修士論文

風景画像に対する画素単位対象物ラベルづけ手法

指導教官　　池田 克夫 教授

京都大学大学院工学研究科修士課程情報工学専攻

椋木 雅之

平成 5 年 2 月 15 日

風景画像に対する画素単位対象物ラベルづけ手法

椋木 雅之

内容梗概

本論文では，風景画像を対象として，画像の対象物ラベルづけを画素単位で行う画素単位対象物ラベルづけ手法を提案する．画像の対象物ラベルづけとは，一つの領域が一つの対象物に対応するように画像を分割し，各領域に対応する対象物のラベルを付加することである．本手法は，画素単位で得られる特徴量と対象物との対応関係の知識である対象物モデルを用いて，画素毎に対象物らしさを与え，対象物ラベルづけを行うものである．対象物に対応する領域は，同一のラベルを持つ画素を集めることにより得られる．

従来の研究では，対象物ラベルづけを領域単位で行っていた．この場合，対象物ラベルづけを行うためには，初期的に領域を生成する必要がある．しかし，対象物と一対一に対応した領域を対象物に関する知識なしに生成することはできない．さらに，仮に対象物と一対一には対応した領域を生成できたとしも，風景画像中の自然物は形状を持たないか不定形のものが多いので，形状，大きさの特徴は対象物ラベルづけに有効に働かないと考えられる．一方，対象物ラベルづけのために利用できる色，テクスチャ，位置，位置関係などの知識は，画素単位でも得られる．従って，対象物ラベルを付加するためには，領域生成を行う必要はない．

画素の評価を行うための対象物モデルには，対象物が画像上で示す特徴量を学習したニューラルネットワークを用いる．この場合，人間が知識を明確に表現することができなくても，例示により対象物モデルが構築可能である．構築した対象物モデルを用いて，非学習データに対して 73.9 % の認識率を得た．この結果は，領域単位で同様に構築した対象物モデルを用いた場合と同程度以上であり，画素単位対象物ラベルづけが領域単位と同程度以上に有効であることが示された．

局所的に得られる特徴量のみで行った対象物ラベルづけでは，画像全体としてみた場合，対象物の位置関係が矛盾していることがある．この問題を解決するためには，対象物ラベルづけに対象物相互の位置関係の知識を導入する必要がある．本論文では，画素単位対象物ラベルづけで位置関係知識を利用する方法を実現した．これは，位置関係知識を満たす画像の対象物ラベルづけの中で，画素毎に定めた対象物ラベルの評価値の和が最大になるものを選択するものである．画素単位対象物ラベルづけでは，画素が規則的に並んでいるため，位置関係を得ることが容易であり，位置関係知識を容易に導入できた．

風景画像の対象物ラベルづけの一つの応用として，画像データベースがある．画素単位対象物ラベルづけの結果を画像に対する検索キーとして利用する場合，認識率が問題になる．検索者が検索したい画像のスケッチを入力し，類似した画像を提示する検索方法で必要となる認識率を評価した結果，10 枚以内の候補の提示で検索が成功するためには，画像の約 50 % 以上が正確にラベルづけされていればよいと見積もられた．本論文で実現した画素単位対象物ラベルづけでは，認識率が 70 %程度であるので，平均的には，画像への検索キーの付加に利用可能であることが示された．130 枚の画像を対象に検索を行った結果，8 割の画像に対して，10 枚以内の候補の提示で検索を行うことができた．

Pixel-based object labeling method of out-door scenes

Masayuki Mukunoki

Abstract

We propose a new approach to the problem on image labeling of out-door scenes. It is called "Pixel-based labeling method." It proceeds as follows: **1.** to calculate feature values of every pixel; **2.** to evaluate the object labels of each pixel with the object models which are mappings between the feature values and the object labels; **3.** to assign an object label to every pixel making use of the result of evaluation.

The purpose of object labeling of an image is to segment the image into regions, each of which corresponds to one object, and to assign an object label to the object in it. Most of the studies so far have taken the approach of region-based labeling, in which they initially segment an image into regions and assign an object label to them. In such approaches, an initial region does not correspond to an object, although the result of the object labeling largely depends on the initial regions.

Since objects in out-door scenes are natural objects, they do not have definite shapes. Pixels do not have the features of shape and size, but these features are not effective for the object labeling in this case. Other features, such as color, texture, position and relations, can be obtained from the pixels. So it is needless to make the initial regions in order to assign an object label to an object in an image. Therefore, Pixel-based labeling method becomes effective.

We construct the object models with Neural-Network which has learned the mapping between the feature values of a pixel and its object label. Although we cannot define the mapping clearly, we can give ex-

amples about the correspondence between a pixel and its object label, and the feature values of the pixel in the examples can be calculated. In this way, we can easily construct the object models. The object models correctly label 73.9% of data in our experiment. The result is better than that of the region-based labeling.

Next, we introduce the knowledge of relations among objects. The result of the object labeling derived from local feature values does not always satisfy the correct relationship among the objects. The knowledge of relations can correct them. We use the output values of Neural-Network, which has learned the object models, as the evaluation values of each object label, and assign the object labels to every pixels in the image which satisfy the knowledge of relations and maximize the sum of the evaluation values. The result of object labeling is improved a little.

One application of the object labeling is to use the result as the index of images in the image database. In this case, the correctness of the object labeling comes into question. The method we use for retrieval is that: **1.** the system labels images in it with Pixel-based labeling method and stores them; **2.** the retriever gives the sketch of an image he want to retrieve; **3.** the system shows original images in order of the similarity between the sketch and the labeled image. The experiments show that when the half of pixels in an image have the correct label, we can find the intended image in the top ten candidates. Since the correctness of the object labeling with the Pixel-based labeling method is over 70%, the result of it can be used as the index for the image database. In our experiments, about 80% of the images in the image database, which includes 130 images, appear in the top ten candidates.

目次

1　はじめに

高度情報化社会の発展に伴い，計算機の応用範囲は飛躍的に広まった．特に，最近の計算機の処理能力の向上と大容量記憶装置の普及により，数値やテキストだけでなく，データ量の多い音声や画像も扱うことが可能となった．音声や画像などを扱う際には，データを加工するのみでなく，与えられたデータの持つ内容をある程度認識し，処理を行う必要が生じる．例えば，画像データベースで画像内容による検索を実現するためには，画像内容を検索キーとして抽出する必要があるが，これを計算機により自動的に行うことができれば，人手で行う場合に比べて非常に効率的である．

画像認識の研究は従来から行われている．その中の一分野に，情景の解釈を行う研究がある．松山ら [1] は，航空写真を対象に，領域の特徴をもとに草地・森・道路・自動車などの抽出を行った．李，辻ら [2] は，屋内情景を対象に，移動カメラを用いて室内の三次元構造を推定した．また，屋外画像を対象とした研究には [3][4][5][6] などがある．航空写真では，対象物が遠距離にあり，ほぼ平面上に存在しているとみなせるため，二次元的な位置関係の知識を用いて解釈が行われる．屋内画像の認識では，自律移動するロボットやマニュピレータによる物体の移動などの応用が考えられる．この場合，近距離にある対象物との正確な距離を得ることが必要であり，複数の画像やセンサー情報を用いて対象物の三次元情報を推定する研究が行われている．一方，屋外画像では，対象物が中距離から遠距離に存在するため，航空写真とは異り三次元的な知識がある程度利用できるが，屋内画像のように正確な三次元情報の抽出は必ずしも必要でない．また，比較的多種類の対象物が様々な見え方で画像に現れるため，それに対処できる柔軟な処理が必要であるなど特徴がある．以下では，屋外画像の一種で，画像内の主な対象物が自然物である風景画像を対象に，画像の対象物

ラベルづけを得ることを考える．

画像の対象物ラベルづけとは，「一つの領域が一つの対象物に対応するように画像を分割し，各領域に対応する対象物のラベルを付加すること」と定義する．そのためには，対象物に対応する領域の生成と対象物ラベルの付加という二つの処理が必要である．従来の研究では，対象物ラベルづけを領域単位で行っているため，初期的に領域を生成していた．初期的な領域は，対象物ラベルづけの後併合・再分割され，最終的に対象物に対応する領域となる．そのため，初期的な領域は，複数の対象物にまたがらないという過分割の仮定を満たす必要がある．一方で，初期的な領域に対して対象物ラベルづけを行う際には，一つの領域が一つの対象物と対応していることが望ましい．このような領域を対象物に関する知識なしに得ることは難しく，対象物ラベルづけを自動化する際の問題点となっていた．

ところで，風景画像の認識では，領域を生成しても領域固有の特徴量 (形状・大きさ) が対象物ラベルづけに有効に働くとは考えられない．従って，対象物ラベルづけのために領域を生成する必要はないと考えられる．

そこで本論文では，初期的な領域分割処理が不要となる対象物ラベルづけ手法を提案する．本手法は，対象物と画素単位で得られる特徴量との対応関係の知識 (対象物モデル) を用いて各画素を評価し，画素毎に対象物らしさを与え，対象物ラベルづけを行う．対象物に対応する領域は，同一のラベルを持つ画素を集めることにより得られる．従来の手法では，領域生成後に対象物ラベルづけを行っていたのに対し，本手法では，対象物らしさを定義した後に領域生成を行う．このため，初期的な領域分割が過分割になるかという問題を回避できる．

画素の評価を行うための対象物モデルには，対象物の画像上での特徴量 (色・テクスチャ性など) を学習したニューラルネットワークを用いる．この場合，対象物モデルの構築が例示により可能である．また，ニューラルネットワークは任意の写像を近似できるため，データの分布が未知であっ

ても対象物モデルの構築に利用できる．

局所的に得られる特徴量のみで行った対象物ラベルづけでは，画像全体としてみた場合，対象物の位置関係が矛盾していることがある．これを解決するためには，対象物ラベルづけに対象物相互の位置関係の知識を加える必要がある．領域に対して位置関係知識を適用する場合，画像上での領域間の位置関係を定義することが困難であった．画素単位ラベルづけでは，画素が画像上で規則的に並んでいるため，位置関係の定義が容易であり，この種の知識が容易に導入できる．

画素単位対象物ラベルづけの応用として，風景画像検索への適用が挙げられる．対象物ラベルづけの結果を用いて画像の索引づけを行う場合，対象物ラベルづけの正解率が 100 %ではないため，正しい索引づけを行うことは困難である．そこで本論文では，検索者が入力した検索したい画像のスケッチ (検索条件画像) と画素単位対象物ラベルづけされた画像 (インデックス画像) を画素毎に比較して検索を行う．この場合，特定の画素の対象物ラベルづけの誤りが検索結果に影響を与えることがないため，全体として正しい対象物ラベルづけが行われていれば，100 %以下の正解率でも検索は可能である．

以下，2 章では画素単位対象物ラベルづけ手法の特徴と一つの実現例を示す．3 章では画素単位の対象物ラベルづけと領域単位の対象物ラベルづけとの比較検討を行う．4 章では，画素単位対象物ラベルづけに大域的な知識を導入する．5 章では，画素単位対象物ラベルづけの応用として，風景画像検索への利用について考察する．6 章では，結論と今後の課題について述べる．

2　風景画像に対する画素単位対象物ラベルづけ

2.1　従来の対象物ラベルづけ手法とその問題点

画像の対象物ラベルづけを，「一つの領域が一つの対象物に対応するように画像を分割し，各領域に対応する対象物のラベルを付加すること」と定義する．

画像の対象物ラベルづけを得るために考えられるもっとも単純な手法は，画像を信号レベルで領域分割し，得られた領域に対して対象物ラベルづけを行うという方法である (図 1)．ここでいう信号レベルの領域分割とは，対象物に関する知識を用いず，画像を数値データとみなし，データの統計的な類似性や画素の隣接関係に基づいて画素をクラスタリングする手法である．例えば，領域拡張法では，画素値の差が閾値以下の連結した画素の集まりを領域とする．このような方法で対象物と一対一に対応する領域を得ることはほとんど不可能である．これは，例えば異る二つの対象物が信号レベルで示す差異が，一つの対象物内での差異よりも小さいこと，信号レベルの領域分割ではテクスチャが扱いにくいこと，対象物に対応するような “類似性” を定めることが困難であることなどが原因である．

信号レベルの領域分割処理結果はそのままでは対象物と一対一に対応しないという問題があった．これを解決するため，対象物に関する知識を対象物ラベルづけのみでなく，領域の生成にも利用する方法が研究されている．大田 [3] は，初期的な領域分割により得られた領域をルールベースにより対象物ラベルづけし，その結果をもとに領域を併合して画像の対象物ラベルづけを得た．平田ら [4] は，道路画像に対して，初期的に領域・エッジを抽出し解釈を行い，その後，三次元的な知識の拘束を利用して解釈の改善と領域の再生成を行った．処理結果としては，画像の対象物ラベルづ

けと三次元情報を得た．Riseman らの Schema System[5] では，対象物が存在するという仮説に基づきその対象物を抽出する処理を起動し，領域の生成と対象物ラベルづけを行った．対象物の抽出は，信号レベルで領域・エッジを生成し，それらから計算された特徴量とシステムの持つ対象物に関する知識 (対象物モデル) とを比較することにより行っている．

これらの研究では，信号レベルの領域分割処理を用いて画像を初期的に領域分割した後，得られた領域の対象物ラベルと対象物ラベルに基づいた領域の生成を繰り返し，最終的に対象物に対応する領域を得ようとしている (図 2)．初期的な領域は，対象物と一対一に対応している必要はないが，対象物ラベルづけを行うためには，幾つかの前提を満たす必要がある．例えば，初期的な領域が複数の対象物にまたがっていると，計算機上の対象物モデルと一致しないため，対象物ラベルづけが行えない．従って，初期的な領域は，複数の対象物にまたがらないという過分割の仮定を満たす必要がある．また，個々の領域があまりに小さいと，対象物を領域の特徴量でモデル化した場合，対象物モデルとの違いが大きくなる．従って，初期的な領域は対象物となるべく対応していることが望ましい．

このような前提を満たす領域を得るためには，どのような領域分割アルゴリズムをどのような特徴量空間上でどのような閾値で適用するかという問題を扱わなければならない．多様な対象物が現れる風景画像においては，全ての画像に対して上記の前提を満たすようにこれらを決定することは困難である．すなわち，初期的な領域分割を行うこと自体難しい問題である．

過分割な領域分割結果が初期的に得られたとしても，領域を単位として対象物ラベルづけを行う場合，知識を適用するうえで問題が生じる．例えば，テクスチャの知識を利用する場合，領域からテクスチャ性の特徴量を得る必要がある．そのためには，領域がある程度の大きさを持ち，領域内では同一のテクスチャ性を示す必要があるが，信号レベルでの領域分割処

理では，テクスチャを扱いにくく，テクスチャ性の強く現れる部分では，細分化された領域が得られることが多い．そのため，個々の領域から得られるテクスチャ性の特徴は対象物モデルと一致しなくなる．また，位置関係の知識を利用する場合も，領域では形状や配列が様々であるため，領域間の位置関係を適切に定めることが問題となる．さらに，対象物ラベルに基づいて領域を併合・再分割するためには，対象物モデルの他に，対象物ラベルをどのように用いて併合・再分割を行うかという知識を与えておく必要があるという問題がある．

2.2　画素単位対象物ラベルづけの提案

対象物ラベルづけを行うための知識には，対象物の形状・大きさの知識，色・テクスチャの知識，位置・位置関係の知識がある．風景画像中に現れる対象物は主として自然物であり，形状を持たないか不定形である．そのため，形状・大きさの知識は，対象物ラベルづけに有効ではないと考えられる．一方，自然物は色・テクスチャの特徴を強く示すため，これらに関する知識は有効であると考えられる．色・テクスチャの情報は，領域からも得られるが，各画素の近傍を考えれば，画素単位でも得ることができる．位置・位置関係の知識についても，領域固有のものではなく，画素単位でも同じ知識が適用できる．

そこで，本論文では，画素単位で対象物ラベルづけを行う手法 (画素単位対象物ラベルづけ手法) を提案する (図 3)[8]．この手法では，対象物と画素単位で得られる特徴量との対応関係の知識 (対象物モデル) を用いて各画素を評価し，画素毎に対象物らしさを与え，対象物ラベルづけを行う．対象物に対応する領域は，同一のラベルを持つ画素を集めることにより得られる．

領域から得られる特徴量をもとに対象物モデルを構築していた従来の手

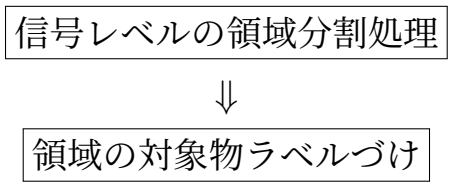


図 1: 単純な手法

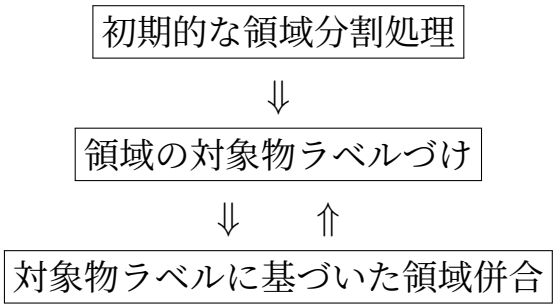


図 2: 従来の手法

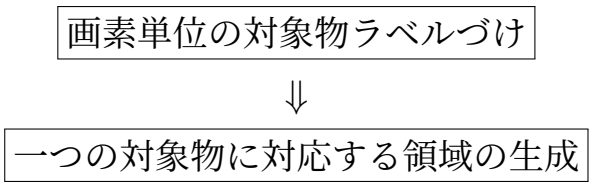


図 3: 画素単位のラベルづけ手法

法に対して，この手法は画素とその近傍から得られる特徴量で対象物モデルを構築する．先に述べたように，風景画像の対象物ラベルづけをするうえで，画素は形状・大きさの特徴以外は領域と同様に得ることができるため，領域単位の対象物ラベルづけと同じ知識を用いて，画素単位対象物ラベルづけが可能である．

画素単位対象物ラベルづけでは，初期的な領域分割なしで対象物ラベルづけを行うことができる．画素は，大きさが一画素の過分割領域とみなすことができるため，初期的な領域分割で生じた過分割の問題は回避できる．また，位置関係の知識を利用する際にも，画素は規則的な配列をしているため，位置関係の定義が容易であり，知識を容易に導入できる．

さらに，領域 (画素) の併合と併合された領域の対象物ラベルづけの繰り返しという手法ではなく，様々な知識を利用して画素単位の対象物ラベルづけを行い，最終的に同一の対象物ラベルを持つ画素の集まりとして，領域を生成するという方法をとるため，領域の併合順序や併合誤りの問題も回避できる．

画素単位対象物ラベルづけの欠点として考えられるのは，扱うデータ量が増えるため，処理時間が増えるという点である．しかし，初期的な領域分割をしない分処理量が減るうえ，リアルタイム性が要求されてはいないため，処理時間はそれほど問題とならない．また，画素単位ラベルづけにより得られる利点は大きく，処理時間が増加しても有用性はあると考えられる．

2.3 画素単位対象物ラベルづけの実現

2.3.1 対象とする風景画像と対象物

本論文では次のような風景画像を処理対象とする．

- 屋外の情景を写した単一のカラー画像.

- 山・木立・湖・草原などの自然物が主な対象物.
- 撮影位置は地面付近．撮影方向はほぼ水平.
- 撮影時間は昼間.
- 特定の対象物の存在は仮定しない.

画像は，地面付近から水平方向に撮影されていることを仮定し，航空写真や天地の反転した画像などは考えない．このため，画像の上側が三次元世界における天に，下側が地に対応すると仮定できる．また，夕焼けや夜・逆光などの照明条件は対象としないので，画像内の対象物の色が極端に偏ることはないと仮定する.

ラベルづけを行う画像中の対象物には，表 1 にあげたものを選んだ．自然物は，対象物自体，形状を持たないか，不定形であり，画像上で色やテクスチャ性を強く示すという特徴がある．従って，自然物の対象物ラベルづけを行う場合，形状よりも画像上での色やテクスチャの特徴を利用する方が妥当である.

風景画像には，これらの対象物以外に，自動車・建物・船・電車などの人工物も現れる．人工物は形状は比較的一定しているが，個体毎に色が異るという特徴がある．このように，自然物とは異る特徴を持つため，同一の方法で自然物と人工物を扱うことは難しい．風景画像では，自然物が主な対象物であるので，本論文では表 1 にあげたもののみを扱う.

対象とする風景画像の例を図 4 に示す.

2.3.2　ニューラルネットを用いた対象物モデル

対象物モデルを保持する方法としては，AI の分野で研究されているルールベースやフレームなど記号的な表現法が多く見られる [3][5][6][11]．記号表現では，人間の持つ論理的な知識は表現しやすく，表現された知識をもとに処理を行うことも容易である．しかし，人間の持つ知識が明確でな

画像 1

画像 2

図 4: 風景画像の例

表 1: 採用した対象物

対象物	具体的内容
SKY	雲以外の青空
CLOUD	曇天の空及び雲
CONCRETE	舗装された路面
LEAVES	緑の葉，草，山肌
SOIL	土
SHADOW	非常に暗い影
MOUNTAIN	遠景の山肌
WATER	映りこみのない水面
D-LEAVES	枯れ草，紅葉，花
ROCK	岩，近景の岩山

い場合には，それを記号的に表現することは困難である．

対象物モデルを色・テクスチャ・位置の特徴で表現する場合，画像から計算されたこれらの特徴を記号的に表現し，それらを用いて対象物モデルを記述する必要がある．しかし，風景画像中の自然物では，人間がこれらの特徴を用いて対象物を正確に表現することができないため，その記号表現を用いて対象物モデルを構築することは難しい．例えば，色・テクスチャ・位置を表現する言葉で，「山」を知らない人に「山」を説明することは困難である．一方，人間が持つ自然物に関する知識のうち，記号的に表現しやすいのは，対象物間の位置関係の知識である．位置関係の知識は，記号的に表現された対象物と画像から得られる位置関係の記号表現を用いて表現される．従って，画像から計算された色・テクスチャ・位置の特徴を記号化せず直接用いて，対象物ラベルという記号的な表現を得ることができれば，記号表現上での処理が行いやすい．

画像から計算される特徴量と対象物ラベルとの写像を得るための方法として，例示による方法がある．この場合，人間が持っている知識が明確でない場合でも，人間は例示を行うのみでよい．人間が暗に利用している知識は，例示されたデータを何らかの手法で解析することにより，計算機上に構築される．例示されたデータから特徴量と対象物との写像を得る方法には，ニューラルネットワークで学習を行う方法や，統計的な手法がある．統計的な手法では対象物毎のデータの分布の型があらかじめわかっている必要がある．また，多次元の入力データにおいて，各次元でデータのもつ性質が異るにも関わらず，全てを同一に扱っている．これに対して，ニューラルネットワークでは，任意の写像を近似することができるため，データの分布が未知でよい．また，どの次元の入力データが有効であるかはネットワーク内部の重みとして学習されると考えられる．

以上の点を考慮して，本論文では，ニューラルネットワークに，各画素毎の特徴量とその画素の正しい対象物ラベルの組を与え学習させる方法

で，対象物モデルを構築する．

ニューラルネットワークには，三層の階層型ネットワークを用いた．階層型ネットワークは，入力層・中間層・出力層からなる．入力層から中間層，中間層から出力層へは任意のノード間に結合があるが，層内の結合や出力側から入力側へのフィードバック結合は存在しない (図 5)．

ノード i は他のノード j からの入力 I_j を，そのノードとの結合の強さ w_{ji} で重みづけした荷重和をとる．さらに内部の閾値 th_i を減じる．ノード i の出力 O_i はこの荷重和 S_i を用いて，

$$\begin{aligned} O_i &= f(S_i) \\ S_i &= \sum_j w_{ji} I_j - th_i \\ f(x) &= \frac{1}{1+\exp(-x/u_0)} \end{aligned}$$

で表される．ノードの出力関数 $f(x)$ には，飽和型の関数を用いている．u_0 は，関数の形を決める定数である．

入力層の各ノードには, 画素から計算されたそれぞれの特徴量を [0.0,1.0] に正規化して与える．出力層は 10 個のノードからなり，各ノードを表 1 のそれぞれの対象物に対応させる．各出力ノードの出力値を，入力された特徴量を持つ画素の対象物らしさの評価値とみなす．

ニューラルネットワークは，いくつかの教師データ (特徴量と，その画素の正しい対象物ラベルの組) を用いて学習させる．学習には誤差逆伝搬法を用いる．これは，出力層の出力と教師データにより与えられる正解との二乗誤差を最急降下法により極小化するものである．入力層と中間層，中間層と出力層のノード間の結合重みの修正量は，図 6 の Δw_{kj}^{ih}，Δw_{ji}^{ho} で与えられる．

学習済のニューラルネットワークは，特徴量と対象物との写像関係を学習しており，対象物モデルをノード間の結合重みとして保持しているとみ

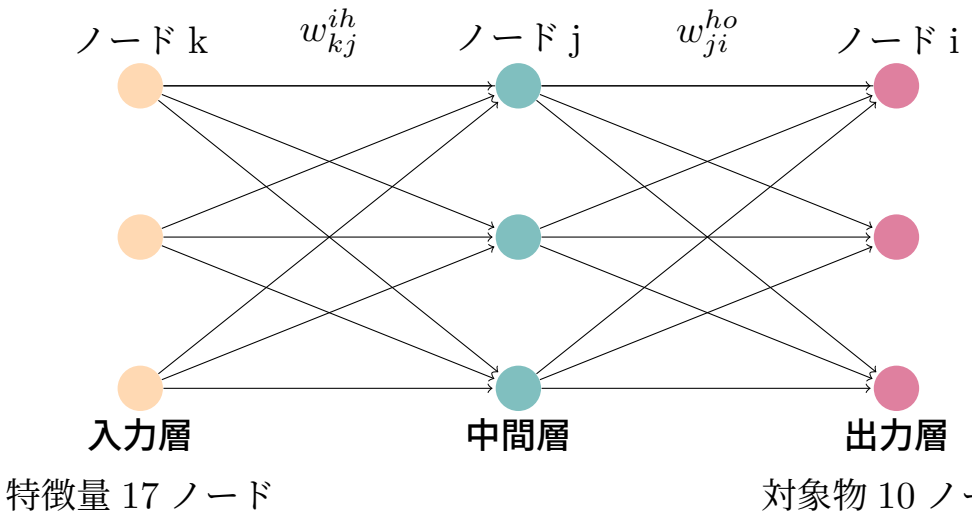


図 5: 階層型ニューラルネットワークの構造

$$
\begin{aligned}
\delta_i^o &= f'(S_i^o)(t_i - O_i^o) \\
\delta_j^h &= f'(S_j^h)\sum_l \delta_k^o w_{jl}^{ho} \\
\Delta w_{ji}^{ho} &= \alpha\delta_i^o O_j^h + \beta\Delta old w_{ji}^{ho} \\
\Delta w_{kj}^{ih} &= \alpha\delta_j^h O_k^i + \beta\Delta old w_{kj}^{ih}
\end{aligned}
$$

δ_i^o	:	出力層のノード i の誤差
δ_j^h	:	中間層のノード j の誤差
w_{ji}^{ho}	:	中間層のノード j と出力層のノード i の結合の重み
w_{kj}^{ih}	:	入力層のノード k と中間層のノード j の結合の重み
Δw_{ji}^{ho}	:	w_{ji}^{ho} の修正量
Δw_{kj}^{ih}	:	w_{kj}^{ih} の修正量
S_i^o	:	出力層のノード i の荷重和
O_i^o	:	出力層のノード i の出力
t_i	:	教師データの出力層ノード i に対する正解の出力
S_j^h	:	中間層のノード j の荷重和
O_j^h	:	中間層のノード j の出力
O_k^i	:	入力層のノード k の出力
$\Delta old w_{ji}^{ho}$	:	w_{ji}^{ho} の以前の修正量
$\Delta old w_{kj}^{ih}$	:	w_{kj}^{ih} の以前の修正量
α	:	学習速度を調節する学習係数
β	:	学習の安定度を調節する安定化係数

図 6: 誤差逆伝搬法

なせる．

この手法を領域分割法としてみると，教師ありの領域分割法とみなせる．教師ありの領域分割法は，教師データから入力の特徴量空間での各クラスタの分布を推定し，未知のデータに対して，そのデータがどのクラスタに属するか判定することにより領域を生成するものである．この方法は，リモートセンシングの分野で盛んに利用されている [9][10]．これらは，衛星から撮影されたマルチスペクトル画像に対して，画素値を特徴量としてニューラルネットワークを学習させ，領域分割を行っている．衛星写真の解析においては，各画素の画素値は地上の利用状況に対応したスペクトル特性をもつため，この情報のみで領域分割が可能である．しかし，風景画像の対象物ラベルづけにおいては，各画素の画素値のみでは十分でなく，テクスチャや画像上での位置の特徴量が必要であると考えられる．また，風景画像では，衛星写真では利用できない三次元的な位置関係の知識が利用できる点で，特徴量のみを用いた単純な教師ありの領域分割法とは異っている．さらに，衛星写真では，画像を地上の利用状況により色分けすることが目標であり特定の “物” を対象としていないのに対し，風景画像では，木や山などの “ 物” も含めて，対象物に対応した領域を生成することが目標であるという違いがある．

3　画素単位と領域単位での対象物ラベルづけの比較

前章で述べたように，画素単位対象物ラベルづけは初期的な領域分割が不要であるとの立場から提案したものである．これは，風景画像中に現れる対象物の形状が，対象物ラベルづけに有効でないという考察から導いたものである．このことを実証するために，本章では，画素単位と領域単位で対象物モデルを実際に構築し，対象物ラベルづけの正解率の比較を行う．さらに，画素単位対象物ラベルづけに関する幾つかの実験結果を示し，その特徴について考察を行う．

3.1　利用する特徴量

対象物ラベルづけは，2.3.2 節で述べたように，ニューラルネットワークに特徴量と正解の対象物ラベルとの写像を学習させて構築した対象物モデルにより行う．対象物モデルの構築に利用する特徴量には，表 2 に示すものを用いる．このうち，色・テクスチャ性・位置の特徴量は，画素単位対象物ラベルづけと領域単位対象物ラベルづけの双方で利用し，形状・大きさの特徴量は領域単位対象物ラベルづけのみで利用する．

色の特徴である H，S，V は，それぞれ色相，彩度，明度を表す．入力画像は RGB の信号で与えられるが，人の感覚により近いといわれる HSV 表色系を用いる．RGB 表色系から HSV 表色系への変換は次式で行う．

$$
\begin{aligned}
H &= \begin{cases} \frac{2(G-min)-(R-B)}{max-min} & (G = max\text{の時}) \\ \frac{4(B-min)-(G-R)}{max-min} & (B = max\text{の時}) \\ \frac{6(R-min)-(B-G)}{max-min} & (R = max\text{の時}) \end{cases} \\
S &= \frac{max - min}{max} \\
V &= max
\end{aligned}
$$

$$\text{ただし } max = \max(R, G, B)$$
$$min = \min(R, G, B)$$

実際の計算は，画素単位ではその近傍の RGB それぞれの平均を，領域単位では同一領域に含まれる画素の RGB の平均を求めた後，上式により変換する．

X,Y は，画像上での座標である．対象とする風景画像は，上下の反転などはないと仮定しているため，画像内での位置の情報も対象物モデルにおいて有効であると考えられる．画素単位では画素の座標を，領域単位では領域の重心の座標を用いる．

テクスチャ性を表す特徴量は，共起行列を用いて求める [15]．共起行列 $P(i, j|d, \theta)$ は，ある領域で画素値 i の画素と方向 θ に距離 d だけ離れた点に画素値 j の画素が現れる確率を表した行列である．画素単位ではその近傍で，領域単位では同一領域で共起行列を計算する．共起行列を求めた後，Energy, Entropy, Inertia の特徴量を

$$\begin{aligned} Energy(d, \theta) &= \sum_{i,j} P(i, j|d, \theta)^2 \\ Entropy(d, \theta) &= \sum_{i,j} P(i, j|d, \theta) \log \frac{1}{P(i, j|d, \theta)} \\ Inertia(d, \theta) &= \sum_{i,j} (i - j)^2 P(i, j|d, \theta) \end{aligned}$$

の式により計算する．Energy はテクスチャの一様性を，Entropy はテクスチャの複雑さを，Inertia はテクスチャのコントラストを表す特徴量である．対象物モデルを構築する際には，距離 $d = 1$ とし，方向については $\theta = 0, 45, 90, 135^\circ$ の 4 通りを用いる．

領域に対する形状の特徴量には，慣性モーメントから計算できるものを用いる．Iangle は慣性主軸と画像の x 軸のなす角度で，領域の細長い方向の傾きを表す．Iratio は重心を通る慣性主軸方向の慣性モーメントとそれ

に直角な方向の慣性モーメントの比で，領域の細長さを表す．ISratio は重心周りの慣性モーメントと領域の面積の比で重心周りの対称性を表す．これらの算出式を図 7 に示す．

領域の大きさの特徴量には，領域の面積 (Size)，画像上での横方向の大きさ (Width) および縦方向の大きさ (Height) を用いる．

3.2 対象物モデルの構築

3.2.1 実験の準備

特徴量計算に用いる近傍の大きさ

画素単位の場合の特徴量計算では，ノイズの影響を押さえ，またテクスチャ性の評価を行うため，画素の近傍を計算対象とする．近傍を小さくとるとテクスチャ性が現れにくくなり，大きくとると近傍内に複数の対象物が含まれやすくなる．近傍の大きさを決定するために，7, 15, 31 のそれぞれの大きさで特徴量計算を行い，対象物モデルを構築して正解率を比較したが，これらの間で大きな差は見られなかったので，計算量が少なくて済む 7×7 近傍を採用した．以下では，7×7 近傍を用いた場合の結果を示す．

領域の生成法

領域単位の対象物ラベルづけでは，初期的な領域の生成が必要である．画像の対象物ラベルづけの自動化を目標とする場合，初期的な領域の生成も自動化する必要がある．ここでは，画素単位と領域単位での対象物ラベルづけの結果を比較することが目的であるので，信号レベルの領域分割処理の閾値を人手により画像毎に適切に定めて初期的な領域分割を行った．具体的には，

1. エッジを保ったスムージング
2. 輝度を用いた領域拡張法による領域分割

表 2: 対象物モデルに用いた画像の特徴量

特徴量の分類	名称
色	H, S, V
画像上での位置	X, Y
テクスチャ性 (４方向)	Energy, Entropy, Inertia
形状	Iangle, Iratio, ISratio
大きさ	Size, Width, Height

$$
\begin{aligned}
I_{xx} &= \iint (x - x_g)^2 dxdy \\
I_{yy} &= \iint (y - y_g)^2 dxdy \\
I_{xy} &= \iint (x - x_g)(y - y_g)\ dxdy \\
\theta &= \frac{\arctan(2I_{xy}/(I_{xx} - I_{yy}))}{2} \\
I_{max} &= I_{xx}\cos^2\theta - 2I_{xy}\sin\theta\cos\theta + I_{yy}\sin^2\theta \\
I_{min} &= I_{xx}\sin^2\theta - 2I_{xy}\sin\theta\cos\theta + I_{yy}\cos^2\theta \\
I &= I_{max} + I_{min} \\
Iangle &= \theta \\
Iratio &= I_{min}/I_{max} \\
ISratio &= Size^2/2\pi I
\end{aligned}
$$

x_g : 領域の重心の x 座標

y_g : 領域の重心の y 座標

$Size$: 領域の大きさ

図 7: 形状の特徴量

3. 孤立点除去

4. 面積を閾値とする領域併合

の手順で処理を行い，ほぼ過分割の領域分割を得た．領域分割結果の例を図 8 に示す．

実験に用いる画像

実験には 66 枚の風景画像を用いた．画像の大きさは 256×256 画素で，写真から 100dpi の精度でスキャナを用いて取り込んだ．スキャナの特性などの影響を取り除く色補正処理は行っていない．

教師用，評価用データの作成

まず，すべての画像からいくつかの点を選び，その点に対する正解の対象物のラベルづけを人間が行った．選んだ点の内訳を表 3 に示す．その後，画像を 33 枚ずつ A,B グループの二組に分けた．上で選んだ点のうち，A グループの画像の点を画素単位対象物ラベルづけの教師用データセット pix-teach とし，B グループの点を評価用データセット pix-eval とした．即ち，教師用データセットと評価用データセットでは，別の画像を用いている．

領域単位対象物ラベルづけのためのデータは，上で人間が選んだ点を用いて作成した．すなわち，人間が与えた正解の対象物ラベルを，その点が含まれる領域に対する対象物ラベルとし，教師用データセット reg-teach と評価用データセット reg-eval を作成した．この際，重複したデータは除いた．また，ここで利用する領域分割結果は完全には過分割となっていないため，同一の領域に異る正解の対象物ラベルがつけられたものがあったが，これは除いた．そのため，データ数は画素単位の場合よりも少なくなっている．データの内訳を表 4 に示す．

3.2.2　ニューラルネットワークの学習

画素単位，領域単位対象物ラベルづけのための対象物モデルを構築するために，教師用データセット pix-teach, reg-teach を用いて，ニューラルネットワークの学習を行った．

学習は基本的には誤差逆伝搬法であるが，収束を早めるために一部改良した．誤差逆伝搬法では，学習係数 α を大きくとると学習が早く進む．しかし，学習係数は教師データの分散と関連しており，分散が大きい場合，学習係数を小さくとらなければ，学習が収束しない [14]．そのため，学習の初期の段階では学習係数を大きく，安定化係数を小さくとり，最急降下法による修正量を優先して，二乗誤差が最もよく減るように修正を行う．学習の進んだ段階では学習係数を減らし安定化係数を増やして，学習を収束させる方法をとった．具体的には，学習係数の初期値を α_{ini}，安定化係数の初期値を β_{ini} とし，最終値 α_{fin}，β_{fin} になるまで，誤差逆伝搬を一回行うごとに，きざみ幅 α_{step} づつ学習係数を減らし，β_{step} づつ安定化係数を増やすという方法を用いた．

また，飽和型関数への入力の絶対値が大きくなると，その微分値が小さくなり，最急降下法による修正量も小さくなる．そのため，飽和型関数への入力の絶対値を定数 $Smax$ 以内に制限し，これを越える場合は，正の場合は $Smax$ に，負の場合は $-Smax$ を入力とした．

中間層のノード数は，多い程，教師データをよく近似できるが，逆に学習にかかる時間は長くなる．また，教師データをノイズも含めて忠実に学習することは，教師以外のデータに対する正解率の向上につながらない．そのため，本論文の実験では，試行錯誤的に中間層数を変えた幾つかのネットワークを学習させ，結果の良かったものを採用した．

ネットワークの結合の初期値は [-0.5,0.5] の範囲で乱数により定めた．

学習は，ニューラルネットの出力と教師データの出力との間のデータあたりの平均二乗誤差が 0.2 になるまで行った．シグモイド関数の入力は，$Smax = 2.0$ に制限した．実験に用いた学習係数などのその他の定数を表 5 に示す．実験では，結合の初期値を変えた 9 個のネットワークで学習を行った．実験結果として示しているものは，その中で評価用データに対する一位正解率が最も高かったネットワークを用いた場合のものである．

3.3 実験結果および考察

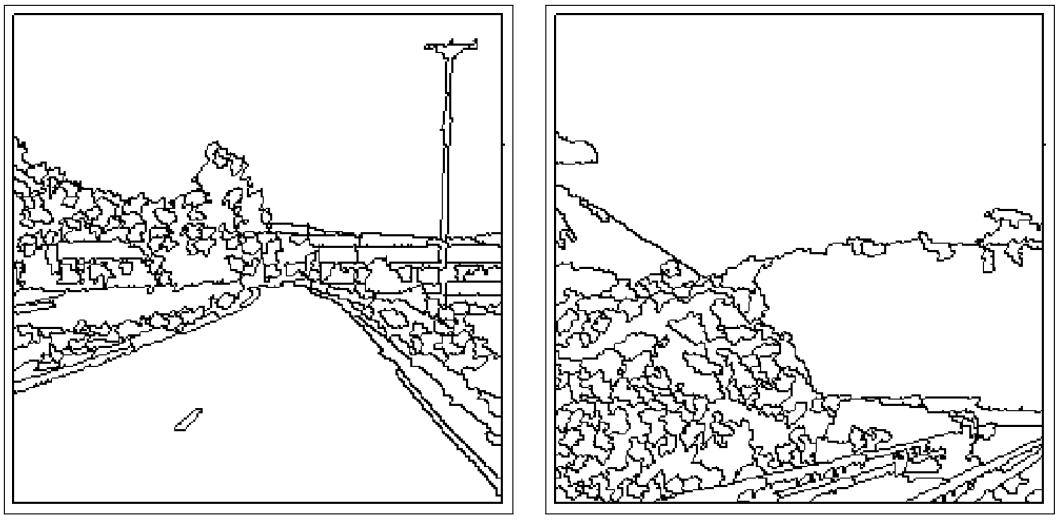
3.3.1 画素単位と領域単位での正解率の違い

画素単位，領域単位対象物ラベルづけの正解率を表 6(1), (2) に示した．表は，教師用，評価用データセットにおいて，ネットワークの出力の上位一，二，三位以内に正解が含まれているデータの割合を示している．以下，評価用データセットに対する一位正解率で評価を行う．

画素単位と領域単位を比較すると，画素単位対象物ラベルづけの方が高い一位正解率を示している．一般にニューラルネットワークでは，教師データの数が多い程教師データの分布を正しく学習するため，学習後のネットワークの正解率は高くなる．表 6(1), (2) では，教師用データの数が異るので画素単位対象物ラベルづけの方が高い正解率が得られると結論することはできないが，この結果から，画素単位で得られる特徴量で対象物ラベルづけが可能であり，領域単位と比べて正解率が極端に低くなることはないことが示された．

表 6(3) は，領域単位対象物ラベルづけで，表 2 にあげた形状・大きさの特徴量を除いて学習を行った場合の正解率である．表 6(2) と比較すると，一位正解率の差はほとんどない．従って，表 2 にあげた形状・大きさの特徴量は，対象物ラベルづけに有効に働いていないといえる．

表 6(3) の結果のみでは，一般に形状・大きさが対象物ラベルづけに有



画像 1　　画像 2

図 8: 領域分割結果

表 3: 画素単位のデータセットの内訳 (対象物毎のデータ数)

対象物名	pix-teach	pix-eval
SKY	97	79
CLOUD	52	93
CONCRETE	116	101
LEAVES	195	251
SOIL	12	37
SHADOW	59	28
MOUNTAIN	102	117
WATER	194	163
D-LEAVES	83	36
ROCK	65	65
Total	975	970

表 4: 領域単位のデータセットの内訳 (対象物毎のデータ数)

対象物名	reg-teach	reg-eval
SKY	16	9
CLOUD	10	21
CONCRETE	45	25
LEAVES	116	171
SOIL	9	24
SHADOW	16	9
MOUNTAIN	27	32
WATER	45	43
D-LEAVES	56	13
ROCK	40	35
Total	380	382

表 5: 実験に用いた定数

定数		値
学習係数初期値	α_{ini}	1.0
学習係数きざみ幅	α_{step}	0.001
学習係数最終値	α_{fin}	0.6
安定化係数初期値	β_{ini}	0.0
安定化係数きざみ幅	β_{step}	0.005
安定化係数最終値	β_{fin}	0.98
中間層数		11

効でないとは結論できない．しかし，対象物の形状が一定でないことを考えれば，形状・大きさの特徴量を除いても正解率が変わらないという結果は妥当である．また，形状・大きさが対象物ラベルづけに有効でないと仮定すれば，画素単位でも正解率が低くならないことも説明できる．以上のことから，画素では得られない形状・大きさの特徴量は対象物ラベルづけに有効ではなく，従って，画素単位対象物ラベルづけは妥当な処理であるといえる．

3.3.2　有効な特徴量の評価

表 2 の特徴量は，色・テクスチャ・位置を表すために一般に用いられているものを選んだ．これらの特徴量が対象物ラベルづけにどの程度有効であるかは，選択の時点では不明であった．本節では，これらの特徴量の有効性を，ニューラルネットを用いて対象物モデルを構築し，評価用データセットに対する一位正解率 (以下認識率と呼ぶ) を比較することにより評価を行う．もし，ある特徴量を除いても，認識率が変化しないならば，その特徴量は対象物ラベルづけに不要である．一方，認識率が低下する場合は，その特徴量が，少なくとも幾つかのデータを対象物ラベルづけする際に有効に働いており，他の特徴量では補うことができない情報を持っていると判断できる．従って，特徴量を除いたときの認識率の低下の度合がその特徴量の有効性を表しているとみなせる．

学習は，結合の初期値を変えた 9 個のネットワークに対し，入力のうち幾つかの特徴量を除いて，20000 回を限度に収束するまで (ネットワークの出力値と教師用データセットの与える正解との平均二乗誤差が 0.2 になるまで) 行う．除いた特徴量とその特徴量を除いた時の評価用・教師用データセットに対する最良の一位正解率，収束したネットワーク数，収束したネットワークでの収束までの学習回数の平均を表 7 に示す．表中の

表 6: ニューラルネットによる対象物ラベルづけの正解率 (単位%)

(1) 画素単位

	一位	二位以内	三位以内
教師用データセット	87.6	96.3	98.9
評価用データセット	73.9	88.8	96.3

(2) 領域単位

	一位	二位以内	三位以内
教師用データセット	88.1	94.6	98.0
評価用データセット	68.5	82.1	92.9

(3) 領域単位

(形状・大きさの特徴量を除く)

	一位	二位以内	三位以内
教師用データセット	88.1	94.9	99.4
評価用データセット	69.0	86.1	96.2

NONE は，全ての特徴量を使った場合の結果，TXT はテクスチャ性の特徴量全てを除いた場合の結果である．また，9 個のネットワークが全て収束しなかったものについては，平均収束回数は示していない．学習のパラメータは表 5 に示したものを用いた．収束したネットワークでは，教師用データセットに対する一位正解率はほとんど同じになっている．

色の特徴では，S を除いても認識率はそれほど低下せず，収束もそれほど変わらない．V を除いても，認識率はそれほど変化しないが，収束は悪くなる．H を除くと 7.1 %認識率が低下し，収束も悪くなる．また，HSV を同時に除くと学習が収束しなくなり，認識率も大きく (23.3 %) 低下する．従って，色の特徴量は対象物ラベルづけで大きな役割を果たしているといえる．また，色の特徴の中では，色相の特徴量 H が他の S,V よりも有効に働いている．これは，SKY と CLOUD のように，色相により識別される対象物が多いことを反映している．

位置の特徴では，X を除いても，認識率は変化していない．一方，Y を除くと 10.5 %認識率が低下する．また，XY を同時に除いても，Y のみを除いた場合と認識率は同程度である．従って，X は対象物ラベルづけにほとんど寄与していない．対象とする風景画像は地面付近から水平方向に撮られたものであるので，画像の上方には三次元世界での天に，下方には地に対応する対象物が現れるのに対し，左右方向ではどの位置にどの対象物が現れるということはない．そのため，位置の特徴では Y のみが対象物ラベルづけに有効である．

テクスチャの特徴量を全て除くと，認識率が 5.3 %低下するが，これは色・位置の特徴を全て除いた場合に比べ低下の度合が低い．また，収束もそれほど悪くならない．LEAVES や ROCK などは，テクスチャを強く示すが，色のみでもこれらを識別することは可能である．従って，テクスチャの特徴を与えなくても認識率はそれほど低下しないと考えられる．

表 7: 幾つかの特徴量を除いて構築した対象物モデルの一位正解率

除いた特徴量	一位正解率		収束ネット数	平均学習回数
	評価	教師		
NONE	73.9	87.6	9	1596
S	70.8	88.0	8	7283
V	70.1	88.3	4	15777
H	66.8	87.8	3	14021
HSV	50.7	66.7	0	—
X	73.6	88.7	9	1463
Y	63.4	88.3	2	18049
XY	62.8	87.6	1	13204
TXT	68.6	89.2	8	4874

3.3.3　構築された対象物モデルに関する考察

対象物ラベルの判定法

ニューラルネットの出力層では，あるノードの出力値が 1.0 の場合，与えられた特徴量を持つ画素にはそのノードに対応する対象物ラベルが正解であり，0.0 の場合，そうでないことを示している．実際のノードの出力値は，(0.0,1.0) の間の値をとるため，この段階で対象物ラベルを決定するためには，

> 「出力層で，出力値が閾値以上かつ最大のノードに対応する対象物ラベルを，そのとき入力層に与えた特徴量をもつ画素に付加する対象物ラベルとする」...(A)

などの方法をとる必要がある．(A) の方法をとった場合，どのような結果が得られるか，ネットワークの出力値と正解率との関係から考察を行う．

(A) の方法で対象物ラベルを決定した時の正解率が図 9 である．出力値が閾値以上のノードが存在しない場合は，対象物ラベルは不明であるとした．横軸に閾値をとり，対象物ラベルが不明と判定されたデータの割合をグラフ Unknown に，不明と判定されたデータを除き，判定結果が正解であったデータの割合をグラフ Right に，誤りであった割合をグラフ Wrong に示した．

閾値を高くとる程，正解率は高く，誤り率は低くなる．従って，閾値を調整することにより，得られる対象物ラベルの正解率を制御することができる．しかし，不明と判定されるデータ数の増加率が，正解率の上昇率より常に大きいため，この間で閾値を細かく設定しても，ラベルづけが行われるデータの数が減る割には正解率は上がらない結果が得られる．従って，上にあげた (A) の方法をとる場合，細かな閾値設定は無意味であり，閾値としては 0.0，0.99 およびその中間の値の三種類程度を考慮すれば十

分である．とくに閾値 0.0 の場合は，「出力値が最大のノードに対応する対象物ラベルを採用する」という判定法になり，全ての画素がラベルづけされる．閾値 0.99 の場合は，ある程度確実なラベルづけを行えるが，不明と判定されるデータ数も大きく増加するため，何らかの処理により不明部分のラベルづけを行う必要がある．

評価値としてのニューラルネットの出力

画像の画素単位対象物ラベルづけでは，対象物ラベルづけのための様々な知識を導入し，最終的に対象物に対応する領域を求める．そのため，ニューラルネットの出力値を，それぞれの対象物ラベルに対する評価値とみなして処理を行うことが考えられる．そのためには，出力値が高い程その対象物である確率が高く，低い程その対象物である確率が低くなっていなければならない．ニューラルネットでは，0.0 と 1.0 の出力値のみを学習させているため，これら以外の出力値でどのような傾向があるかは自明ではない．

図 10 は，ネットワークの出力値とその時の正解の割合の関係を示している．横軸にはネットワークの出力値 O を 0.1 刻みにとり，縦軸にはネットワークの出力値が $(O, O + 0.1)$ の範囲にあるデータに対する正解の割合をとっている．ネットワークの出力値とともに正解の割合が増加していることから，ネットワークの出力値を評価値として利用してよいことがわかる．

画像への適用

図 4 の画像の各点について，表 2 の特徴量を計算し，学習済のニューラルネットワークで評価した．ニューラルネットの出力値から，(A) の判定法により対象物ラベルづけを行った結果を，図 11 に示す．

画像 1 で閾値 0.0 とした場合，全ての画素が対象物ラベルづけされてい

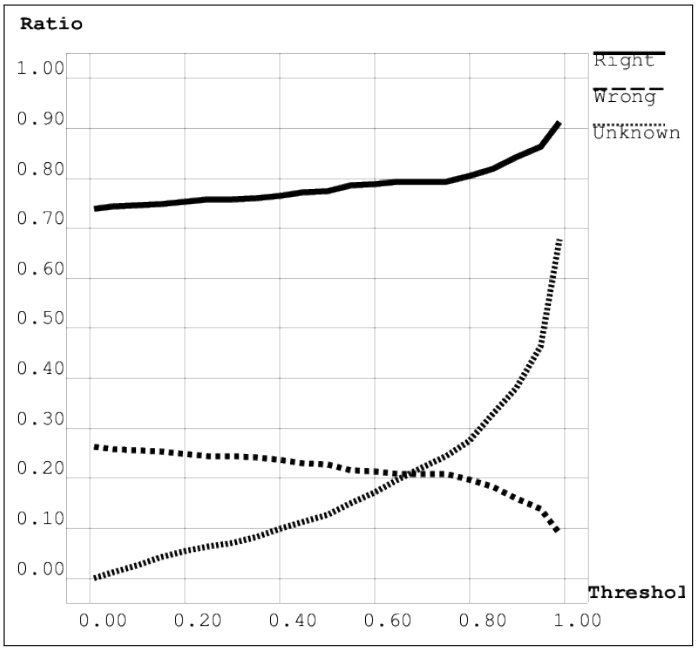


図 9: 閾値と正解・誤り・不明率

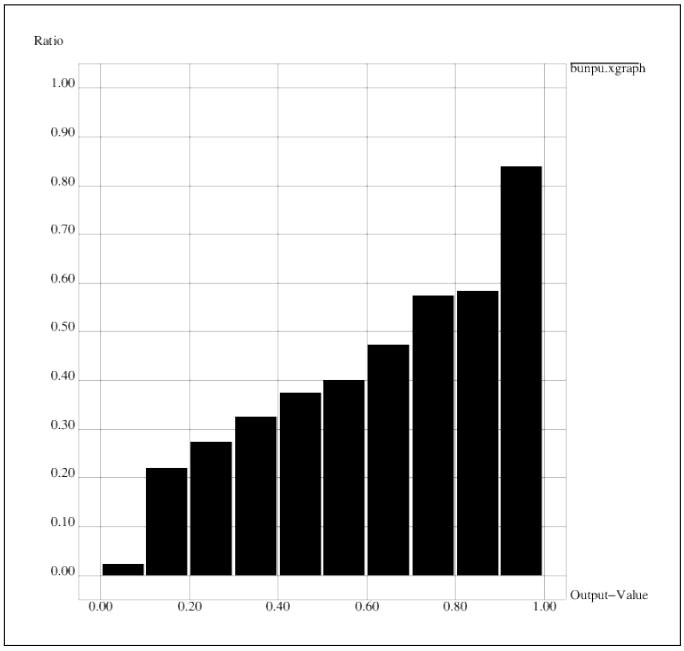


図 10: 出力値と正解率

るが，CONCRETE を WATER とラベルづけしている部分がある．閾値 0.5 では，この部分の対象物ラベルは不明と判定されており，ラベルづけされた画素での正解率は高くなっている．閾値 0.99 では，約 85 %の画素の対象物ラベルが不明と判定されている．しかしこの場合でも正解率は 100 %ではなく，LEAVES を ROCK とラベルづけしている部分がある．閾値の設定により，対象物ラベルの誤りを減らすことができるが，画像によってはかなりの誤りが残ることがあり，何らかの後処理が必要である．

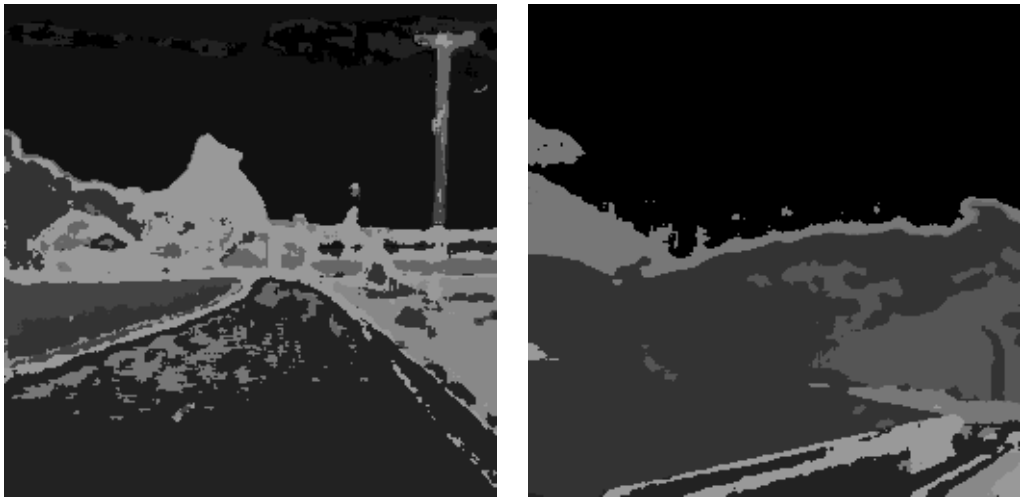
画素単位対象物ラベルづけでは，各画素点毎に特徴量の計算とその評価が必要である．処理時間は HP9000/720 を用いて，画像から全画素の特徴量を計算する処理に 890.95 秒，特徴量から全画素の対象物ラベルを得るための処理に 14.05 秒かかる．処理時間の大半は，特徴量の計算処理であり，特にテクスチャの特徴量の計算時間がほとんどである．テクスチャの特徴は対象物ラベルづけにそれほど有効に働いていないので，この部分を省略すれば，高速に特徴量が計算できる．また，特徴量の計算，対象物ラベルづけ処理はともに画素点毎に独立に行えるので，並列処理により処理時間が改善できる．

誤り方の傾向

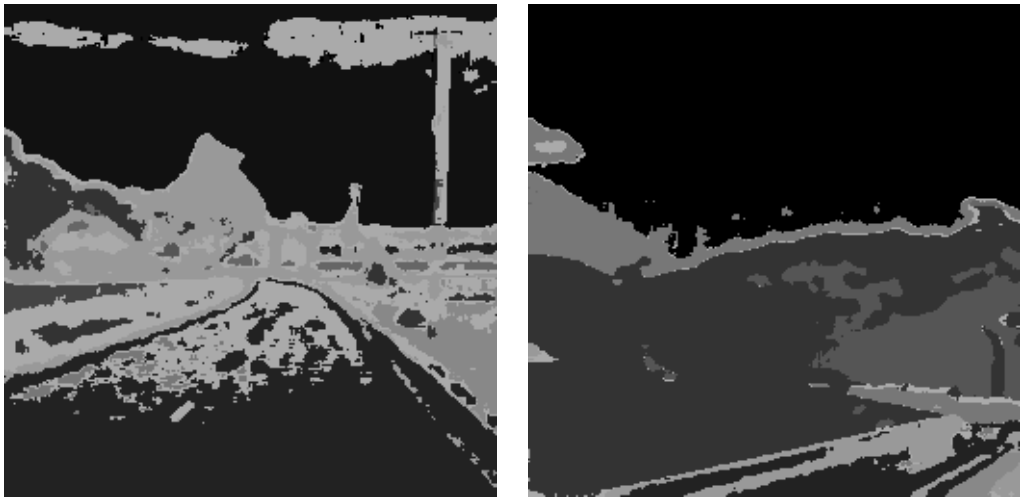
構築したニューラルネットは，評価用データに対して一位正解率で 7 割程度の性能を示した．この結果を改善するためには，どのような誤りが多いかを調査する必要がある．表 8 は，(A) の判定方法で，閾値を 0.0 とした場合の誤り方を示したものである．例えば，正解が CLOUD である時，SKY に対応するノードの出力値が最大であったものが 7 個，CLOUD が最大であったデータが 80 個というように読む．

誤りが各対象物に対するデータの 15 %以上である組み合わせは，

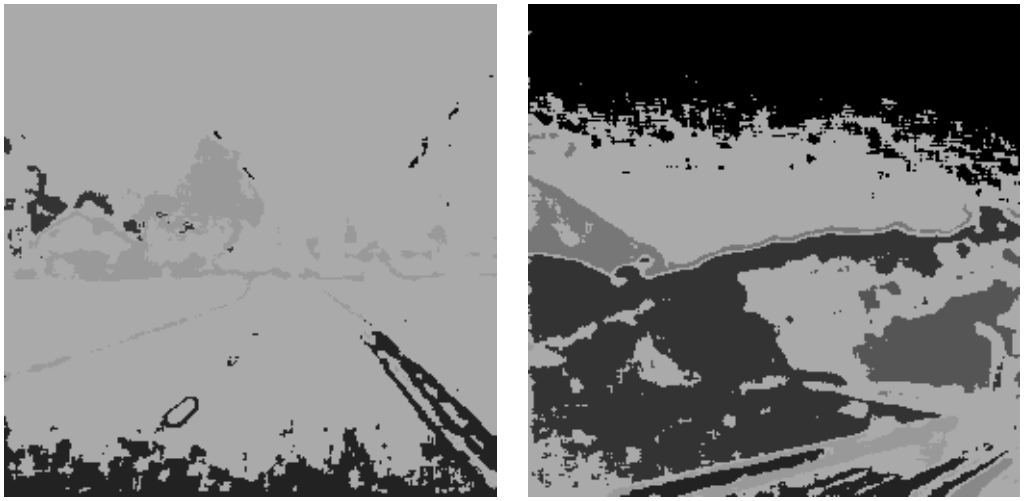
CONCRETE を WATER に，SOIL を D-LEAVES に，
ROCK を WATER に，MOUNTAIN を WATER に，SOIL



閾値 0.0



閾値 0.5



閾値 0.99

画像 1　　画像 2

図 11: 画像の対象物ラベルづけ

を LEAVES に，ROCK を SOIL に，ROCK を MOUNTAIN に，D-LEAVES を LEAVES に，

にラベルづけしたものである．

誤りは，基本的に局所的に得られる特徴量のみを用いていることが原因で生じている．細かく分類すると，誤り方には二種類ある．

一つは，色・テクスチャ性の特徴量を個々の画素の近傍のみを利用して計算していることが原因で生じるものである．人間はこれらの特徴量を絶対的な量として知覚しているのではなく，周囲との相対的な量で知覚している．そのため，例えば，枯れ野原で局所的に緑の葉があっても枯葉と判断するが，同じ緑の葉が緑の草原にあれば，緑の葉とみなす．このように，対象物は絶対的な特徴量のみで定まるものではないので，対象物モデルを構築する際には，特徴量の相対的な関係も考慮する必要がある．

もう一つは，個々の画素を独立に処理していることが原因で生じるものである．例えば，道路と曇天の下の凪いだ水面では，人間でも，かなり広い範囲を見ても識別できない．これは人間が画像全体を見て，他の対象物との関係をもとにして判断しているためであると考えられる．次章では，位置関係知識を導入して，この誤りの一部を改善する処理について述べる．

表 8: ニューロによる対象物ラベルづけ結果と正解

正解 (データ数)	ネットワークの出力									
	SKY	CLD	CNC	LVS	SOI	SHD	MNT	WTR	D-L	RCK
SKY (79)	78	0	0	0	0	0	0	1	0	0
CLOUD (93)	7	80	0	0	0	0	3	3	0	0
CONCRETE (101)	0	0	73	5	1	0	2	16	2	2
LEAVES (251)	0	0	4	202	0	6	9	7	2	21
SOIL (37)	0	0	2	7	6	0	0	0	22	0
SHADOW (28)	0	0	0	0	0	24	4	0	0	0
MOUNTAIN (117)	13	5	0	1	0	0	73	25	0	0
WATER (163)	1	0	17	0	0	0	4	129	0	12
D-LEAVES (36)	0	0	0	11	0	1	0	0	24	0
ROCK (65)	0	0	0	0	11	0	10	15	1	28

4　画素単位対象物ラベルづけにおける位置関係知識の利用

本章では，対象物ラベルづけに位置関係知識を導入し，局所的に得られる特徴量を用いた対象物ラベルづけの改善を目指す．位置関係知識は，対象物間に成り立つ位置関係を規定したもので，画像全体に対して対象物ラベルづけを行う際に，ラベルづけの拘束条件として働く．この知識を画素単位対象物ラベルづけで利用する方法を実現した．

4.1　位置関係知識の特徴

位置関係知識は，色・テクスチャ・位置の知識とは異る特徴を持つ．

前章で利用した色・テクスチャ・位置の知識は，画素の近傍から計算される特徴量とその特徴量を持つ画素の対象物ラベルとの対応関係の知識であり，個々の画素に対して独立に適用される．そのため，各画素の対象物ラベルは，他の画素の対象物ラベルとは無関係に定まる．一方，位置関係知識は，ある位置関係にある画素につけられる対象物ラベルが満たさなければならない拘束条件として働く．そのため，それのみでは対象物ラベルづけは行えないが，個々の画素に付加された対象物ラベルの整合性を検査することができる．一般に，ある知識を利用して得られた結果が誤っていた場合，その知識のみでは誤りを訂正することはできない．色・テクスチャ・位置の知識と位置関係知識は異る知識であるため，位置関係知識を導入することにより，前章で行った対象物ラベルづけを改善することができる．

位置関係知識はまた，知識が人間により明確に定義できるという特徴がある．色・テクスチャ・位置の特徴量と対象物ラベルとの対応関係の知識は，人間が暗に用いているもので，人間自身が明確に表現できないもので

ある．そのため，人間が例を与え，計算機がそれを学習することにより対象物モデルを構築する必要があった．一方，位置関係知識は比較的明確であるので，人間が知識を表現できる．そのため，例示を用いる必要はない．逆にもし，色・テクスチャ・位置の知識と同様な方法で対象物モデルを構築し，利用しようとすると，画像全体の正しいラベルづけを学習データとして与える必要がある．このようなデータを作ることは困難であるうえ，ネットワークの規模が大きくなるため，学習を行うことが困難になる．従って，位置関係知識を利用する際には，2.3.2 節とは別な，より効率的な枠組を利用する方がよい．

4.2 利用する位置関係知識

本論文で用いる位置関係知識は，カメラと三次元世界との位置関係に関する前提から得られるものである．対象とする風景画像には，

- 画像は地面近くからほぼ水平方向に撮影されている．
- 上が天，下が地にあたる．

という前提をおいている．そのため，例えば画像上で道路が空より上に来ることはない．この性質を知識として表すと，上下方向に関するものと左右方向に関するものの二種類で表現できる．

上下方向の知識は，画像上のある列をとったときの対象物の現れる順番に関するものである．先の例を一般化すると

1. 画像上では，天に存在する対象物が最も上に，地に存在する対象物が最も下に現れる．
2. 天と地の間では，遠くにある対象物ほど上に現れる．

と表現できる．1. の知識は画像の撮影条件から得られるものであり，2. の知識は「近くの対象物は遠くの対象物を隠す」ことから得られるものである．

左右方向の知識は，水平線・地平線に関する知識である．一般に地面以外の対象物は，三次元世界で地面より上に存在する．対象物が近くにある場合，カメラの高さによっては，対象物の向こう側の地面が見え，画像上で対象物より上に地面が現れることがある．しかし，対象物が遠くにある場合はカメラの位置が地面に近いことを仮定しているので，対象物の向こう側の地面が見えることはない．また，空は，常に水平線・地平線より上に存在する．従って，画像上の左右方向についての知識は，遠くの対象物や空と地面が同一線上に現れることはないと表現できる．特に，空と地面が接している場合，画像上のこの行が水平線または地平線になる．

これらの知識を扱うために，対象物ラベルに地・中景・遠景・天の位置属性をつける．それぞれの位置属性を持つ対象物ラベルを表 9 に示す．以降，混乱のない場合は，地・中景・遠景・天の位置属性をもつ対象物ラベルのことをそれぞれ地・中景・遠景・天の対象物ラベルと呼ぶ．天は空に対する位置属性で，SKY，CLOUD が対応する．天は地面より上にある対象物の背景になる．遠景は遠くにある高さのある対象物に対する位置属性で，MOUNTAIN が対応する．遠景より遠くにある高さのある対象物は遠景の対象物に隠されて画像上では見えない．中景は遠景より前にある地面以外の全ての高さのある対象物に対する位置属性である．地は地面に対する属性である．LEAVES，D-LEAVES，SHADOW，ROCK は，中景と地の両方の位置属性を持っている．これは例えば，LEAVES は，草，木の葉，近くの木におおわれた山を表すが，このうち草は地に分類され，木の葉，近くの木におおわれた山は中景に分類されるためである．

この位置属性により本論文で用いる知識を表すと，

上下方向： 対象物は画像上で，上から天，遠景，中景または地の順に並んでいる．

左右方向： 天，遠景と地の対象物は，画像上の同一行に存在しない．

となる．

4.3 位置関係知識を利用した処理

前節で示した位置関係知識は，画像全体の対象物ラベルづけが満たさなければならない拘束条件として働く．対象物ラベルづけを正しく行うことができれば，条件は満たされる．一方，条件を満たさない対象物ラベルづけは誤りを含んでいるので，条件を満たすように対象物ラベルを変更する必要がある．この場合，どの対象物ラベルを変更すれば正しい対象物ラベルづけを行うことができるかが問題となる．一般に，二つの画素 A,B に対する対象物ラベルが位置関係知識に反していた場合，画素 A の対象物ラベルのみを変更しても，画素 B の対象物ラベルのみを変更しても，両方の対象物ラベルを変更しても，条件を満たす対象物ラベルづけが行える．これらのうちのどの対象物ラベルづけを行えば正しい対象物ラベルが得られるかは自明ではない．

本論文では，画像全体に対する対象物ラベルづけのうち，位置関係知識を満たすもので，対象物ラベルの評価値の和が最大のものを求める．対象物ラベルの評価値は，各画素に対し，その画素がその対象物ラベルをもつ確率が高いほど大きな値をとるものであり，前章で構築した学習済のニューラルネットの出力値を用いる．3.3.3 節で示したように，ニューラルネットの出力値は評価値とみなしてよい．得られる画像全体に対する対象物ラベルづけ結果は，位置関係知識を満たすものの中で評価値の和が最大であるという意味で最適なものである．

画像全体の最適な対象物ラベルづけを求めるためのアルゴリズムは三段階からなる．第一段階で，画像の各列毎に，上下方向の知識を満たす最適な対象物ラベルづけを求める．第二段階で，その中で左右方向の知識を満たす最適な対象物ラベルづけを求める．第三段階で，各画素の対象物ラベ

表 9: 対象物と位置属性

位置属性	対象物名
地	CONCRETE, SOIL WATER LEAVES,D-LEAVES SHADOW, ROCK
中景	LEAVES, D-LEAVES SHADOW, ROCK
遠景	MOUNTAIN
天	SKY, CLOUD

ルを出力する．

上下方向の知識は，画像上で上から天，遠景，中景または地の順に対象物ラベルが現れることを示している．ここで，カメラの仰角から定まる画像上での地面の開始位置を考える．地面の開始位置は，地と天以外の対象物が存在しない場合には，画像上での地平線・水平線の位置である．一般には，地平線・水平線は中景の対象物により隠されている可能性がある．従って，ある行 k が地面の開始位置であると仮定すると，k 行目より下の画素の対象物ラベルは，中景または地の対象物ラベルでなければならない．また，k 行目より上では，地以外の対象物ラベルが上から順に天，遠景，中景の順でつけられなければならない．ただし，天，遠景，中景の対象物ラベルのうち一つないし二つは実際につけられていなくてもよい．従って，第一段階の処理では，このような対象物ラベルづけの中で最適なものを求めればよい．また，第二段階で画像の各行を地面の開始位置と仮定した時の対象物ラベルづけの評価値が必要となるので，各行毎に，その位置が地面の開始位置であると仮定し，最適な対象物ラベルづけを求める必要がある．

列 j で，k 行目が地面の開始位置であると仮定した時の，評価値が最大になる対象物ラベルづけの求め方は以下のものである (リスト 12 参照)．

列 j を上から走査して行く．$i-1$ 行目 $(1 \leq i \leq k-1)$ までの評価値の和のうち，天の対象物ラベルのみでラベルづけをした場合の最大値を SUM_ten, 遠景の対象物ラベルを含み, 天と遠景の対象物ラベルでラベルづけをした場合の最大値を SUM_far，中景の対象物ラベルを含み，天と遠景と中景の対象物ラベルでラベルづけをした場合の最大値を SUM_mid とする．また，列 j の i 行目の画素での天の対象物ラベルの最大の評価値を EVL_ten[j][i] とする．遠景，中景，地についても最大の評価値をそれぞれ EVL_far[j][i]，EVL_mid[j][i]，EVL_chi[j][i] とする．i 行目では，次の式により SUM_ten，SUM_far，SUM_mid を更新する．

```
SUM_ten = SUM_ten + EVL_ten[j][i]
SUM_far  = max(SUM_ten, SUM_far) + EVL_far[j][i]
SUM_mid  = max(SUM_ten, SUM_far, SUM_mid) + EVL_mid[j][i]
```

これにより，i 行目についても，SUM_ten, SUM_far, SUM_mid がそれぞれ，天のみ，遠景を含み天と遠景のみ，中景を含み天と遠景と中景のみでラベルづけした場合の評価値の和の最大値になる．

k 行目では，SUM_ten, SUM_far, SUM_mid の中で，最大のものを SUM_retu とする．これにより，k 行目までの天と遠景と中景のみによるラベルづけのうち評価値の和が最大のものが得られる．

i 行目 ($k \leq i \leq n : n$ は画像の上下方向の画素数) では，その画素での地の対象物ラベルの最大の評価値 EVL_chi[j][i] と中景の対象物ラベルの最大の評価値 EVL_chi[j][i] の大きい方を SUM_retu に加算して行く．

このように一列を走査することにより，$1 \leq i < k$ では，天，遠景，中景の対象物ラベルによるラベルづけ，$k \leq i \leq n$ では，地，中景の対象物ラベルによるラベルづけを行ったもののなかで，評価値の和のが最大のものが得られる．

以上の処理を各列 $j(1 \leq j \leq m : m$ は画像の左右方向の画素数) について行う．アルゴリズムは行 k を地面の開始位置と仮定し，それぞれの k に対して処理するように示したが，実際のプログラムでは，SUM_ten, SUM_far, SUM_mid は $k-1$ 行目が地面の開始位置と仮定して計算した値を利用している．また，仮定した地面の開始位置以降で評価値の和を求める部分も簡略化できるため，計算量は画像全体で nm 回程度の加算になる．従って，処理量はそれほど多くない．

第一段階の結果として，列の評価値の和が，各列で n 個，画像全体では $n \times m$ 個得られる．

第二段階は，上下方向の知識を満たし，かつ，左右方向の知識を満たす

```
for(j=1; j<=m; j++){
/* 各列 j について */
  for(k=1; k<=n; k++){
/* k 行目が地面の開始位置と推定 */
    SUM_ten, SUM_far, SUM_mid = 0 に初期化;
/* k 行目より上では，天，遠景，中景でラベルづけ */
    for(i=1; i<k; i++){
      EVL_ten[j][i], EVL_far[j][i], EVL_mid[j][i] を求める;
      SUM_ten=SUM_ten+EVL_ten[j][i];
      SUM_far=max(SUM_ten,SUM_far)+EVL_far[j][i];
      SUM_mid=max(SUM_ten,SUM_far,SUM_mid)+EVL_mid[j][i];
    }
/* k 行目では，評価値が最大のラベルづけを選択 */
    EVL_mid[j][k], EVL_chi[j][k] を求める;
    SUM_retu=max(SUM_ten,SUM_far,SUM_mid)
              +max(EVL_chi[j][i],EVL_mid[j][i]);
/* k 行目より下では，地，中景でラベルづけ */
    for(i=k+1; i<=n; i++){
      EVL_chi[j][i], EVL_mid[j][i] を求める;
      SUM_retu=SUM_retu+max(EVL_chi[j][i],EVL_mid[j][i]);
    }
/* 結果を保存 */
    SUM_retu を列 j で地面の開始位置が k 行目と仮定した時の
        評価値の和の最大値として保存する;
  }
}
```

図 12: 第一段階のアルゴリズム

最適な対象物ラベルづけを求める処理である．左右方向の知識を満たすためには，ある行 i が画像全体で共通な地面の開始位置になっていればよい．第一段階で，各列 j について，i 行目が地面の開始位置であると仮定した場合の最適な対象物ラベルづけの評価値 SUM_retu[j][i] を求めた．これを左右方向に和をとれば，画像全体で i 行目が地面の開始位置であると仮定した場合の評価値の和 SUM_all[i] が求まる．この時，SUM_retu[j][i] は上下方向の知識を満たした対象物ラベルづけにより得られる．従って，評価値の和が SUM_all[i] になる対象物ラベルづけで，i 行目が地面の開始位置であり，上下方向，左右方向の位置関係知識を満たす画像全体に対する対象物ラベルづけが存在する．SUM_all[i] の最大値を与える地面の開始位置 l を求め，これを画像全体に対する地面の開始位置とする (リスト 13 参照)．

第三段階は，各画素の対象物ラベルを決定する処理である．第二段階までで，地面の開始位置が画像に対して一つ定まる．地面の開始位置が定まると，第一段階と同様な処理でその時の最適な対象物ラベルづけが一意に定まる．第一段階で，評価値の和を計算する際に用いた各画素の対象物ラベルを記憶しておけば，この処理は不要である．

4.4 実験および結果

前章で用いた風景画像に対して位置関係知識を適用し，対象物ラベルづけを行った．図 4 の原画像に対する処理結果を図 14 に示す．処理時間は HP9000/720 を用いて 15.32 秒であった．

画像 1 では，空にあった CONCRETE とラベルづけされた部分が SKY に変更された．また，推定された地面の開始位置より下にあった MOUNTAIN の対象物ラベルは，WATER または CONCRETE に変更された．例 2 では，地面の開始位置が空と接する位置に推定されたため，

```
for(i=1; i<=n; i++){
    SUM_all[i]=0 に初期化;
    for(j=1; j<=m; j++){
        SUM_all[i] += SUM_retu[j][i];
    }
}
SUM_all[i] の最大値を与える地面の
    開始位置をもとめる;
```

図 13: 第二段階のアルゴリズム

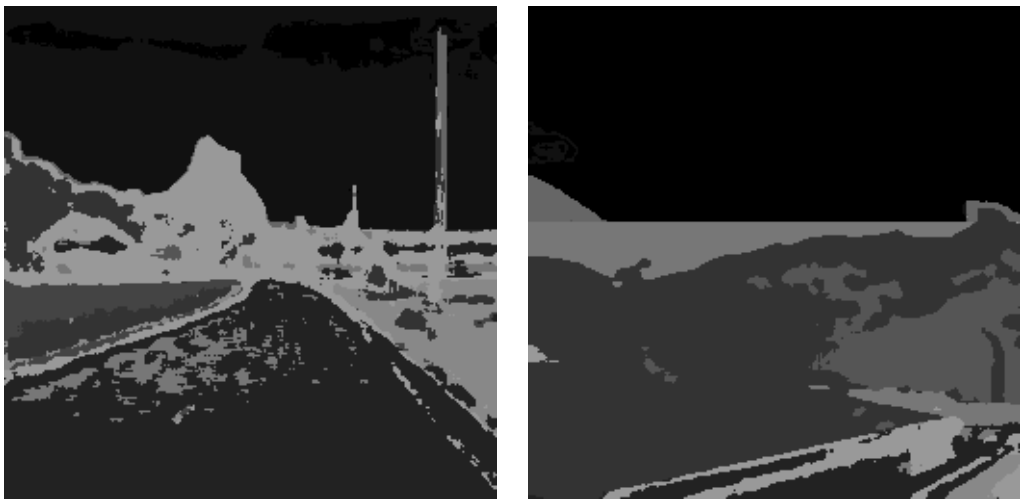
SKY と対象物ラベルづけされた部分の一部が WATER に変更されたが，推定された地面の開始位置より上の部分で，WATER の対象物ラベルが，MOUNTAIN，CLOUD に正しく変更された．他の画像に対しても位置関係知識を適用して対象物ラベルづけを行った．3 章で用いた評価用データセットに対する対象物ラベルづけの正解，誤り数の変化を表 10 に示す．

位置関係知識を導入することにより改善された対象物ラベルのうち最も多かったものは，本来 MOUNTAIN であるべき画素が WATER とラベルづけされていたもので，13 データある．本章で利用した位置関係知識で改善できる対象物ラベルづけは，天, 遠景の対象物ラベルと地, 中景の対象物ラベルの間での誤りが主である．表 8 にあげた誤り方の傾向のうち，この組み合わせの誤りは，MOUNTAIN を WATER と誤る場合が最も多く，他の誤りはそれほど多くない．そのため，評価用データに対する正解率は，それほど大きくは改善されていない．しかし，画像全体としてみると，一画素のみの対象物ラベルが改善されるのではなく，同じ誤り方をした画素の対象物ラベルが全て改善されるため，正解率はより大きく改善されていると考えられる．

本章の処理では，推定された地面の開始位置により，画像の対象物ラベルづけ結果が大きくかわる．地平線・水平線が見える画像では，地面の開始位置が正しく推定できる傾向がある．それ以外の場合では，空に水平線を推定してしまうことがあり，対象物ラベルづけが正しく行えなくなる．推定された地面の開始位置が正しいものであるか検証するなどの方法で，より正確な推定を行う必要がある．

他の位置関係知識に関する考察

本章では，カメラと三次元世界との位置関係に関する前提から得られる位置関係知識を，画像全体に対する対象物ラベルづけの拘束条件として用いている．位置関係知識には，この他にもいくつかのものが考えられる．



画像 1　　画像 2

図 14: 位置関係知識を用いた対象物ラベルづけ

表 10: 位置関係知識の導入による対象物ラベルづけの改善

誤り $\longrightarrow$ 正解	23
誤り $\longrightarrow$ 誤り	10
正解 $\longrightarrow$ 誤り	12
変化なし (正解) (誤り)	925 (705) (220)

一つは，近傍の対象物ラベル間に成り立つ位置関係知識である．これは例えば，「道路の中に一画素のみ水面が存在していることはない」という知識である．一般に対象物はある程度の広がりを持って存在している．従って，一画素のみの対象物ラベルが近傍と異る場合には，ノイズが原因で対象物ラベルづけを誤ったと推定され，周囲の対象物ラベルや，対象物ラベルづけの誤り方の傾向などから対象物ラベルをつけ直す処理を行うことができる．

他の位置関係知識としては，特定の対象物間に成り立つ位置関係の知識があげられる．例えば，「道路があればその近くに道路標識がある」という知識である．この知識を利用して，対象物ラベルづけ処理の順序を決定することができる．上の例では，道路の対象物ラベルづけを行った後に，道路標識の対象物ラベルづけのための処理を起動するという方法がとれる．このようにすると，対象とするすべての対象物を同時に扱うよりも，処理対象が限定でき，また，画像中に存在しない対象物のラベルづけのための処理をある程度省略できるため，処理が効率的になる．しかし，風景画像中の自然物には，道路と道路標識のような密接な関係があるものはないため，このような知識は適用しにくい．

他にも，カメラの位置とは無関係に，対象物の三次元世界での妥当な配置の知識を与え，利用することが考えられる．この場合，カメラの位置に関する前提が変わっても，知識自体は同じものを用いることができるが，各カメラ位置毎に，位置関係知識を画像上で利用できる形に変換する必要がある．本論文で仮定したカメラ位置に対する条件はそれほど厳しくないため，処理対象となる画像はそれほど制限されない．

5　対象物ラベルづけ結果の風景画像検索への利用に関する評価

5.1　従来の画像検索研究

対象物ラベルづけの応用分野の一つに，画像データベースの検索への利用がある．画像データベースでは，画像に検索キーを付加し，検索時には，検索者が与えた検索条件を満たす検索キーをもつ画像を提示する．画像データベースでは，データあたりの情報量が多く，また，検索要求が多様であるため，検索キーの付加，利用が様々な観点から行われる．例えば，X 線写真を対象に画像のスケッチを自動抽出し検索を行う研究 [16]，画像認識により抽出した山の特徴量により検索を行う研究 [17]，人間が画像から受ける曖昧な印象を数値化して検索に利用する研究 [18] などが行われている．しかし，特定の対象物の存在を仮定していない風景画像を対象とした研究はあまりなされていない．

画像データベースを構築する際の問題点の一つは，画像に検索キーを付加する処理である．新たな画像をデータベースに登録するためには，検索キーを付加する必要がある．画像の登録が頻繁な場合，検索キーの付加を自動化したいという要求が生じる．そのため，[16] では，画像処理の閾値を人手により設定することにより，検索キーである原画像のスケッチを半自動的に生成している．[17] では，山の形状，位置，大きさ，色などの特徴量を検索キーとするために，画像中の山を認識処理により自動的に抽出している．また，[18] では，画像から得られる特徴量と人間の受ける印象との相関関係を解析して，検索キーを自動的に付加している．この際問題となるのは，検索キーがどの程度の信頼度で付加されるかである．現在の計算機による画像認識では，認識率が 100 %ではないため，検索キーの付

加を対話的に行うなどの工夫が必要である．本章では，風景画像を対象に，画素単位対象物ラベルづけを用いて，画像への検索キーの付加を自動化することを目指す．

5.2 対象物のスケッチによる風景画像検索

5.2.1 対象物ラベルづけの画像検索への利用

風景画像の検索では，検索キーとして画像の撮影場所や撮影時期を利用することが考えられる．このような検索キーをデータベースに蓄積された画像から自動的に得ることは不可能である．一方，検索したい画像がある時に，画像内容によりアクセスしたい場合がある．例えば，個人のアルバムでは，検索したい写真が存在することは分かっているが，それがどのようなファイル名で格納されているか分からない場合がある．このような時，画像内容による検索を行うことができれば，すべての画像を表示するより，効率的に目的の画像をみつけることができる．画像内容には，画像内の対象物とその位置，位置関係，色や形状，大きさなどが考えられ，画像内容の指定法も，キーワードによる指定，特徴量による指定，アブストラクト画像による指定などが考えられる．このうち，アブストラクト画像による指定では，画像内容のうち，対象物とその位置，形状，大きさが同時に指定できる．特に，データベース内の検索したい画像が既知で，映像として思い浮かべることができる場合，アブストラクト画像による検索は有効であると考えられる．

この場合，本論文で示した画素単位対象物ラベルづけ手法を，画像に対する検索キーの自動的な付加のための手段として利用できる．画像の対象物ラベルづけは，画像中のどの位置にどの対象物が現れているかを示している．このラベルづけ結果をアブストラクト画像とし，検索キーとして利用する．対象物ラベルづけの正解率 (以下単に認識率) は，7 割程度であ

る．この認識率で画像検索のための検索キーの付加手段として有効であるか調べるため，簡単な画像検索システムを構築し，評価を行った．

5.2.2　検索キーの付加および検索方法

検索は，画像内の対象物の指定により行う (図 15 参照)．まず，データベース内の風景画像を，前章までに述べた画素単位対象物ラベルづけ手法により，画素毎にラベルづけする．原画像の大きさは，256 ×256 画素であったが，検索のためにはそれだけの精度は必要ないと考えられるため，縦横 4 画素毎にラベルづけを行い，64×64 画素に圧縮する．このラベルづけされた画像をインデックス画像と呼ぶ．

検索のためには，検索者が検索したい画像 (検索目標画像) を，画像中の対象物とその位置によりスケッチする．これは，塗り絵を行う要領で画素毎の対象物ラベルを指定することにより行う．作成した画像を検索条件画像と呼ぶ．対象物ラベルが指定されなかった画素は，対象物ラベルが不明であるとし，UNKNOWN というラベルを付加する．検索条件画像は 64×64 画素で与える．

データベース側は対象物ラベルづけされたインデックス画像と検索条件画像との間で対応する位置にある画素毎に対象物ラベルを比較し，指定された全画素に対する一致した画素の割合 (一致度) を原画像のスコアとする．この際，UNKNOWN とラベルづけされた画素は計数しない．

画像データベース内の全ての画像のインデックス画像と検索者が入力した検索条件画像との比較を行い，スコアの良いものから順に提示を行う．

5.3　検索実験と評価

検索実験

まず，画素単位対象物ラベルづけ手法により作成したインデックス画像

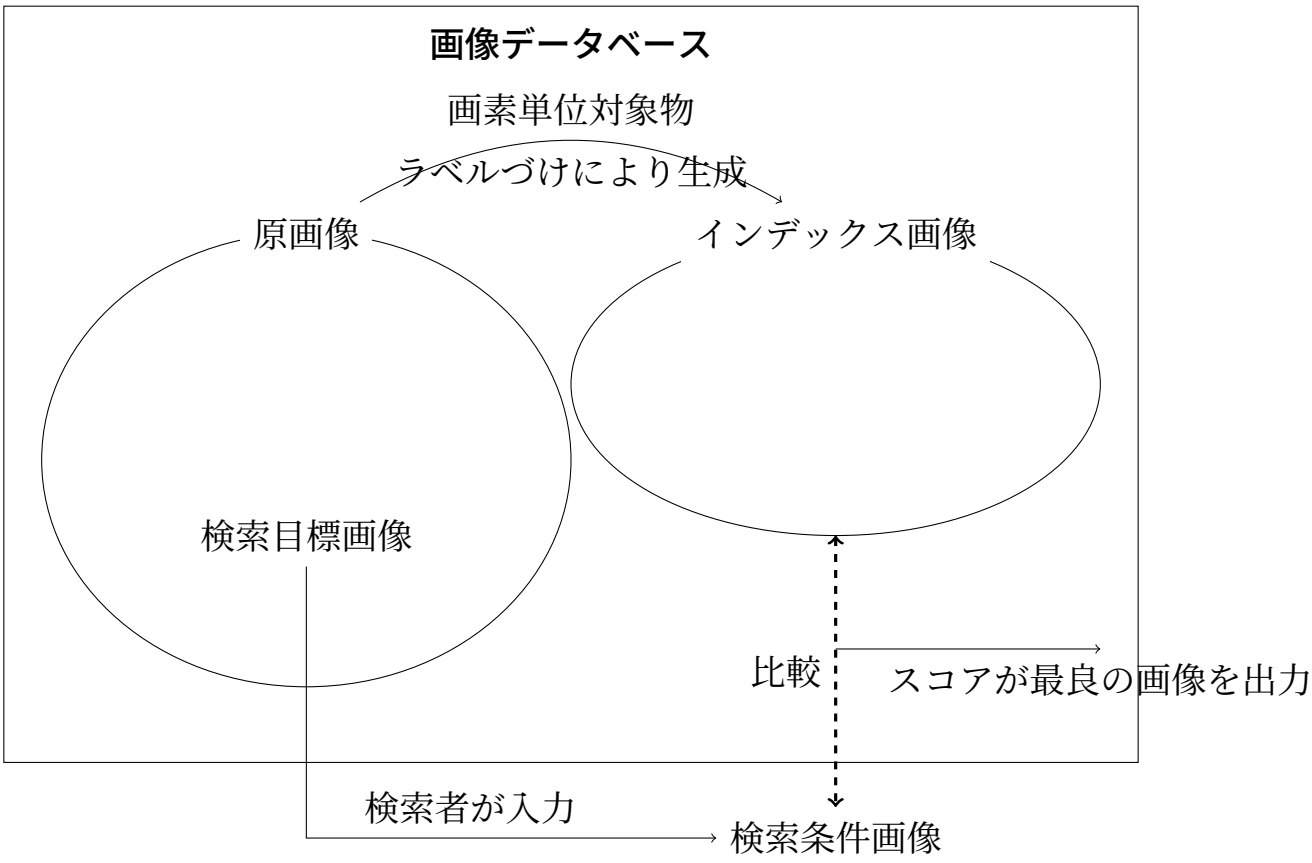

画像データベース
画素単位対象物
ラベルづけにより生成
原画像
インデックス画像
検索目標画像
比較
スコアが最良の画像を出力
検索者が入力
検索条件画像

図 15: 検索方法

が，画像検索に有効に利用可能であるか調べるために，原画像を見ながら検索条件画像を入力し，検索を行った．これを実験 1 とする．ここで作成した検索条件画像を正解検索条件画像とよぶ．正解検索条件画像は，原画像の理想的な対象物ラベルづけである．検索目標画像が提示された順位の分布と各順位での検索目標画像のスコアの平均を表 11 に，同一の原画像に対するインデックス画像と正解検索条件画像との一致度の分布を表 12 に示した．データベース中には，3 章で教師用，評価用データセットを得るために用いた 66 枚の画像に，新たに 64 枚の画像を加えた 130 枚の画像が存在する．これら全ての画像に対し検索を行った．検索条件画像を入力してから，最もスコアのよい画像を選択するまでの処理時間は HP9000/720 を用いて，10.94 秒であった．

次に，曖昧な記憶をもとにした検索への適用を考え，実験を行った．データベースから任意に一枚の画像を選択し，被験者に短時間提示した．その後，提示した画像を見ずに検索を行った．この実験を実験 2 と呼ぶ．実験では，11 人の被験者に対し，のべ 20 枚の画像の検索を行った．検索目標画像が提示された順位を表 13 中の実験 2 に，入力された検索条件画像が原画像をどの程度正確に反映しているかを目視により判定した結果を表 13 中の検索条件画像の正確さに示した．また，各検索目標画像について，インデックス画像と正解検索条件画像との一致度を表 13 中の正解検索条件画像との一致度に示した．

認識率に関する考察

原画像をみながらの検索 (実験 1) では，検索目標画像の約 5 割が 1 位，約 8 割が 10 位以内に提示されており，多くの場合少数の画像の提示で検索目標画像が得られた．しかし，21 位以降に提示される画像が約 14 %あり，最悪で 81 位まで提示する必要があった．画像が早い段階で提示されないのは，インデックス画像の生成をうまく行うことができなかったことが

表 11: 原画像をみながら検索した場合の順位の分布 (実験 1)

順位	画像数	スコアの平均 (%)
1	67	68.6
2	11	62.0
3	4	63.5
4	7	61.2
5	4	53.0
6	5	51.6
7	3	53.0
8	0	—
9	1	49.0
10	1	60.0
11〜20	9	40.8
21〜30	11	33.3
31〜40	1	31.0
41〜50	5	26.3
81	1	18.0

表 12: 一致度の分布

一致度 (%)	画像数
～20	1
～30	5
～40	13
～50	18
～60	24
～70	34
～80	24
～90	8
～100	3

表 13: 曖昧な記憶に頼った検索の成功率 (実験 2)

検索例	提示順位 (実験 2)	正解検索条件画像との一致度	検索条件画像の正確さ (誤り方)
1	1	84	正確
2	1	78	正確
3	1	77	正確
4	4	72	正確
5	1	68	正確
6	1	61	正確
7	1	56	正確
8	1	56	正確
9	6	51	正確
10	18	32	正確
11	38	32	正確
12	24	79	不正確 (SKY と CLOUD)
13	2	78	不正確 (対象物誤り)
14	8	64	不正確 (位置のずれ)
15	8	61	不正確 (対象物誤り)
16	16	56	不正確 (SKY と CLOUD)
17	26	55	不正確 (位置のずれ)
18	7	54	不正確 (位置のずれ)
19	97	49	不正確 (SKY と CLOUD)
20	6	47	不正確 (位置のずれ)

原因である．表 11 のスコアの平均をみると，スコアがおよそ 50 %以上であれば，10 位以内に提示されている．3 章での評価用データセットに対する認識率は 73.9 %であったので，全画像に対して平均的にこの認識率が得られれば，10 位以内での提示が可能である．しかし，表 12 から分かるように，実際には，画像により認識率に差が生じており，一致度が 50 %以上の画像は約 7 割である．これは，画像間で色の偏差が大きいためと考えられる．対象物ラベルづけにニューラルネットワークを用いるので，色補正処理は不要としていたが，色補正処理を行わずに高い認識率を得るためには，多様な品質の画像から学習データを選ぶことが必要である．検索目標画像のスコアの最悪順位をあげるためには，データベース内のすべての画像から学習データをとるなどして，より強靭な対象物モデルを構築し，インデックス画像の生成に失敗しないようにする必要がある．この点については改善の余地があるが，約 70 %の認識率の対象物ラベルづけ結果を用いて，ほとんどの画像を検索することが可能である．

検索方法に関する考察

曖昧な記憶をもとにした検索 (実験 2) では，20 例中 14 例で 10 位以内で提示を行うことができた．画素単位対象物ラベルづけ手法により生成したインデックス画像と正解検索条件画像との一致度が 50 %以上で，検索条件画像がほぼ正確に与えられた場合はすべて 10 位以内に提示されている．しかし，一致度が 30 %台の 2 例では提示できなかった．

検索条件画像の入力時に生じる誤りは，位置のずれと対象物の憶え違いである．位置のずれについては，対象物の境界位置が多少ずれていても対象物の指定が正しければ，ほとんどの場合提示されている．対象物の憶え違いでは，とくに SKY と CLOUD を混同することが多い．これらの対象物は画像上の大きな部分を占めるため，この誤りが生じた場合，提示を行うことができない．同じ空の部分でも，人によって SKY ととる場

合と CLOUD ととる場合がある．対象物を指定法として，「SKY または CLOUD」のような方法を可能とすることにより，この誤りは対処できる．

インデックス画像では，原画像に比べ画素あたり 24bit から 3.5bit に，画像の大きさで 1/16 になっており，全体の情報量を 1,572,864bit から 13,607bit に 1/115 に圧縮している．インデックス画像ファイルの読み込み時間を除いた，画像同士の比較のための処理時間は一組あたり 2.8 ミリ秒 (HP9000/720 使用時) 程度である．本論文で用いた検索方法では，画像データベース内の画像数に比例して検索時間が増加するため，本質的には，大規模な画像データベースの検索に向かないが，あらかじめすべてのインデックス画像を読み込んでおけば，1000 枚程度の規模のデータベースを実用的な時間で検索することができる．

本論文で用いた検索手法の問題点の一つは，画素数の少ない対象物が検索キーとして有効に働いていないことである．検索者がこのような対象物を指定するのは，その対象物を重要視したためと考えられるので，有効に活用する必要がある．また，検索条件画像と類似した画像を提示する手法であるので，検索目標画像が提示されなかった場合，画像データベース内に検索目標画像が存在していないのか，記憶に誤りがあるのか判定することができないという問題点もある．画像に記号的なインデックスを付加し，キーワードによる検索を併用するなどの方法で，これらの問題点に対処する必要がある．

6　おわりに

本論文では，風景画像の対象物ラベルづけにおいて，画素を単位としてラベルづけを行う手法を提案した．この手法では，領域を単位とする場合に必要な初期的な領域分割を行わないため，信号レベルでの領域分割が過分割になるかなどの問題を扱う必要がない．また，画素に付加した対象物ラベルに基づいて領域生成を行うため，対象物と一対一に対応した領域を得ることができる手法である．

画素単位対象物ラベルづけを実現するためには，画素単位で得られる特徴量と対象物ラベルとの対応関係を対象物モデルとして与える必要がある．人間はこの対象物モデルを明確に表現することができないため，本論文では，人間が例示を行い，ニューラルネットワークを利用して対象物モデルを構築する方法を用いた．構築された対象物モデルを用いて，非学習データに対して 73.9 %の認識率を得た．この結果は，領域単位で同様に構築した対象物モデルを用いた場合と同程度以上であり，画素単位対象物ラベルづけが領域単位と同程度以上に有効であることが示された．

画像の対象物ラベルづけには，画素単位で独立に得られる局所的な特徴のみでなく，画像全体としてのラベルづけの整合性が満たされる必要がある．ラベルづけの整合性は，位置関係知識として与えられる．この知識を画素単位対象物ラベルづけで利用する方法を実現した．その方法では，位置関係知識を満たす画像の対象物ラベルづけの中で，画素毎に定めた対象物ラベルの評価値の和が最大になるものを選択する．風景画像での位置関係知識は，画像の上下方向と左右方向に関する拘束条件に大別できるが，画素を単位とすることにより，これらの拘束条件を容易に導入でき，処理も高速に行うことができた．

風景画像の対象物ラベルづけの応用の一つとして，画像データベース検

索への利用がある．画像に対する対象物ラベルづけを検索キーとする場合，認識率が問題になる．検索者が検索したい画像のスケッチを入力し，類似した画像を提示する検索方法で必要となる認識率を評価した結果，10枚以内の候補の提示で検索が成功するためには，画像の約 50 %以上が正確にラベルづけされていれば十分であると考えられる．実験のために用意した 130 枚の画像のうち 7 割はこの条件を満たしており，多くの画像に対して，少数の候補の提示で検索が可能であった．

今後の課題としては，

- 風景画像中の人工物の認識
- 位置関係知識と局所的な特徴量とを同時に利用した対象物ラベルづけ処理の実現
- 色の偏差に強い対象物モデルの構築

などがあげられる．

謝辞

本研究を進めるにあたり，終始御指導，御鞭撻を賜わりました池田教授に心から感謝の意を表します．

熱心な御指導，有益な御示唆を頂きました美濃助教授に感謝の意を表します．

また，日頃から有益な御意見を頂き，画像検索の実験にご協力頂きました広瀬助手，天野氏，並びに池田研究室の皆様に感謝いたします．

参考文献

[1] 松山 隆司, 長尾 真: 航空写真の構造解析, 情報処理学会誌, Vol.21, No.5 (1980-05), 468-480.

[2] 李 仕剛, 辻 三郎, 今井 正和: 移動カメラで連続観測した線による環境の 3 次元構造の決定, 信学論 (D-II), Vol.J74-D-II, No.2 (1991-2), 175-183.

[3] Yu-ichi Ohta: A Region-Oriented Image-Analysis System by Computer, 京都大学大学院工学研究科博士論文 (1980-3).

[4] 平田 真一, 白井 良明, 浅田 稔: 単一カラー画像から得られる 3 次元情報を利用したシーンの解釈, 信学論 (D-II), Vol.J75-D-II, No.11 (1992-11), 1839-1847.

[5] Bruce A.Draper etc: The Schema System, IJCV Vol.2, No.3 (1989), 209-250.

[6] Thomas M.Strat and Martin A.Fishler: Context-Based Vision: Recognizing Objects Using Information from Both 2-D and 3-D Imagery, IEEE Trans. on PAMI, Vol.13, No.10 (1991-10), 1050-1065.

[7] 椋木 雅之, 美濃 導彦, 池田 克夫: 対象物モデル構築に有効な特徴量のニューラルネットワークによる評価, 秋季信学全大 SD-11-5 (1992).

[8] 椋木 雅之, 美濃 導彦, 池田 克夫: 対象物認識処理と領域分割処理の相互作用についての一考察, 画像の認識・理解シンポジウム MIRU'92 講演論文集 I-17 (1992-7).

[9] 鎌田 清一郎, 辻 正文, 河口 英二: 観測時期の異なるランドサット画像の識別, 画像の認識・理解シンポジウム MIRU'92 講演論文集 II-487 (1992-7).

[10] Jon A.Benediktsson, Philip H.Swain, Okan K.Ersoy: Neural Network Approaches Versus Statistical Methods in Classification of Multisource Remote Sensing Data, IEEE Trans. on GE, Vol.28, No.4 (1990-7), 540-551.

[11] 松山 隆司, ビンセント ハング: 画像理解システム SIGMA, 情処論, Vol.26, No.5 (1985-9), 877-889.

[12] 高屋 出, 美濃 導彦, 池田 克夫: ニューラルネットワークを利用した対象物モデルの構築, 秋季信学全大 D-239 (1991).

[13] 栗田 多喜夫: ニューラルネットにおけるモデル選択の試み, 信学技報 PRU89-16 (1989), 17-22.

[14] 渡辺 澄夫, 福水 健次: ニューラルネットワークの統一理論と新しいモデルの提案, 信学技報 NC91-122 (1991), 179-186.

[15] 長尾真：画像認識論, コロナ社 (1983).

[16] 長谷川 純一, 福村 晃夫, 鳥脇 純一郎: 胸部 X 線写真データベースのためのスケッチ画像の作成と利用, 信学論 (D), Vol.J65-D, No.9 (1982-9), 1121-1128.

[17] 美濃 導彦, 岡崎 洋, 坂井 利之: 対象物の属性特徴による画像検索法, 情処論, Vol.32, No.4 (1991-4), 513-522.

[18] 栗田 多喜夫, 加藤 俊一, 福田 郁美, 坂倉 あゆみ: 印象語による絵画データベースの検索, 情処論, Vol.33, No.11 (1992-11), 1373-1383.