

## 局所特徴の相対位置情報を用いたカテゴリ分類 Object Categorization Based On Relative Position of Local Features

吉村 功二†  
Koji Yoshimura

堀田 一弘†  
Kazuhiro Hotta

高橋 治久†  
Haruhisa Takahasi

### 1. はじめに

カテゴリ分類とは、与えられた入力画像中に含まれる物体が属するカテゴリ(車, 携帯電話)に分類することを指す。画像に含まれる対象物体の大きさや位置, 向きや見え方は未知であるので, 従来法では画像から特徴的な局所領域を検出し, 認識に用いていた。

現在この分野において, カルフォルニア工科大学の Caltech101 という 101 カテゴリを含む評価画像データが無償で公開されている<sup>†</sup>。このデータセットに対して, H. Zhangらは K 最近傍法と SVM を組み合わせた SVM-KNN を用いて 6 割以上という最も良い結果を出している [1]。ただし, 検出された局所特徴量と特徴点の絶対座標を距離関数に組み込んで識別を行っている。また, S Lazebnik らも画像を位置情報を持ついくつかの領域に分割し, 対応する位置の領域毎に類似度を測ることにより高い識別率を出している [4]。しかし, 絶対位置情報を用いることは, 対象物体が画像中の同位置にあるという前提の上に成り立っており, 一般的な画像については成立しない。

そこで本論文では画像の持つ位置情報を相対位置情報として使用する。検出された特徴点の重心を計算し, その重心から特徴点までの傾きを位置情報として用いる。これは特徴点の対象物体に集中しており, 特徴点の重心が対象物体の中央付近にあるという仮定に基づいている。これにより画像中の対象位置に不変な位置情報を用いることができる。得られた傾き情報を従来の距離関数に組み込んだものを新しい距離関数として提案した。

本論文では位置や向きなどにロバストに局所特徴を検出し, 記述する手法である Scale-Invariant Feature Transform(SIFT) [3] を用いる。さらに検出された全局所特徴を用いるのではなく, 訓練データごとに特徴量のマッチングを行うことにより, カテゴリ分類に重要な特徴点だけを選択した。これにより計算量が低減すると共に識別性能の向上も期待できる。

計算機実験では二つの識別器を用いた。一つ目の方法では, 訓練データから主成分分析を用いて各カテゴリに特化した空間を作成し, この空間からの距離 (Distance-From-Feature-Space) [2] を用いる。この DFFS と提案した距離関数を組み合わせたものを K 最近傍法で識別を行う。これを DFFS-KNN と呼ぶ。二つ目の方法は, 提案した距離関数をカーネル関数として拡張したサポート

ベクトルマシン (SVM) である。Caltech-101 に対して, SVM と DFFS-KNN の二つの識別器を用いて相対位置情報の有無を比較した結果, 提案手法の優位性を得ることができた。

### 2. SIFT 特徴

Scale-Invariant Feature Transform(SIFT) は, 特徴点の検出と特徴量の記述を行うアルゴリズムである。まず特徴点の検出は, 入力画像  $I(u, v)$  にスケールの異なるガウス関数  $G(x, y, \sigma)$  を畳み込んだ平滑化画像  $L(u, v, \sigma)$  の差分 (DoG 画像) から求める。DoG 画像の極値となる場所を特徴点とする。また特徴点周辺領域で勾配方向ヒストグラムをとり, その最大方向を特徴点の方向 (オリエンテーション) とする。従来の SIFT ではこの最大方向にしたがって特徴量を回転させることにより, 回転に不変な特徴量としている。しかし, 回転に不変な特徴量とすることは, 方向成分の情報を欠落させていることと同じである。実際に回転有無のデータを後述する識別器 (カテゴリ数は 101, 学習数は 20) にかけて評価したところ, 回転なしの方が 10% 以上識別率が高いという結果が得られた。そのため, 本研究では特徴点周辺領域の回転は識別性能に悪影響を与えると考え, 使用しなかった。特徴量の記述にも同様に勾配情報を用いる。周辺領域を一辺を 4 ブロックの計 16 ブロックに分割し, ブロックごとに 8 方向の勾配方向ヒストグラムを作成するため,  $4 \times 4 \times 8 = 128$  次元の特徴量が得られる<sup>‡</sup>。

表 1: 回転の有無による識別率の違い

	回転有り	回転無し
KNN による識別率	19.41	32.97

### 3. 類似度

#### 3.1 距離関数

各画像から SIFT で抽出される特徴点数は一定ではない。一般に, 距離関数は対称性  $D(x, y) = D(y, x)$  をもつ必要があるため, 2 枚の加増から抽出された特徴点に基づく距離関数には工夫が必要である。

従来の距離関数

$$D(I_L, I_R) = D'(I_L \rightarrow I_R) + D'(I_R \rightarrow I_L) \quad (1)$$

<sup>‡</sup>Rob Hess により提供されているソフトウェアを使用した。  
<http://web.engr.oregonstate.edu/~hess/index.html>

<sup>†</sup>電気通信大学, The University of Electro-Communications

<sup>‡</sup><http://www.vision.caltech.edu/feifeili/Datasets.htm>

$$D'(I_L \rightarrow I_R) = \frac{1}{m} \sum_{i=1}^m \min_{j=1..n} \| f_i^L - f_j^R \|^2$$

$f_i^L, f_j^R$  は距離を計算する左右2枚の画像の  $i, j$  番目の SIFT 特徴を表し,  $m, n$  は各画像で抽出された特徴点数を表す. すべての特徴量間のユークリッド距離を求め, それが最小となる組み合わせを選び, その総和を距離としている.

提案する距離関数

従来の距離関数は SIFT 特徴量のみから算出される距離であるが, SIFT では特徴量だけでなく特徴点の位置も抽出される. 物体認識において, 局所特徴の位置というのは重要な情報である (例えば, 人間の顔で目は上部, 鼻は中央, 口は下部にあるなど). そこで, 特徴点の位置を考慮に入れた距離関数を考える.

しかし, 特徴点の位置情報を距離関数に加える際, 直接特徴点の座標を用いることは出来ない. なぜなら, 距離を測る2枚の画像のサイズは異なっており, 対象物体の大きさや位置は未知であるからである. そこで, 各画像から特徴点の重心を求めて, その重心からの特徴点までの角度を位置情報として用いる. これは特徴点に対象物体に集中しており, 特徴点の重心が対象物体の中央付近にあるという仮定に基づいている.

$$r_C^L = \frac{1}{m} \sum_{i=1}^m r_i^L \quad (2)$$

$$\begin{bmatrix} dx_i^L \\ dy_i^L \end{bmatrix} = r_i^L - r_C^L \quad (3)$$

$$\theta_i^L = \tan^{-1} \left( \frac{dy_i^L}{dx_i^L} \right) \quad (4)$$

ここで  $r_i^L$  は左画像の  $i$  番目の特徴点の座標,  $r_C^L$  は左画像の特徴点の重心である. 同様に右画像の特徴点への角度  $\theta_j^R$  を求め, その差の絶対値  $\theta$  を左右の特徴点の相対的な位置のずれとした.

$$\theta = \| \theta_i^L - \theta_j^R \| \quad (0 \leq \theta \leq \pi) \quad (5)$$

この  $\theta$  を用いて,  $\theta = 0$  で最小の値をとり,  $\theta = \pi$  で最大の値をとる関数  $A$  を定義する.

$$A(\theta_i^L, \theta_j^R) = 1 + a(1 - \cos \| \theta_i^L - \theta_j^R \|) \quad (6)$$

従来の距離関数に積の形で補正できるように最小値が1をとるように正規化してある. ここで  $a$  は局所特徴の相対位置情報の重みを決める係数であり, 関数  $A$  の定義域を定めている.  $a$  が大きいほど特徴点の位置のずれを重

E

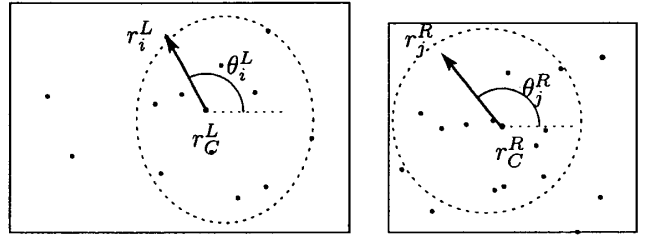


図1: 特徴点の相対位置情報

要視していることとなる. この関数と従来の距離関数式(1)を積の形でとり, 以下の式

$$D_A(I_L, I_R) = D'_A(I_L \rightarrow I_R) + D'_A(I_R \rightarrow I_L) \quad (7)$$

$$D'_A(I_L \rightarrow I_R) = \frac{1}{m} \sum_{i=1}^m \min_{j=1..n} [ \| f_i^L - f_j^R \|^2 \cdot (1 + a(1 - \cos \| \theta_i^L - \theta_j^R \|))] \quad (8)$$

を新たな距離関数  $D_A(I_L, I_R)$  とした.  $a = 0$  の場合は, 従来の距離関数と同じになっていることがわかる.

3.2 特徴点数の削減

上記のように距離関数を定めたが, すべての局所特徴量間の距離を求めてマッチングを行うので計算量が非常に大きくなる. そこで, 各カテゴリごとに訓練データ同士で距離を測り, その際に, 特徴点と同カテゴリ内の他の画像の特徴点と何回マッチしたかを調べる. すべての訓練画像に対して, マッチした回数が上位の特徴点だけを残した. 残す特徴点の個数は実験的に求め, 上位100個の特徴点を残した. 同一カテゴリ内の他の画像とマッチしない特徴点は, その画像固有の特徴点といえカテゴリ分類に不必要な特徴点と考えられる. 図2に削減前と削減後の特徴点を示した. 削除後はアコーディオンの特徴である鍵盤, ボタン, 蛇腹などに特徴点が集まっていることがわかる. これにより計算量の削減と識別率の向上にも繋がった.

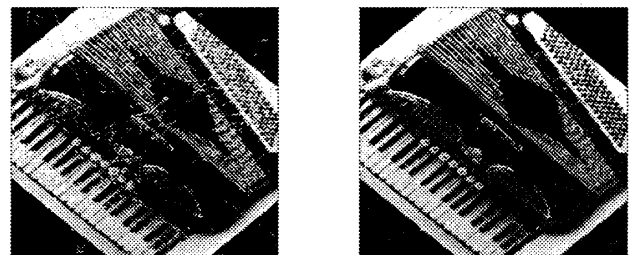


図2: 特徴点の表示 (左:削減前 右:削減後)

4. 識別器

4.1 DFFS(distance-from-feature-space)

主成分分析 (PCA) は与えられたデータ集合に対して, 情報を最も良く近似する方向に対して軸を取ることによ

り、データ集合に適した特徴空間を作成する。従って、あるカテゴリに対する特徴空間に、カテゴリの異なるサンプルを射影した場合、情報が正しく表現されない。一度射影した情報を元の空間に戻した際、元情報と大きな差が生じてしまう。この生じた差を、特徴空間からの距離 (DFFS) と呼ぶ。

各カテゴリの訓練集合に対して PCA をかけて変換行列  $P$  を求める。入力画像  $I_m$  の SIFT 特徴  $F = \{f_1, \dots, f_m\}$  に変換行列  $P$  の転置をかけることで特徴空間へ射影し、特徴量  $Y = \{y_1, \dots, y_m\}$  を得る。この特徴量  $Y$  に再び変換行列  $P$  をかけることで元の空間に戻すことができる。

$$\begin{aligned} DF(P, I_m) &= \frac{1}{m} \|F - F'\|^2 & (9) \\ Y &= P^t F \\ F' &= PY \end{aligned}$$

入力情報との差を測る関数を DF とする。これは、入力画像が訓練集合と離れているほど大きい値をだす測度となる。この測度だけでも識別は可能である。

#### 4.2 DFBS-KNN

入力画像とあるカテゴリとの類似性を DFBS を用いて測ることができる。この類似度に従来法の KNN を組み合わせることにより DFBS-KNN を実現した。DFBS-KNN の距離関数は以下の式で表される。

$$D_D(I_R, I_L) = D_A(I_R, I_L) \cdot DF(P_k, I_R) \quad (10)$$

この距離関数は、第 3.1 節で述べた距離関数  $D_A$  と訓練サンプル  $I_L$  の属するカテゴリ  $k$  との DFBS の積からなる。この距離を用いて近傍  $K$  個のサンプルを選択し、その多数決で識別すればよい。DFBS は特徴空間へ射影して戻すだけなので計算量もかからず、高々カテゴリ数だけ実行すれば良いので高速である。従って DFBS を用いてある程度カテゴリの候補を絞ってから、時間のかかる訓練データとの KNN を実行することも出来る。

#### 4.3 サポートベクトルマシン (SVM)

本研究では前章で定めた距離関数  $D_A(I_L, I_R)$  を用いて RGF カーネルに似た以下のようなカーネルを使用した。

$$K(\mathbf{x}, \mathbf{z}) = \exp\left(-\frac{D_A(\mathbf{x}, \mathbf{z})}{\sigma^2}\right) \quad (11)$$

SVM は 2 クラスのパターン識別器であるが、複数の SVM を組み合わせることにより多クラスでの識別も可能になる。本研究では全クラスの 1 クラス対 1 クラスの SVM を行い、その識別結果の勝利数をカウントする投票方式を用いた<sup>1</sup>。

<sup>1</sup>本論文では LIBSVM を使用した。  
http://www.csie.ntu.edu.tw/~cjlin/libsvm/

## 5. 実験

本論文では Caltech-101 データセットを用いて評価を行う。前処理としてカラー画像を濃淡画像に変換した。

### 5.1 相対位置情報の有効性

相対位置情報の有効性を確認するために、位置情報を含む提案手法と従来の距離関数を用いた手法の比較を行った。表 2 に識別結果を示す。全 101 カテゴリからラ

表 2: 位置情報の有無による識別率 [%] の変化

	位置情報なし	位置情報あり
KNN	32.48	43.47
DFBS-KNN	38.12	45.94
SVM	42.97	53.90
DFBS	33.47	

ンダムに 30 枚ずつ画像を選出し、学習サンプル数 20 枚、テストサンプル数 10 枚、各パラメータは  $K = 1$ ,  $a = 1$ ,  $\sigma = 0.9$ ,  $C = 1.4$  である。全ての手法において位置情報を加えた提案手法の方が高い識別能力を示した。

また Caltech-101 から 10 カテゴリを選出して、式 (7) 中の傾き情報の重み係数  $a$  の最適なパラメータを求めるべく、係数  $a$  を 0.2 刻みで変化させながら、訓練サンプル数 10 で KNN, DFBS-KNN ( $K = 1$ ) で識別を行った。図 3 に傾き情報の重み係数  $a$  を変化させたときの異なるカテゴリ同士と同じカテゴリ同士の距離の比を示した。距離の比は以下の式で与えた。

$$\text{距離の比} = \frac{\text{異なるカテゴリ同士の距離平均}}{\text{同じカテゴリ同士の距離平均}} \quad (12)$$

重み係数  $a$  を変化させたとき、距離の比が増加していれば相対位置情報が正しく働いているといえる。図 3 を見てわかるように距離の比が上昇している。図 4 に識別率のグラフ (data1:KNN data2:DFBS-KNN) を示す。傾き情報をまったく使わない場合 ( $a = 0$ ) に比べて、使った場合 ( $a > 0$ ) の方が 2 手法とも良い識別性能を示している。また、図 3 と図 4 の形から距離比と識別率の相関関係ははっきりと見て取れる。

以上のことから相対位置情報の有効性が確認できた。また、この実験により重み係数  $a$  は一定値以上であれば識別結果があまり変化しないことが見て取れ、次節の実験では  $a = 3$  で行った。

### 5.2 評価実験

実験では各カテゴリからランダムに 30 枚の画像を選出し、一部を訓練データとし、残りをテストデータとして用いた。ただし、訓練サンプル数が 30 の場合には、テストサンプル数を 10 とするが、テストサンプルが 10 枚に満たないカテゴリは残り全ての画像をテストデータとした。各カテゴリでテストサンプル数が異なるため、各カテゴリの識別率の平均を最終的な識別率とした。訓練

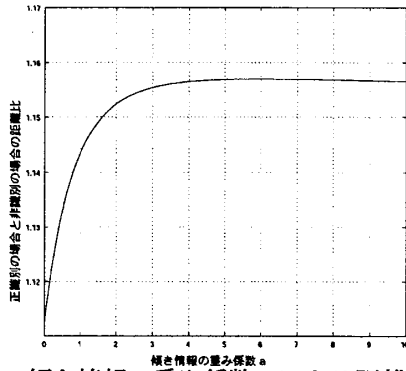


図 3: 傾き情報の重み係数  $a$  による距離の変化

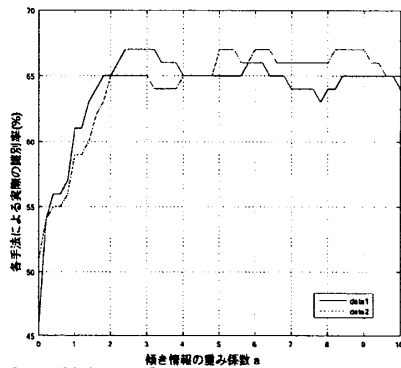


図 4: 傾き情報の重み係数  $a$  による識別率の変化

データの特徴点数を 100 まで削減し、各特徴量の次元は主成分分析を用いて 128 次元から寄与率 0.9 となる次元に削減した。また、DFFS を測る際に用いた各カテゴリごとの特徴空間の次元は 20 次元とし、 $K = 1$ 、 $a = 3$  とした。SVM のパラメタは  $\sigma = 0.9$ 、 $C = 1.3$  である。これらのパラメタは実験から経験的に求めた。訓練サンプル数を変化させた時の DFFS 単体での識別率、KNN による識別率、DFFS-KNN による識別率、SVM の識別率をそれぞれ表 3 に示した。

表 3: 各手法による識別結果 [%]

学習枚数	KNN	DFFS-KNN	DFFS	SVM
5	30.1	33.7	24.4	36.1
10	38.3	41.3	28.2	45.6
15	43.0	45.8	31.1	50.9
20	45.7	48.3	30.1	51.9
30	49.4	52.4	35.0	56.7

どの手法も訓練サンプル数を増やすことによって性能が向上していることがわかる。また、提案手法を組み込んだ SVM が最も良い性能を示している。図 5 に最も高い識別率が得られた SVM と最近発表された他の手法の識別率とを比較した。最も良い結果を出している Zhang[1] には及ばないが、絶対位置情報を使わない手法としては高水準の性能となった。

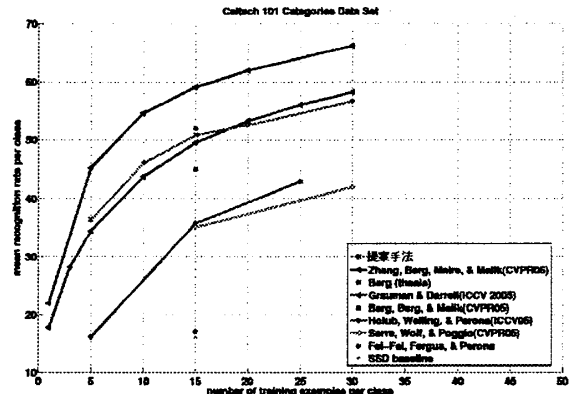


図 5: 提案手法の識別率

## 6. おわりに

本研究では、重心からの傾きを用いることにより、局所特徴点の位置情報を相対位置情報として利用する手法を提案した。また、SIFT の最大勾配方向への回転の廃止、特徴点の削減なども本手法独自のものである。提案した距離関数を用いた場合、従来よりも高い識別率を得ることができた。本手法では局所特徴およびその位置のみに着目したが、さらに識別率を向上させるためには、画像全体の特徴と組み合わせていくことも検討していく必要がある。

また、SIFT 特徴点の個数は画像によって大きな開きがあり、それが識別に悪影響を及ぼしている可能性がある。特徴点数が多くなると、計算時間も増大する。今後は特徴点の検出に別の手法を用い、特徴量の記述だけに SIFT を用いることも考えられる。

2007 年にカルフォルニア大学からカテゴリ数を増やした Caltech-256 が公表された。このデータセットは Caltech-101 で行った向きや大きさなどの統一が行われなかったため、より一般的な画像データセットとなっている。これらのデータセットに対して評価実験を行うことにより、本手法の優位性をさらに確認していきたい。

## 参考文献

- [1] H.Zhang et al. "SVM-KNN:Discriminative Nearest Neighbor Classification for Visual Category Recognition", Proc. IEEE Computer Vision and Pattern Recognition, Vol.2, pp.2126-2136, 2006.
- [2] B.Moghddam et al. "Probabilistic Visual Learning for Object Representation", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol.19, No.7, 1997.
- [3] D.G.Low."Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, Vol.60, No.2, pp.91-110, 2004.
- [4] S.Lazebnik,et al."Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories", Proc. IEEE Computer Vision and Pattern Recognition, Vol.2, pp.2169-2178, 2006.