記述子 (local descriptor) によって得られる特徴ベクトルを用いて画像を表現し、特徴ベクトル同士を比較し、検索・識別をするものがある。SIFT(Scale-Invariant Feature Transform) [3] や PCA-SIFT [4] は局所記述子による特徴ベクトル抽出法の一例である。この手法には、特徴ベクトルが局所的に得られるため、隠れや変動に対して比較的ロバストであるという利点がある。現在では、SIFT の他にも、SIFT を改良した GLOH(Gradient Location and Orientation Histogram) や、高速化を図った SURF(Speeded Up Robust Features) [5] を始めとする複数の局所記述子が提唱されている。

局所記述子を用いて画像の局所特徴量を記述する場合、記述する際に、その局所特徴量を記述するための特徴点/領域 (interest point/region) を与える必要がある。一般に、認識したい入力画像について、被写体の撮影角度がデータベースに登録されているものと全く同一の方向から撮影されるとは限らない。そこで、特徴点/領域抽出には、スケール変化や回転、角度変化などのアフィン変換に耐性をもつ手法を用いる。具体的には、Harris-Affine detector, Hessian-Affine detector [6] などである。これらの特徴点/領域に、局所特徴量を記述することにより、入力画像に対して、スケール変化やアフィン変換が起こっても、高精度で認識をすることが可能となる。

Mikolajczyk らは、画像認識のための特徴点/領域の抽出手法と、11 種の局所記述子を比較、検証した [7] [8]。この際、回転、視点、照明など、6 種の変化を加えた画像を用いて、認識の精度を検証しているが、検証に用いている画像の数は、48 枚であり小規模にとどまっている。識別対象の画像を大規模化したとき、特徴点/領域の抽出方法、並びに局所記述子の優劣がどのように変化するのは不明である。

そこで本稿では、1 万枚の画像データベースで、視点を様々に変化した特定物体の平面画像を撮影したものに対して、認識精度の検証を行う。検証には、複数のアフィン不変の特徴点/領域抽出手法、複数の局所記述子を用いる。実験の結果、特徴点/領域単独で比較した場合、Hessian-Affine 領域、局所記述子単独で比較した場合、GLOH や Shape Context [9] が最も高速で高精度の結果が得られることが分かった。また、特徴点/領域、局所特徴量記述の双方を組み合わせる場合、PCA-SIFT が最も高精度な結果であることが分かった。

さらに、同一の画像に、同一の記述子を用いて局所特徴量の記述を行った場合でも、記述に用いる特徴点/領域抽出手法が異なれば、画像の誤認識の傾向が異なり、組み合わせることにより、認識精度が向上することが分かった。

## 2. 局所記述子を用いた画像認識

識別する対象物体が写っている入力画像が与えられたとき、画像が多数収められているデータベースから、特徴ベクトルを用いて、対応する画像を検索することを考える。

この手法では、図 1 に示すように、データベースへの登録希望画像から、それぞれ特徴ベクトルを抽出しておき、それを画

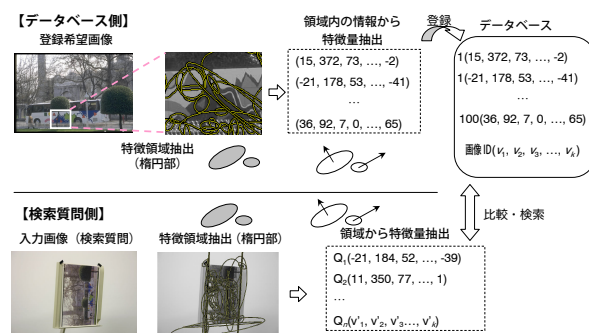


図 1 局所記述子を用いた画像認識の概要

像 ID と共にデータベースに登録しておく。認識時には、認識したい入力画像 (以後、検索質問と呼ぶ) から特徴ベクトルを抽出し、これをデータベースに登録してある特徴ベクトル群と比較・照合して認識を行う。特徴ベクトルを抽出する手順は以下の通りである。

- (1) 対象画像上で、複数の特徴点/領域を抽出し、選択する。
- (2) 画像上に複数決定された領域の画像情報を、それぞれ局所記述子を用いて特徴ベクトルで表現する。

特徴ベクトルの検索は、検索質問から抽出した特徴ベクトル  $Q_i$  ごとに、データベース中の特徴ベクトルの内で、ベクトルの距離が近いものを検索する。そして、該当する特徴ベクトルに対応する画像 ID に投票することで、対応画像を判定する。

本稿では、画像認識のタスクを「特徴点/領域抽出」「特徴量記述」「検索」の 3 ステップに分割する。そして、3 節で「特徴点/領域抽出」「特徴量記述」について述べ、4 節で「検索」について述べる。

## 3. 比較する特徴点/領域と記述子の概略

本節では、本稿で比較する「特徴点/領域抽出法」「特徴量記述法」について述べる。特徴ベクトル抽出手法には、

- (1) 「特徴点/領域抽出」「特徴量記述」をそれぞれ独立にして扱う手法
- (2) 「特徴点/領域抽出」「特徴量記述」を組み合わせ、一括して行う手法

がある。(1) では、「特徴点/領域抽出処理」と「特徴量記述」の処理が独立しているため、それぞれを別々の手法として考える。「特徴点/領域抽出法」「特徴量記述」にはそれぞれ複数の手法が存在するため、それらを任意に組み合わせることが可能である。(2) では、「特徴点/抽出処理」と「特徴量記述」を組み合わせ、1 つにまとめているため、「特徴点/抽出処理」と「特徴量記述」の組合せで 1 つの手法である。

3.1 節で (1) の手法の「特徴点/領域抽出」、3.2 節で (1) の「特徴量記述」について述べ、3.3 節で、(2) の手法について述べる。

### 3.1 特徴点/領域抽出

特徴点/領域抽出では、画像から、特徴量記述に必要な領域を局所的に複数抽出する。本稿では、アフィン変換に耐性のある 4 つの特徴点/領域抽出手法を用いる。これにより、被写

体が真正面からではなく、斜めから撮影された場合においても、高精度で認識をすることが可能となる。本稿で用いる特徴点/領域抽出手法は、文献 [7] [8] と同じく Harris-Affine 領域, Hessian-Affine 領域, MSER(Maximally Stable Extremal Regions [10], IBR(Intensity extrema based Regions) である。これらの手法を用いて、アフィン変換に耐性を持つような楕円領域を抽出する。

以下では、Harris-Affine 領域, Hessian-Affine 領域, MSER について、概略を説明する。

#### (1) Harris-Affine 領域, Hessian-Affine 領域

座標  $(x, y)$  の画素の値を  $I(x, y)$  とし,  $x, y$  を連続変数とみなすと,  $I(x, y)$  は曲面を描く。基本的な考え方は、この曲面の点の値が極値となるような点を、微分によって求め、特徴点とするものである。

Harris-Affine 領域を抽出する場合、まず Harris detector で特徴点を求める。画像の各画素について、以下の式 (1) の行列  $M$  を求める。

$$M = G(\sigma) * \begin{bmatrix} I_x^2(u, v) & I_x I_y(u, v) \\ I_x I_y(u, v) & I_y^2(u, v) \end{bmatrix} \quad (1)$$

ここで、 $G(\sigma)$  は、標準偏差  $\sigma$  のガウス関数、 $I(u, v)$  は注目画素  $(u, v)$  における画素の濃度値、 $I_x, I_y$  はそれぞれ  $I$  の  $x$  方向、 $y$  方向の 1 階微分である。また、 $*$  は畳み込みを表す。

この行列  $M$  の 2 つの固有値を、 $\lambda_1, \lambda_2$  とすると、共に所定の閾値より大きければコーナ点、片方が所定の閾値より大きければエッジ、共に所定の閾値より小さければ平面となる。

ここで、この判定のため、行列  $M$  の行列式とトレースを用いて以下の式 (2) のように定義される  $R$  を求める。

$$R = \det M - k \cdot (\text{trace} M)^2 = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 \quad (2)$$

$k$  は定数で、実験的に 0.04~0.06 である。 $R$  が大であれば、コーナ点である。そこで、 $R$  が、設定された閾値よりも大きければ、その点を特徴点と見なす。本稿では、この値として、1000 を用いる。

次に、求めた点(注目画素)に対して、周囲の画素の輝度勾配を計算し、注目画素を中心とする楕円領域を求める。

Hessian-Affine 領域を抽出する場合、Harris-Affine 領域の抽出と同じように、まず、Hessian Detector で特徴点を求める。画像の各画素について、以下の式 (3) の行列  $H$  を求める。

$$H = G(\sigma) * \begin{bmatrix} I_{xx}(u, v) & I_{xy}(u, v) \\ I_{xy}(u, v) & I_{yy}(u, v) \end{bmatrix} \quad (3)$$

ここで、 $I_{xx}$  は  $I$  の  $x$  方向の 2 階微分、 $I_{xy}$  は  $I_x$  を  $y$  方向に微分したものである。今、

$$\det H = \sigma^2 (I_{xx} I_{yy}(u, v) - I_{xy}^2(u, v))$$

$$\text{trace} H = \sigma (I_{xx}(u, v) + I_{yy}(u, v))$$

を考え、これらの値が共に極値となるような点を特徴点とする。Hessian detector は、Blob 検出に用いられる。求めた点(注

目画素)に対して、周囲の画素の輝度勾配を計算し、注目画素を中心とする楕円領域を求める。

#### (2) MSER 領域

MSER の基本的な考え方は、グレースケール画像で画素値に基づいた画像の領域分割を行い、分割された領域を特徴領域とするものである。画素値が類似している近傍領域を連結し、領域を得る。一般に、MSER で得られる領域は本来楕円領域ではないため、アフィン変換に対する耐性が保存されるようにこの領域を楕円に変換する。

#### 3.2 特徴量記述

特徴量記述では、前節で求めた特徴領域から、特徴量を抽出し、特徴ベクトルとして記述する。本稿では、文献 [8] と同じく、Cross Correlation, Complex Filters, GLOH, Steerable Filters, Moments, PCA-SIFT, Shape Context, SIFT, Spin Image の 9 種類の局所記述子を用いる。

以下では、そのうち、SIFT, PCA-SIFT<sup>(注2)</sup>, GLOH, Shape Context の概略を説明する。

#### (1) SIFT

SIFT は、照明、ノイズ、回転などにロバストな記述子で、高い識別性能を持つ。まず、グレースケール画像において、局所正方形領域の輝度変化の大きさ、変化方向のヒストグラムを作成する。そして、それを  $4 \times 4$  の格子領域で 8 方向に量子化する。これにより、128 次元の特徴量として記述される。

#### (2) PCA-SIFT

PCA-SIFT は、 $39 \times 39$  の正方形領域の画素に対して、画像の輝度変化の大きさを  $x, y$  方向で計算し、3042 次元の特徴ベクトルを得る。それに対して、主成分分析した結果を射影し、36 次元の特徴量を得る。

#### (3) GLOH

GLOH は、SIFT のロバスト性、識別性を向上させたものである。2 次元座標を対数極座標 (Log-Polar) に変換して計算する。まず、半径方向を 3 領域に分割、角度方向を 8 方向に分割する。ただし、半径方向において、特徴点に最も近い部分は角度方向を分割しない。この結果、領域は 17 領域に分割される。この 17 領域のそれぞれに対して、16 方向に量子化した方向の輝度変化の大きさヒストグラムを作成する。その結果、272 次元の特徴量が得られる。この 272 次元の特徴量に主成分分析した結果を射影し、128 次元の特徴量を得る。

#### (4) Shape Context

Shape Context は、SIFT と類似した記述子である。まず、GLOH のように、2 次元座標を対数極座標に変換し、半径方向を 3 領域、角度方向を 4 方向に分割する。ただし、半径方向において、特徴点に最も近い部分は角度方向を分割しない。この結果、領域は 9 領域に分割される。そして、方向を 4 つに量子化した上で、輝度変化の大きさヒストグラムを作成し、36 次元の特徴量を得る。

(注2): 本来の SIFT, PCA-SIFT は、特徴点/領域抽出、特徴量記述を共に行う手法であるが、本節では、特徴量記述のみを用いている。

### 3.3 特徴点/領域抽出, 特徴量記述の組合せ

特徴点/領域抽出と特徴点記述を一括して扱うものも存在する。本稿では, SIFT, PCA-SIFT, SURF を用いる。なお, 3.2 節の SIFT, PCA-SIFT との混同を避けるため, 以後, 組合せの記述子には記述子名の最後に ' を付与して区別をすることにする。具体的には, 本節が指す SIFT, PCA-SIFT を, それぞれ, SIFT', PCA-SIFT' と表記する。

## 4. 検索方法

2 節で述べたように, 検索質問から抽出した特徴ベクトル  $Q_i$  が, データベース中にある特徴ベクトルのいずれに近いかを検索し, データベース中の対応したベクトルを持つ画像に投票処理を行う。その結果, 最大得票数を得た画像を認識結果とする。ベクトル同士の距離計算にはユークリッド距離を用いる。

投票処理によって認識を行うため, データベースの画像が得た票数によって認識結果が決まる。つまり, 正解の画像が最大得票数となっていれば, 特徴ベクトル  $Q_i$  の照合時に多少の誤りが発生しても問題がないといえる。そこで, 本稿では, ベクトルの照合に ANN (Approximate Nearest Neighbor) [11] を用いる。近似を行うことにより, ベクトル照合の精度は低下するものの, 検索にかかる処理時間を削減することが可能となる。

## 5. 実験

### 5.1 実験条件

実験には以下に述べる画像データベース, 検索質問画像を用いた。以下に示す処理時間は, 検索質問画像 1 枚あたりの認識に要した時間を表す。ただし, 特徴点領域抽出, 特徴量記述に必要な時間は含まれていない。特徴点/領域抽出, 特徴量記述には, Affine Covariant Features のサイト<sup>(注3)</sup>, SIFT のサイト, PCA-SIFT のサイト<sup>(注4)</sup>, SURF のサイト<sup>(注5)</sup>で提供されているものを用いた。検索には ANN のサイト<sup>(注6)</sup>で提供されているものを用いた。また, 使用計算機は, CPU が AMD Opteron 2.8GHz, メモリ 64GB のものである。

#### 5.1.1 画像データベース

実験に用いた画像について説明する。まず, Google のイメージ検索, PCA-SIFT のサイト, Flickr のサイトなどから収集した画像 10,000 枚を用いた。Google 画像検索の検索キーワードとしては, “ポスター”, “雑誌”, “表紙” などを用いた。画像のサイズはおおよそ VGA サイズである。

次に, 3 節冒頭で述べた, 以下の 2 通りの方法でデータベースを作成した。

(1) 「特徴点/領域抽出」, 「特徴量記述」をそれぞれ独立にして扱う手法

(2) 「特徴点/領域抽出」, 「特徴量記述」を組み合わせ, 一括して行う手法

まず, (1) について説明する。3.1 節で述べた 4 つの手法で

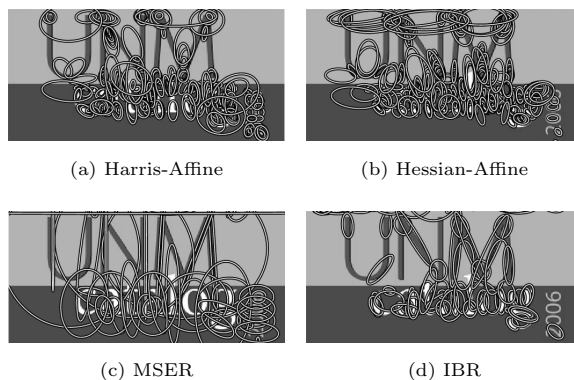


図 2 抽出領域の例

表 1 各手法での平均特徴点/領域数 (DB)

	Harris-Affine	Hessian-Affine	MSER	IBR
平均特徴点数	702	679	354	263

表 2 各手法での平均特徴点/領域数 (DB)

	SIFT'	PCA-SIFT'	SURF'
平均特徴点数	2034	2069	571.5

表 3 各記述子の次元数

記述子	次元	記述子	次元
SIFT	128	Complex Filters	15
GLOH	128	Moment Invariants	20
Shape Contexts	36	Cross-Correlation	81
PCA-SIFT	36	SURF	64+1
Spin Images	50	SURF128	128+1
Steerable Filters	14		

特徴点/領域を各々抽出した。各手法で抽出した楕円領域の例を図 2 に示す。なお, 特徴点抽出の各手法で求めた, 画像の平均特徴点/領域数は, 表 1 の通りである。

そして, 得られた特徴点/領域に対して, 3.2 節で述べた 9 種類の記述子でそれぞれ特徴量記述を行い, データベースを作成した。この結果 (領域 4 種類) × (特徴量 9 種類) で, 36 通りのデータベースが作成された。

次に (2) について説明する。SIFT', PCA-SIFT', SURF', SURF128' の 4 つのデータベースを作成した。ここで, SURF128' とは, SURF' の拡張版であり, 次元を 128+1<sup>(注7)</sup>としたものである。

1 万の画像から抽出された平均特徴点/領域数は, 表 2 の通りである。また, 今回の実験で用いた各記述子が持つ次元数は表 3 の通りである。

#### 5.1.2 検索質問画像

検索質問として, 次の手順で作成した画像 2,000 枚を用いた。まず, 画像データベース内に含まれる画像 500 枚を無作為に選択し, A4 の紙面に印刷した。次に, カメラを用いて印刷した

(注3): <http://www.robots.ox.ac.uk/~vgg/research/affine/>

(注4): <http://www.cs.cmu.edu/~yke/pcasift/>

(注5): <http://www.vision.ee.ethz.ch/~surf/>

(注6): <http://www.cs.umd.edu/~mount/ANN/>

(注7): SURF', SURF128' では, 64, 128 次元の特徴ベクトルの他にラブラシアン符号を判定するための情報 (1 又は -1) を使用しているため, それを +1 と記した。

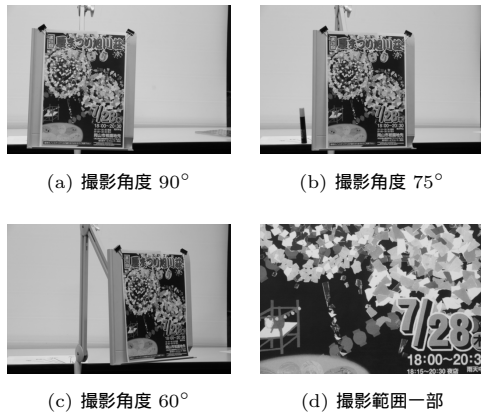


図 3 検索質問の例

表 4 特徴領域抽出に要した平均時間 [sec]

特徴領域	平均時間	特徴領域	平均時間
Harris-Affine	0.57	Hessian-Affine	0.35

表 5 特徴量記述に要した平均時間 [sec]

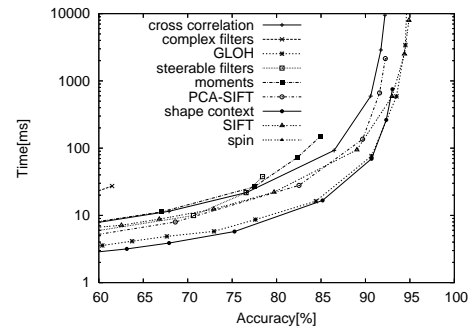
記述子	平均時間	記述子	平均時間
SIFT	0.83	Complex Filters	0.71
GLOH	0.90	Moment Invariants	0.73
Shape Contexts	0.97	Cross-Correlation	0.67
PCA-SIFT	0.91	Steerable Filters	0.71
Spin Images	0.69		

紙面を撮影した。撮影した画像の例を図 3 に示す。図に示す通り、紙面全体が写る配置で、紙面に対するカメラの光軸の角度  $\theta$  を  $90^\circ, 75^\circ, 60^\circ$  に変化させた。また角度を  $90^\circ$  として、紙面の一部を撮影した。これにより、1 枚の紙面に対して、合計 4 通りの画像を得た。さらに、データベースと同様に、(1), (2) の 2 通りで、同じように特徴ベクトルを求めた。画像 1 枚に対して、特徴点/領域抽出、特徴量記述に要した時間の平均をそれぞれ、表 4, 表 5 に示す。

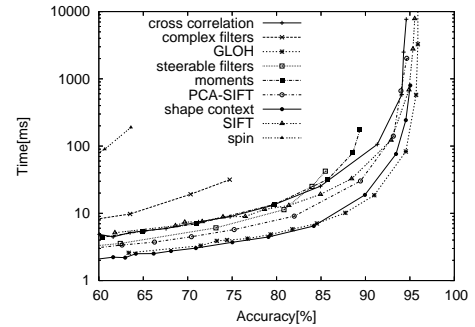
### 5.2 特徴点/領域抽出、記述子の比較

前節 (1) のデータベースと検索質問に対して、ANN の許容誤差パラメータ  $\varepsilon$  を 2, 3, 5, 10, 20, 30, ..., 90, 100, 200 に変化させた時の検索時間と認識率を調べた。その結果を図 4 に示す。縦軸が処理時間 [ms]、横軸が認識率 [%] である。右にプロットされているものほど認識率が高く、下にプロットされているものほど処理時間が短い。そのため、右下にプロットされているものほど優れているといえる。特徴点抽出では、Hessian-Affine, Harris-Affine, MSER の順でよかった。IBR では、他のものより結果が悪かった。以上より、特徴点の取り方や特徴点の数によって、認識結果が変化することが分かった。記述子では、Shape Context, GLOH の結果が概ね最も良好であった。

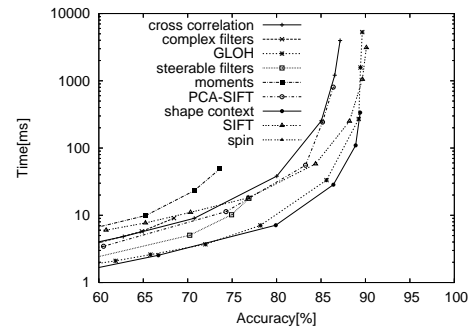
ここで、好結果を示した、Harris-Affine, Hessian-Affine, MSER を特徴点/領域として採用した場合において、誤認識の傾向がどのように異なるか調べた。その結果、Harris-Affine と Hessian-Affine では、誤認識の傾向が比較的類似していることが分かった。一方、MSER では、Harris-Affine, Hessian-Affine



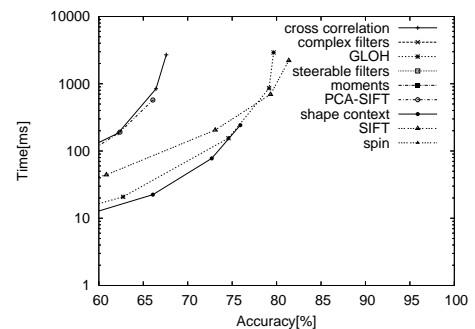
(a) Harris-Affine



(b) Hessian-Affine



(c) MSER



(d) IBR

図 4 各領域での記述子間の精度と検索時間の比較

と誤認識の傾向がやや異なっていた。誤認識した対応画像の得票数を調べてみると、得票数が 1 や 2 と非常に少ないことが分かった。従って、複数の手法で抽出した特徴領域を組み合わせることにより、判定の為の票数を増やすことが出来るため、より認識率を高めることができる可能性があるといえる。

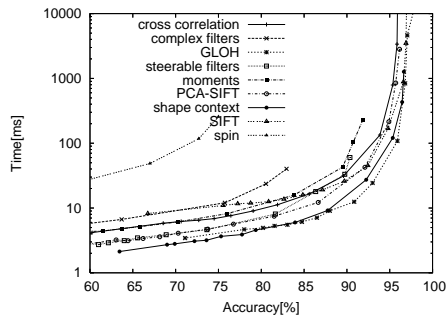


図 5 Hessian-Affine 領域, MSER を組み合わせた場合の結果

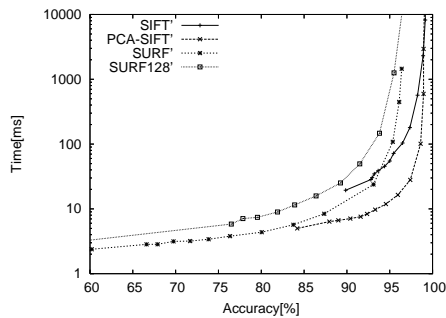


図 6 各手法の精度と検索時間の比較

### 5.3 特徴点/領域を組み合わせた場合

前節で述べたことを踏まえ, Hessian-Affine と MSER の領域を組み合わせ, 結果の比較を行った. 具体的には, Hessian-Affine と MSER の領域に対して特徴量記述を行い, 個別に, それぞれ ANN で検索する. その結果, Hessian-Affine, MSER でそれぞれ, 画像別に得票数が得られる. そして, これらの得票数を最後に足しあわせ, 最大となったものを認識結果とするものである. 結果を図 5 に示す. 図 4(b), (c), すなわち, Hessian-Affine や MSER で抽出した領域 1 つのみを用いる場合と比較すると, 認識率が向上していることが分かる.

### 5.4 SIFT', PCA-SIFT', SURF', SURF128' の比較

データベース, 検索質問は, 5.1 節 (2) のものである. 前節と同様に, ANN で検索を行い, 処理時間と認識率を調べた. その結果を, 図 6 に示す. 図より, PCA-SIFT' で最もよい結果となったことが分かる. そして, これらを前節のものと比較すると, SIFT' や PCA-SIFT' では, 前節のものよりも精度が高いことが分かる. 考えられる原因として, 特徴点/領域の数の多さが挙げられる. 表 1, 表 2 から分かるように, SIFT', PCA-SIFT' では, それぞれ 2034, 2069 の特徴点/領域が抽出されている一方, 前節の Hessian-Affine, MSER の組み合わせでは,  $679 + 354 = 1033$  にとどまっている. 平均特徴点/領域数が 571.5 である SURF'/SURF128' では, Hessian-Affine 領域での認識率と類似して, 最高でも 95% 程度の性能にとどまっていることから, 特徴点/領域の数が認識率に大きく関与している可能性があるといえる.

## 6. まとめ

本稿では, 局所領域 (特徴点) と, 局所記述子を用いて画像

の局所特徴量を記述し, ANN で検索したときの比較検討結果を述べた. 実験によって分かったことは以下の通りである.

(1) Hessian-Affine と MSER で抽出された特徴点/領域を組み合わせることにより, それぞれを単独で用いる場合よりも, 認識率が向上した.

(2) PCA-SIFT よりも PCA-SIFT' の方が高性能であった.

(3) PCA-SIFT' が最も高い認識率であった.

以上の事実について, 特徴点/領域の数を調べてみると, 特徴点/領域数が認識率に大きく関与している可能性があることが分かる. (1) では, 特徴点/領域を組み合わせることにより, (2) では, 異なる特徴点抽出手法を用いることにより, 特徴点/領域数が増加している. (3) の PCA-SIFT' では, 平均特徴点/領域数も最も多く, 2069 であった.

特徴点/領域抽出手法を統一し, 特徴点/領域数を同等とした場合, PCA-SIFT よりも, Shape Context や GLOH の結果が良好であった. このことから, 特徴点/領域の数を増やした上で, Shape Context や GLOH を局所記述子として用いれば, それらは PCA-SIFT' の認識率を上回る可能性があるといえる.

今後の課題として, 各特徴点/領域抽出のパラメータを変化させることによって, 特徴点/領域の数と認識率の関係を比較, 検討することや, データベースの登録画像数, 検索質問数をさらに増やして実験を行うことが挙げられる.

謝辞 本研究の一部は, 科学研究費補助金 (基盤研究 (B) 19300062) の補助による.

## 文 献

- [1] 高木, 藤吉, “SIFT 特徴量を用いた交通道路標識認識”, 第 13 回画像センシングシンポジウム SSII07, LD2-06, June 2007.
- [2] J. Kim, T. H. Rhee, K. Kim, and J. H. Kim, “Signboard Recognition by Consistency Checking of Local Features”, 電子情報通信学会技術研究報告, PRMU2007-118, October 2007.
- [3] D. G. Lowe, “Distinctive image features from scale-invariant keypoints”, Internal Journal of Computer Vision, 60, 2, pp.91-110, January 2004.
- [4] Y. Ke, and R. Sukthankar, “PCA-SIFT: A more distinctive representation for local image descriptors”, Proc. CVPR'04, vol.2, pp.506-513, 2004.
- [5] H. Bay, T. Tuytelaars, and L. V. Gool, “SURF: Speeded Up Robust Features”, Proc. of the 9th European Conference on Computer Vision, May 2006.
- [6] K. Mikolajczyk and C. Schmid, “Scale & Affine Invariant Interest Point Detectors”, International Journal of Computer Vision 60(1), pp.63-86, January 2004.
- [7] K. Mikolajczyk, et al., “A Comparison of Affine Region Detectors”, International Journal of Computer Vision, Springer Science, January 2006.
- [8] K. Mikolajczyk and C. Schmid, “A Performance Evaluation of Local Descriptors”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.27, No.10, pp.1615-1630, October 2005.
- [9] S. Belongie, J. Malik, and J. Puzicha, “Shape Matching and Object Recognition Using Shape Contexts”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol.2, no.4, pp.509-522, April 2002.
- [10] J. Matas, O. Chum, M. Urban, and T. Pajdla, “Robust wide baseline stereo from maximally stable extremal regions”, in Proc. BMVC, 2002.
- [11] S. Arya, D. Mount, R. Silverman and A. Y. Wu, “An optimal algorithm for approximate nearest neighbor searching”, Journal of the ACM, vol.45, no.6, pp.891-923, 1998.