

スパースコーディングによる基底画像を用いた風景画像認識

佐々木翔大[†] 篠沢佳久[‡]

慶應義塾大学大学院 理工学研究科 開放環境科学専攻[†]
慶應義塾大学 理工学部 管理工学科[‡]

1 はじめに

近年、自動車の自動運転やロボットの行動認識などにも画像認識技術が応用されるようになってきた。そのため、対象物に制限を設けず、物体を認識する技法（一般物体認識）の研究が進められている。本研究では、特に風景画像を対象とした認識手法の確立を試みている。

これまで風景画像を対象として、さまざまな認識手法が考案されてきた[1]。特に Olshausen らは、人間の初期視覚の情報処理をモデル化した結果より、自然画像は空間的な局所性や方向性を持つ基底画像から構成されていることを示した[2]。すなわち網膜上の画像は、初期視覚の段階で基底画像の線形和の形に分解されることが分かっている。さらには、この分解において基底画像を重みつける確率変数の情報量を最小にすることで、確率過程の非定常性や不連続性の影響を表現するために特徴的な局所性をもった基底画像を抽出できることも示されている。

基底画像を抽出する手法として、独立成分分析（Independent Component Analysis, ICA と略す）が挙げられる。ICA においては、画像を分解する際、統計的な独立性のみを仮定しているため、必ずしもスパースネス性の高い特徴が得られるとは限らない。一方で、スパースネス性の高い特徴を得る手法の一つに、スパースコーディングが挙げられる。本研究においては、スパースコーディングを用いて基底画像を抽出し、画像の再構成時における特徴を用いて、風景画像を認識する手法を考案する。

2 スパースコーディングを用いた画像認識

図1に提案手法の流れを示す。図1の左図には学習用画像を対象に、正規化、領域分割を行なった後（2.1 節）、スパースコーディングにより基底画像と特徴ベクトルを抽出する流れ（2.2 節）を、右図には、抽出した基底画像を用いて、未知画像を再構成し（2.3 節）、その際に得られた特徴を用いて認識を行なう流れ（2.4 節）を示す。

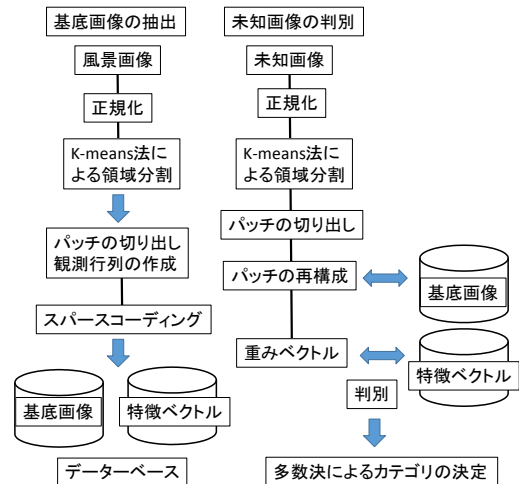


図1 提案手法の流れ

2.1 前処理および領域分割

風景画像を対象に、前処理として一定の大きさ（ 360×270 もしくは 270×360 ）に正規化する。次に k-means 法を用いて同一のカテゴリ領域に分割し、学習用画像とする（図2）。本研究においては雲、空、木、道、岩、砂、水、建物を認識対象のカテゴリとした。類似した色、近い位置関係にある領域を同じカテゴリとして分割するため、k-means 法に用いる特徴量としては、RGB 値（3 次元）、HSV 値（3 次元）、位置座標（2 次元）の合計 8 次元とした。



図2 k-means 法による領域分割

2.2 基底画像および特徴ベクトルの抽出

カテゴリごとに分割した学習用画像から観測行列を作成する。カテゴリ数を C 個（本研究においては $C=8$ ）、一つのカテゴリにつき L 枚の学習用画像を用意するものとする。まず一枚の学習用画像から $N \times N$ の大きさの正方領域（パッチ）をランダムに P 個切り出す。一つのカテゴリからは $L \times P$ 個、全ての学習用画像からは $C \times L \times P$ 個のパッチを抽出する。切り出したパッチを $\mathbf{x}_i$ ($i=1, 2, \dots, C \times L \times P$) とする。ここでパッチ $\mathbf{x}_i$ は M ($N \times N \times T$) 次元ベクトルで表現できる。

Recognition of landscape images using basis images extracted by sparse coding

[†]Shodai Sasaki, Keio University

[‡]Yoshihisa Shinozawa, Keio University

T はパッチの各画素の特徴数である．本研究においては，パッチの特徴量として RGB 値を用いる（さらにはパッチごとに位置情報（2 次元）も用いているため， $M=N \times N \times 3 + 2$ となる）．

次に観測行列を対象に Dictionary learning を行ない，S 枚の基底画像を抽出する．同時に，基底画像を用いて，元の学習用画像のパッチの再構成（基底画像との線形和で復元）を行ない，係数ベクトル $\mathbf{a}_{clp}$ （特徴ベクトル（S 次元）と呼ぶ， $c=1,2,\dots,C$ ， $l=1,2,\dots,L$ ， $p=1,2,\dots,P$ ）を求め，カテゴリごとに $L \times P$ 個の特徴ベクトルを保存する．

2.3 パッチの認識

未知画像についても，正規化処理，k-means 法を行ない，K 個の領域に分割する．各領域 U_k ($k=1,2,\dots,K$) から P 枚のパッチを切り出す．そして 2.2 節で抽出した基底画像を用いて，各パッチの再構成を行ない，係数ベクトル $\mathbf{a}_{kp}$ （重みベクトルと呼ぶ， $p=1,2,\dots,P$ ）を求める（図 3）．

各パッチがどのカテゴリかを認識するため，重みベクトル $\mathbf{a}_{kp}$ と特徴ベクトル $\mathbf{a}_{clp}$ とのユークリッド距離を求め，最も距離の短いカテゴリを認識結果とする（最近傍法）．

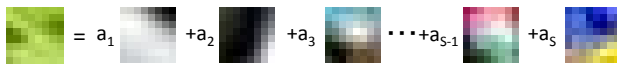


図 3 基底画像によるパッチの再構成

2.4 カテゴリの認識

領域分割後，各領域 U_k からは P 枚のパッチを切り出している．そこで，2.3 節のパッチの認識処理を P 回繰り返し，多数決法によって，カテゴリを決定する（図 4）．

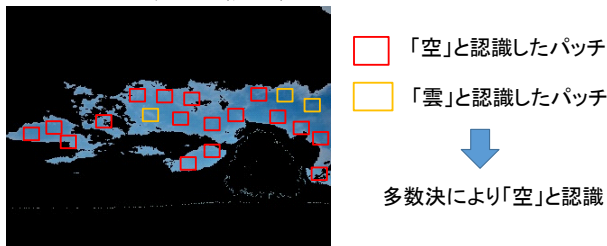


図 4 多数決によるカテゴリの認識

3 評価実験

提案手法を評価するため，カテゴリごとに学習用画像と評価用画像を 50 枚ずつ収集した（2-cross Validation）．提案手法におけるパラメータを表 1 に示す．パッチの大きさ N，切り出すパッチの枚数 P を変え，全カテゴリに対する平均認識率を求めた．その結果， $S=4$ ， $N=8$ ， $P=50$ において，71.6%の認識率が得られた．

表 1 評価実験におけるパラメータ

パラメータ	値
カテゴリ：C	雲，空，木，道，岩，砂，水，建物
カテゴリごとの画像数：L	学習用，評価用ともに 50
パッチの大きさ：N×N	N=8, 12, 16
パッチの枚数：P	P=25, 50, 75, 100, 150, 200
パッチの画素の特徴量	RGB 値，位置情報
認識時の基底画像数：S	S=4, 8, 12, 16, 20, 24, 28

そこで，同じ条件下で ICA を用いて抽出した基底画像による認識手法[3]との比較を行なった（ICA の場合，基底画像は $M=194$ 個抽出される）．その結果を表 2 に示す．スパースコーディングにおいては，認識時に用いる基底画像数を変更した際の認識率も示す．

表 2 提案手法と ICA による平均認識率

基底画像数 S	平均認識率 (%)
S=4	71.6
S=8	71.1
S=12	70.0
S=16	69.8
S=20	68.4
S=24	68.3
S=28	67.4
ICA	69.8

表 2 より， $S=4,8,12$ の時，スパースコーディングの方が高い認識率が得られた．スパースコーディングを用いた場合では，認識時に利用する基底画像の個数を少なくすることで ICA の場合よりも高い認識率を得ることができた．

4 結論

本研究においては，スパースコーディングによって抽出した基底画像による風景画像の認識手法を提案した．今後は，パッチの切り出し方（位置）や，カテゴリの認識方法の改良を行ない，認識精度の向上を図る予定である．

参考文献

- [1] 藤田隆二郎：車載カメラ映像による風景特徴解析，第 4 回情報科学技術フォーラム（FIT2005），pp.137-138，2005．
- [2] B.Olshausen and D.Field：Wave-let-like Receptive Fields Emerge from a Network That Learns Sparse Codes for Natural Images, Nature, Vol.381, pp.607-609, 1996.
- [3] 園朋也，篠沢佳久：基底画像を用いた風景画像認識，第 15 回画像の認識・理解シンポジウム（MIRU2012），IS1-42，2012．