

テクスチャと色情報を用いた森林画像の領域分割と種別

山崎 公俊^{*1}

Segmentation and Classification of a Forest Image Using Texture and Color Information

Kimiotoshi Yamazaki^{*1}

^{*1} Shinshu University
4-17-1 Wakasato, Nagano, Nagano, 380-8553, Japan

This paper describes a method of image segmentation and classification intended for forest images. We consider remotely-operated agent who is imposed to seek existences such as disaster victim. Image streams captured from the agent are our input data. The difficulty of image segmentation and classification is that categories defined in forest environments tend to have various size and complicate shape. Pixel-wise classification should be needed for such situation. However, it leads misunderstanding because of low expressive power on each pixel. To overcome this issue, we take two-stage classification algorithm combining grid-based classification and pixel-based classification. Filter bank is used to describe image features. Experiments using real forest images show the effectiveness of the proposed method.

Key Words : Forest image, texture feature, segmentation, classification

1. はじめに

森林などの屋外環境で探索活動をおこなう場合、まず、その環境を理解できる必要がある。それには高度な認識能力が必要となるため、探索活動の多くは人間に頼っているのが現状である。しかしながら、その活動は探索者にとっても危険な作業であるため、移動体を遠隔操作することにより、その危険を避ける方法も検討されている⁽¹⁾。そこでの課題の一つは、遠隔にいる操作者が探索環境を素早く把握することである。それが不十分であると、移動体の操作の質にも影響を与えると考えられる。例えば、探索対象の見落としなどといった望ましくない事態を引き起こしかねない。そこで、移動体側にある程度の環境認識機能を搭載し、遠隔操作者の判断の援けとすることが考えられる。

本研究の目的は、例えば図1に示されるような森林環境で探索をおこなうことを想定し、遠隔移動体によって撮影される画像から、環境構造を把握することである。具体的には、森林画像から地面、木の幹や葉、空、移動体(犬)などの存在と配置を推定する。これらの情報は、遠隔操作者が周囲環境の構造を効率的に把握したり、探索対象のような異質な部分をいち早く検出するための基礎処理となると期待される。

そのような推定をおこなうには、画像の領域分割と、



Fig. 1 An example of forest image captured from a remote-controlled agent

分割された領域の種別が必要である。ただし、森林画像の場合は以下のような難しさに対処する必要がある。

1. 木や地面など、森林の構成物には決まった形がなく、かつ複数の構成物が重なり合ってひとつの景色を構成している。
 2. 遠近の違いにより、構造物の見えが様々に変わる。例えば、一つの葉がはっきり見える場合と、多数の葉がひとかたまりにしか見えない場合がある。
- これらの課題に対応できる領域分割・種別手法を構築することが重要である。

本稿では、テクスチャ情報と色情報を用いた手法を提案する。テクスチャ情報に関しては、フィルタバンクを用いた特徴抽出法を利用する。また、種別の成功率を高めるために、矩形の画像領域をひとつの単位と

^{*1} 信州大学学術研究院工学系 (〒 380-8553 長野県長野市若里 4-17-1) kyamazaki@shinshu-u.ac.jp

したグリッドベースの識別と、個々の画素に対するピクセルベースの識別を組み合わせた方式を提案する。これにより、遠くに映り込んだ細い幹の存在や、地面に落ちた枯葉と地面から生えた草の区別など、細かい境界の推定に対する実現可能性を示す。

本稿の構成は次のようである。2章では、関連研究と本研究の立場について述べる。3章では、本研究で提案する識別処理の構成について述べる。4章では、特徴量表現の算出方法について述べる。5章では、実画像を用いた実験結果について述べ、6章でまとめる。

2. 関連研究

屋外画像のセグメンテーションは従来から多く取り組まれている課題である。自動運転車等への搭載を目的とした方法が様々に提案されている⁽²⁾⁽³⁾。Miscusikら⁽⁴⁾は、入力画像からスーパーピクセルを求めたのち、それらに対して算出されたビジュアルワードの共起性を利用した領域分割の手法を提案した。Zhangら⁽⁵⁾は、KNN-MRF マッチングを提案し、画像の領域分割と意味づけを可能にした。Chengら⁽⁶⁾は、入力画像をスーパーピクセルに分割し、画像エッジや色等の情報に基づいてスーパーピクセルを融合していく方式を提案した。Shottonら⁽⁷⁾は、入力画像へフィルタバンクを適用し、テキストマップを作成したのちに、ピクセルレベルでの分類を行う方法を提案した。

本研究は、フィルタバンクと色情報を利用した特徴量記述をおこなっている点ではShottonらの手法と同様である。一方で、森林環境のテクスチャパターンに着目した特徴量表現や、識別器を2段階にする部分が従来とは異なる。これらにより、森林画像に対して機能する領域分割と種別の手法を構築する。

この他のアプローチとして、文献⁽⁸⁾で提案されたようなCNNを用いた手法も魅力的ではあるが、大量の学習データを要することが一つの課題である。本研究では災害時の探索活動を一つの応用先としているが、そのような環境は多種多様であり、それぞれの環境で大量の学習データを集めることは容易ではない。そこで本研究では、比較的少数の学習データしか与えられない場合であっても、その環境下で機能する手法の構築に重点を置いた。

3. 識別処理の構成

森林環境での探索行動を考えたとき、地面の落ち葉と草の見分けができれば、移動体の経路計画を立てる一つの手がかりとなる。また、木の幹と葉が分離でき、幹の太さが分かれば、遠近情報を知る一つの手がかりとなる。ただし、それらは大小様々かつ複雑な形状の

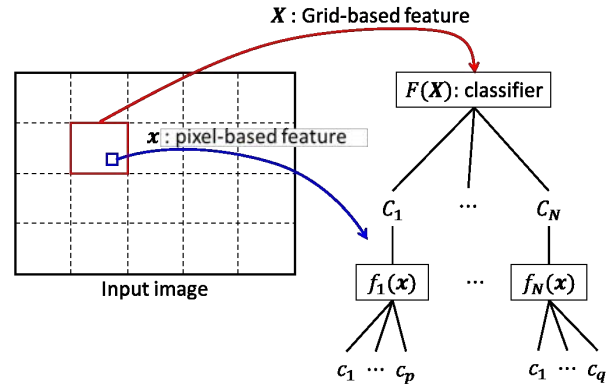


Fig. 2 Classification framework

画像領域を構成していることが多い。よって、セグメンテーションはピクセル単位で詳細におこなうことが望ましい。ただし、そのためには、各ピクセルごとに特徴量を計算し、それをベースとして識別をおこなうといった流れになる。しかしながら森林画像では、各ピクセルで弁別性の高い特徴量を得ることが難しい。そこで、まず画像をグリッドに分割し、各グリッドごとの識別をおこなう。その後、各グリッド内部に含まれるピクセルを一つ一つ識別する。これにより、各カテゴリの存在領域をピクセル単位で得る。

Fig. 2は、本研究で構成した識別処理の流れを示している。ここには2種類の識別器がある。ひとつは、グリッド単位で切り出した画像領域から特徴ベクトルを求め、それを入力として識別をおこなうものであり、 $F(\mathbf{X})$ と表現した。ここで、 \mathbf{X} はグリッド単位の特徴ベクトルを意味する。もう一つは、ピクセルごとに求めた特徴ベクトルを入力とする識別器 $f_i(\mathbf{x})$ ($i=1, \dots, N$)である。グリッド単位の識別器は一つのみ用意されるのに対し、ピクセル単位の識別器は、グリッドベース識別でのカテゴリの数だけ用意される。 $f_i(\mathbf{x})$ が出力するクラスは、グリッドのクラスごとに定められる。例えばグリッドベースで地面と判断されると、ピクセルベースでは草、落ち葉などといった限られたクラスのみが識別の対象となる。 \mathbf{x} はピクセル単位の特徴ベクトルを意味する。

識別器の構築は事前におこなっておく。すなわち、テスト用の画像セットを用意しておき、画像の一部を切り出してクラスごとに分類することで、学習データを作成する。それらを用いて多クラス識別器を構成する。識別処理では、入力画像が得られると、まず画像をグリッドに分割し、各グリッドに対してカテゴリ識別をおこなう。次段の処理としてピクセル単位の識別をおこなう。以上により、最終的にはピクセル単位で

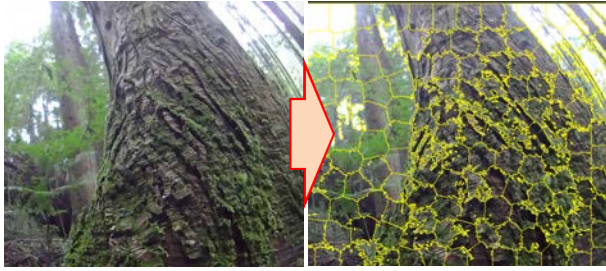


Fig. 3 Superpixels generated on barks of a tree

の識別結果を得る。

ここで重要となるのは、グリッド単位、ピクセル単位のそれぞれでどのような特徴量を用いるかである。すでに述べたように、森林画像では再現性のある形状情報を得ることが難しいため、形状依存の少ない特徴量を適用することが望ましい。そこで、テクスチャ情報と色情報の組み合わせを識別に利用することとし、テクスチャに関してはフィルタバンクを利用して特徴量の算出をおこなう。さらに、表現力の向上のため、フィルタバンクから得られた反応値の列から頻度ヒストグラムを算出し、形状不変性を有する特徴ベクトルを生成する。この詳細は次章に述べる。

なお、グリッド単位で分割するのではなく、領域の境界線を重視したアプローチとして、スーパーピクセルを適用する方式も考えられる。しかしながら、例えば樹皮などは、領域内部の大きな輝度変化に着目することが識別をおこなう一つの方法であり、樹皮らしさをよく記述できる可能性がある。一方で、スーパーピクセルは輝度差の高い部分で境界を生成するため、同一カテゴリ領域を不要に分断する可能性がある。TurboPixels⁽⁹⁾の例を Fig.3 に示す。また、処理速度の観点からも、グリッドのほうが利用しやすい。

4. 特徴量の表現

4.1 特徴量の構成 森林画像は、木の幹、葉、落ち葉、空などのカテゴリに分けられる。いずれも決まった形状がなく、決まった模様もない。それぞれのカテゴリらしさを表現するためには、テクスチャ変化の方向や密度などを手掛かりとするのが一つの手段である。また、色情報を利用することで種別が容易になるカテゴリも存在する。

前章で述べたように、提案手法では、グリッド単位の識別とピクセル単位の識別を組み合わせる。画像特徴量には輝度変化と色を利用することとし、このうち輝度変化については、様々なスケール、様々な方向のエッジ検出フィルタを組み合わせたフィルタバンク

を利用する。そして、一つのフィルタに対する一つの反応値をフィルタの個数分計算し、それを一列に並べて特徴ベクトルとする。ピクセル単位の識別をおこなうとき、各ピクセルがもつ特徴量はこのベクトルとなる。すなわち、ピクセルごとに特徴ベクトルを求めるという表現を用いてはいるものの、畳み込み演算により周囲のピクセルも考慮しているため、一つの画素ではなく、小領域を対象とした特徴量ととらえることができる。

一方で、グリッド単位での識別をおこなうときには、各ピクセルごとに算出された特徴ベクトルを複数まとめて一つの特徴量とする。算出方法の詳細は 4.3 節で述べる。

4.2 ピクセル単位の特徴量算出 4.1 節で述べたように、輝度変化に関する特徴量を得るために、フィルタバンクを利用する。フィルタバンクでは、様々なフィルタカーネルを用意しておき、それらを入力画像に適用したときの反応値を集める。画像テクスチャの分析などの目的に利用されている。以下では、本研究で利用を検討したものを概説する。

山崎と稲葉⁽¹⁰⁾は、布製品の種別を行うためにフィルタバンクを用いた。反応の最大値を得るという意味では下に述べる MR8 と似ているが、フィルタの大きさによって反応値を正規化する点と、異方性の一次微分フィルタのみを使用している点が異なる。本稿では、ガボール関数の標準偏差の最小値を 0.4 とし、そこから 0.2 づつ大きくして、8 種のスケールを設定する。一方で、8 種類の方向を設定する。すなわち、特徴ベクトルの要素数は 64 になる。以下、このフィルタを YI と呼ぶ。

Leung-Malik (LM) フィルタバンク⁽¹¹⁾は、異方性の一次微分フィルタを 18 種類、二次微分フィルタを 18 種類、等方性のフィルタを 12 種類用意し、それぞれのフィルタを画像に適用したときの反応値を一列に並べたものである。1 画素につき 48 要素の特徴ベクトルが得られる。

Maximum Response (MR)⁽¹²⁾を利用するフィルタバンクとして、MR8 がある。これは、異方性の 3 種類の一次微分フィルタ、3 種類の二次微分フィルタ、そして 2 種類の等方性のフィルタを組み合わせたものである。それぞれの異方性フィルタでは 6 種類の方向が定義され、それらを入力画像に適用したとき、最も大きな反応が得られたときの値を代表値とする。1 画素につき 8 要素の特徴ベクトルが得られる。

この他にも、LM フィルタバンクに似た構成の Root Filter Set (RFS)、何種類かの回転不変なフィルタを組

み合わせた Schmid (S) などでも試行する. RFS からは 38 次元, S からは 13 次元の特徴ベクトルが得られる. また, 色情報については, RGB と $L^*a^*b^*$ 表色系を利用する. それぞれ 3 種類の値を持つため, 一つのピクセルからは 3 要素の特徴ベクトルが得られる.

これらの特徴量については, 単独での利用だけでなく, 複数の特徴を組み合わせることも検討する. 具体的には, 複数の特徴ベクトルを連結して一つの特徴ベクトルとする. 例えば, ある画素座標 (u_i, v_j) において異なる特徴量が k 種類あるとすれば, 合計 K 個の要素からなる連結特徴ベクトル $\mathbf{x}^K(u_i, v_j)$ は $\mathbf{x}^K(u_i, v_j) = \{\mathbf{x}_1^T | \dots | \mathbf{x}_k^T\}^T$ と表される. ここで, T は転置を意味する.

4.3 グリッド単位の特徴量算出 グリッド単位での特徴量算出では, 前節で述べた特徴量計算を各ピクセルについておこなった後, 次のようにする. あるピクセルについて, 特徴ベクトル $\mathbf{x}^K(u_i, v_j)$ が得られたとする. ここで, グリッドの左上座標を (u_1, v_1) , 右下座標を (u_n, v_m) とする. すなわち $i = 1, \dots, n, j = 1, \dots, m$ である. 以上の記述に則れば, あるグリッドで得られる特徴ベクトルの集合は, $\mathbf{F}^K = \{\mathbf{x}^K(u_1, v_1), \dots, \mathbf{x}^K(u_n, v_m)\}$ と表現できる.

Fig. 4 に示すように, この \mathbf{F}^K は, $n \times m$ の矩形領域を K 段だけ重ねたものとみなすことができる. そこで, それぞれの段から頻度ヒストグラムを計算する. すなわち, グリッド内部の (u_i, v_j) 座標での特徴ベクトルの k 番目の要素に対して,

$$h_d = N_d \frac{x^k(u_i, v_j) - x_{\min_k}}{x_{\max_k}} \quad (1)$$

を計算し, h_d を整数化したのち, その値が示すヒストグラムの要素位置に 1 を加算する. ここで, N_d はヒストグラムの最大要素数である. x_{\min_k} と x_{\max_k} は特徴ベクトルの要素 $x^k(u_i, v_j)$ が持ちうる最小値と最大値をそれぞれ表しており, これらはあらかじめ与えておく. 一つのヒストグラムごとに $n \times m$ 個の投票がなされるので, その結果を特徴ベクトルとして得る.

一つのグリッドに対する特徴ベクトル \mathbf{X} は, 上述の手順で計算した特徴ベクトルを, K 個並べることで生成する. すなわち, ベクトルの要素数は, $K \times N_d$ となる. この算出方式は, グリッド内部での特徴ベクトルの位置関係には依存しないため, 例えば多数映り込む木々や葉などの存在を, それらの位置によらず統一的に扱うことができる.

5. 実験

5.1 実験準備 移動体 (犬) によって森林環境の探索実験をおこない, そのときに移動体に取り付

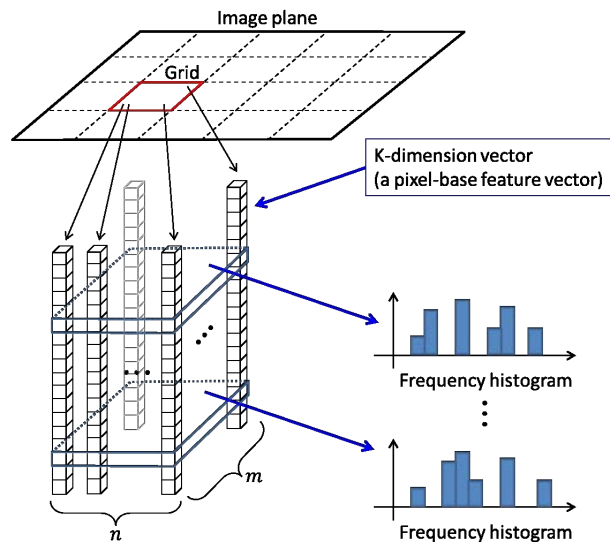


Fig. 4 Making frequency histogram from a set of feature vectors calculated at every pixels

けたカメラで撮影した動画を利用して, 検証実験をおこなった. 特定すべきカテゴリをグリッド単位で 4 種類, ピクセル単位で 8 種類設定した. 具体的には, グリッド単位では, C_1 : 移動体, C_2 : 地面, C_3 : 遠くに見える木々, C_4 : 木の幹, とし, ピクセル単位では, c_1 : 樹皮, c_2 : 犬のボディ, c_3 : 雑草, c_4 : 木の葉, c_5 : 空, c_6 : 落ち葉・枯れ枝, c_7 : 犬の舌, c_8 : 木の幹, とした. そして, クラス間の対応として, $C_1 = \{c_1, c_4, c_7\}$, $C_2 = \{c_3, c_6\}$, $C_3 = \{c_4, c_5, c_8\}$, $C_4 = \{c_2, c_5, c_7\}$ と設定した. 木の幹や木々のカテゴリに空が含まれる理由は, ほぼすべての画像において, 空は木々にさえぎられており, グリッド単位で空が独立に観測できる場合がほとんどなかったためである.

学習データの収集では, 7 枚の画像を学習用に振り分け, そこから学習データを取得した. グリッド単位の識別のために, 各カテゴリごとに 100 から 200 の画像領域を切り出した. それらから識別器を得る手段として, 非線形 SVM を用いる方法と, それとの比較として, Conditional Random Field (CRF) の学習による方法⁽¹³⁾の 2 種類を試した. 一方で, ピクセル単位の識別のために, 各カテゴリごとに 50000 程度のピクセルを選択し, それらにグリッド単位の識別と同様の手続きをおこない, 非線形 SVM により識別器を得た.

以下の実験結果では, グリッドの大きさを 50×50 pixel に設定した. 4.3 節で述べた頻度ヒストグラムについては, ピンの数を 10 に設定した. 非線形 SVM には C-SVC を使い, 多項式カーネル, $C = 1.0$ を経験的に設定した. 4 章で述べたように, 6 種のテキストチャ

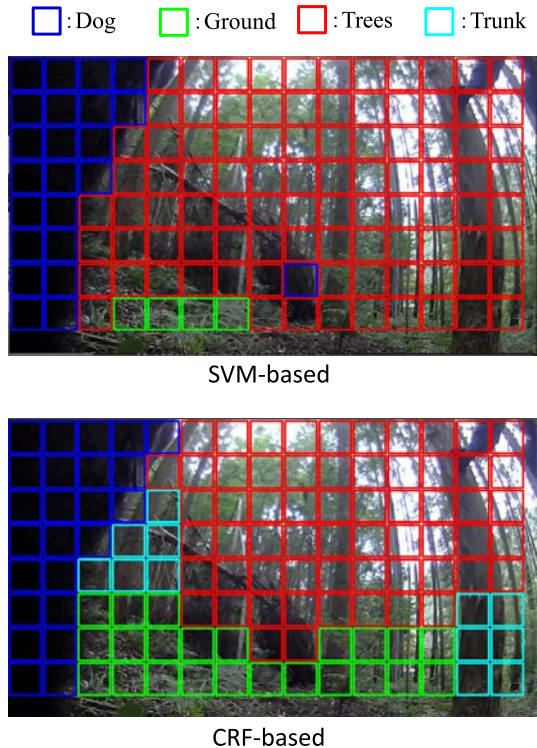


Fig. 5 Results of grid-base segmentation and classification

特徴量と2種の色特徴量を実装し、それら単独もしくは連結の特徴量を用いて、識別の程度を調べた。

5.2 実験結果の例 Fig. 5は、グリッド単位での領域分割の結果を示している。ここでは、YI, LML, $L^*a^*b^*$ の3種を連結させた特徴ベクトルを用いた。SVM利用, CRF利用のそれぞれの場合において、4種の領域がおおむね識別できていることがわかる。ただし、「木々」と「幹」などのように、境界があいまいなカテゴリも存在するため、図中にあるように赤と水色が交互に見られることもあった。

SVMで構成した識別関数を用いて識別をおこなう場合は、各グリッドごとに識別をおこなうので周辺のグリッドは考慮されていない。結果として、例えばFig. 5の倒木部分が1グリッドだけ犬と判断されるといった誤識別が見られた。一方で、CRFを用いる場合は周辺が考慮されるため、そのような誤りは少なくなる。ただし、カテゴリ間の境界部分での誤識別が比較的多く見られた。

Fig.6は、画像の一部を拡大して、ピクセル単位での識別結果の例を示している。図中の円マークはグリッドベースでの識別結果を色によって示したものである。その結果に従って、ピクセルごとの識別をおこなった。上段右側では、緑の部分が木の葉を、黄色の部分が空

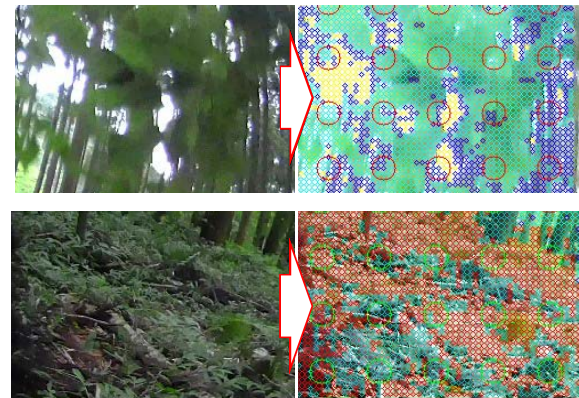


Fig. 6 Results of pixel-base segmentation and classification

を、青の部分が木の幹を示している。葉と空の境界部分で誤識別がみられるが、3種類の識別が概ねできていることがわかる。下段右側は、地面の識別結果である。赤が草を、水色が落ち葉・枯れ木を示している。正解を定義することが困難な例ではあるものの、地面のくぼみに落ち葉や枯れ木が溜まっている部分などがわかるような識別ができており、目的に即した結果が得られていると考える。

5.3 手法の比較と識別性能の評価 グリッドベースの識別手法について、30枚の画像で識別をおこない、識別率の評価をおこなった。Fig.7は、8種の単独特徴量と、性能向上が見られたいくつかの連結特徴量について、その結果を示したものである。白抜きの丸印は識別率の平均値であり、縦方向のバーは標準偏差を示している。このグラフより、単独特徴量としては、YIと $L^*a^*b^*$ 表色系で高い識別率が得られていることがわかる。ただし、色情報とテクスチャ特徴量には識別において得意不得意がある。例えば、木の幹と地面、草と木の葉などでは色が似ているため、色情報を用いた場合は誤識別が起こることがあった。

そこで、色ベースの特徴量とテクスチャベースの特徴量を組み合わせることで、性能改善を試みた。Fig.7からわかるように、性質の異なる2種の特徴量の組み合わせでは画像ごとの識別率にばらつきが見られる。一方で、3種の組み合わせではそのばらつきが少なくなった。これより、提案する特徴表現に対して、機械学習手法による重みづけが適切に機能したと考えられる。試行した中では、YI特徴量, LML特徴量と $L^*a^*b^*$ 表色系の組み合わせが最も効果が高かった。SVMとCRFでの比較においては、前者の場合は識別率が83%、後者は81%となり、前者のほうが若干高い値となった。

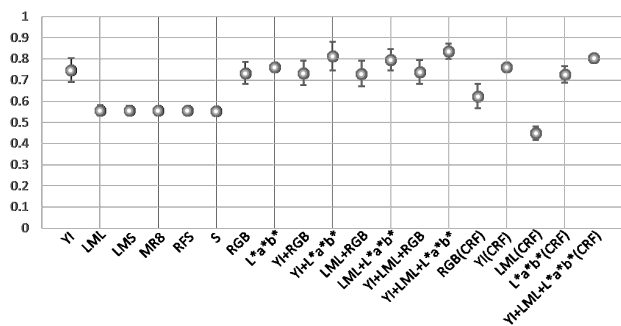


Fig. 7 Classification success rate resulting by the use of various feature combinations

本論文で提案したヒストグラム算出方式のほかに、文献⁽¹⁰⁾で提案された方式についても同様の評価をおこなった。この方式は、領域中心のピクセルに対して同心円状のピンを設定し、回転不変の頻度ヒストグラムを得るものである。無造作に置かれた布製品を対象とする課題では、SIFT 特徴量で Bag of Features を構成する方式⁽¹⁴⁾よりも高い識別性能が確認されていた。しかしながら本研究の課題においては、提案手法よりも識別率が 10% 程度低下した。

6. ま と め

本稿では、森林画像の領域分割と種別について述べた。形状の多様性や遠近による見えの変化など、森林画像での難しさについて述べ、それらに対応するための特徴量記述の方法を提案した。識別率を向上させる方策として、頻度ヒストグラムの計算方法や、グリッドベースの識別とピクセルベースの識別の組み合わせを提案した。実際の森林画像を用いて、適切な特徴量の組み合わせにより、グリッドベースの識別では 83% 程度の成功率で森林画像の領域分割と種別ができることを示した。

今後の課題として、ピクセルベースの識別の評価をおこなうことがある。また、手法を改良して識別率の向上を目指す。さらに、提案手法による処理結果を利用したり、異物発見手法へ展開するなどして、探索行動を効率化する方法についても検討していく。

謝 辞

本研究は内閣府 ImPACT の補助を受けたものである。実験用の画像を提供いただいた東北大学大野和則先生に感謝する。

参 考 文 献

- (1) N. Sakaguchi, K. Ohno, E. Takeuchi, S. Tadokoro: "Precise Velocity Estimation for Dog using its Gait," The 9th Conference on Field and Service Robotics, 2013.
- (2) J. M. Alvarez, A. M. Lopez: "Road detection based on illuminant invariance," IEEE Trans. on ITS, Vol. 12, No. 1, pp. 184 – 193, 2011.
- (3) B. Kulis, K. Saenko, T. Darrell: "What you saw is not what you get: Domain adaptation using asymmetric kernel transforms," In Proc. of Computer Vision and Pattern Recognition, pp. 1785 – 1792, 2011.
- (4) B. Micusik and J. Kosecka: "Semantic segmentation of street scenes by superpixel co-occurrence and 3-D geometry," in Proc. IEEE Workshop VOEC, 2009.
- (5) H. Zhang, J. Xiao and L. Quan: "Supervised label transfer for semantic segmentation of street scenes," in Proc. of the 11th European conference on Computer Vision, pp. 561 – 574, 2010.
- (6) C. Cheng, C. Chen, D. Page and M. Abidi: "Outdoor Scene Image Segmentation Based on Background Recognition and Perceptual Organization," IEEE Trans. on Image Processing, pp. 1007 – 1019, 2012.
- (7) J. Shotton, J. Winn, C. Rother, and A. Criminisi: "Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context," Int. J. Computer Vision, vol. 81, no. 1, pp. 2 – 23, Jan. 2009.
- (8) J. Alvarez, T. Gevers, Y. LeCun and A. Lopez: "Road Scene Segmentation from a Single Image," in Proc. of European Conference on Computer Vision, pp. 376 – 389, 2012.
- (9) A. Levinshstein, A. Stere, K. Kutulakos, D. Fleet, S. Dickinson and K. Siddiqi: "TurboPixels: Fast Superpixels Using Geometric Flows," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31, No. 12, pp.2290 – 2297, 2009.
- (10) 山崎, 稲葉:「布地, しわ, 布の折れ重なりに着目した画像特徴量による無造作に置かれた布製品の個体識別」, 計測自動制御学会論文集, Vol.49, No.7, pp. 661-669, 2013.
- (11) T. Leung and J. Malik: "Representing and recognizing the visual appearance of materials using three-dimensional textons," International Journal of Computer Vision, Vol. 43, No. 1, pp. 29 – 44, 2001.
- (12) M. Varma and A. Zisserman: "Texture classification: Are filter banks necessary?," in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 691 – 698, 2003.
- (13) Krahenbuhl and V. Koltun: "Efficient inference in fully connected crfs with gaussian edge potentials," In NIPS, 2011.
- (14) G. Csurka, C. Bray, C. Dance, and L. Fan: "Visual categorization with bags of keypoints," in Proc. of ECCV Workshop on Statistical Learning in Computer Vision, pp. 59 – 74, 2004.