

遠近変化による見えの違いに対応した 屋外画像からの路面領域の検出

山崎公俊 (信州大学) 入江清 (千葉工業大学) 友納正裕 (千葉工業大学)

1. はじめに

移動ロボットや自動走行車両に持たせるべき環境認識機能として、走行可能領域の検出がある。屋外を走行する移動ロボットのカメラで撮影した画像には、路面、建物、壁、歩行者など様々なものが混在するが、その中から路面を誤りなく検出することは移動ロボットの自律走行にとってきわめて重要である。

本稿では、移動ロボットに載せたカメラで撮影した画像から路面領域を検出するのに適した特徴量を求めることを目的とする。市街地では、多くの場合、路面は舗装された平坦な領域であり、その素材はアスファルトやコンクリート、タイルなどが多い。移動ロボットに水平に載せたカメラから撮影すると、図1のように、カメラの近くではテクスチャが粗く、カメラの遠くではテクスチャが滑らかに見える。このように遠近によってテクスチャが変化するため、従来のような均等さを仮定したテクスチャ解析では路面領域をうまく検出することは難しい。

これに対処するために、本稿では、画像を多重解像度表現し、各解像度で抽出したテクスチャ特徴量を用いて路面検出を行う方法を提案する。この方法の利点は2つある。1つは、遠近のテクスチャの違いの吸収である。低解像度の画像は遠方から撮影したのと同じ効果が得られ、そこで得た近景のテクスチャ特徴量は遠景のテクスチャ特徴量と類似すると期待される。これにより、遠近によるテクスチャの粗さの違いを吸収して同じ領域として検出することが可能になる。もう1つの利点は、遠近の区別である。アスファルトなどの素材を仮定した場合、近景であれば、原画像ではテクスチャの木目が鮮明なので、解像度によって粗さが大きく変化するのである。一方、遠景ならば、どの解像度でも滑らかである可能性が高い。したがって、各解像度のテクスチャ特徴量をもつことは遠近解析の1つの手がかりになりうる。

なお、路面と似た素材が建物の壁や塀などに使用されることがある。本稿ではこれらを見分けることは行わず、テクスチャの類似性に基づいて路面領域の候補を見つける。

2. テクスチャ特徴量

一般的な路面は、アスファルト、コンクリート、タイル、土、砂利などの素材でできている。これらを撮影した画像からは、繰り返しパターンのテクスチャしか得られないことが多く、SIFT [3] のように識別性の高い特徴量を得ることは期待できない。また、色情報は照明条件により大きく変化するため、屋外での領域検出には向かない。そこで、本研究では、テクスチャ

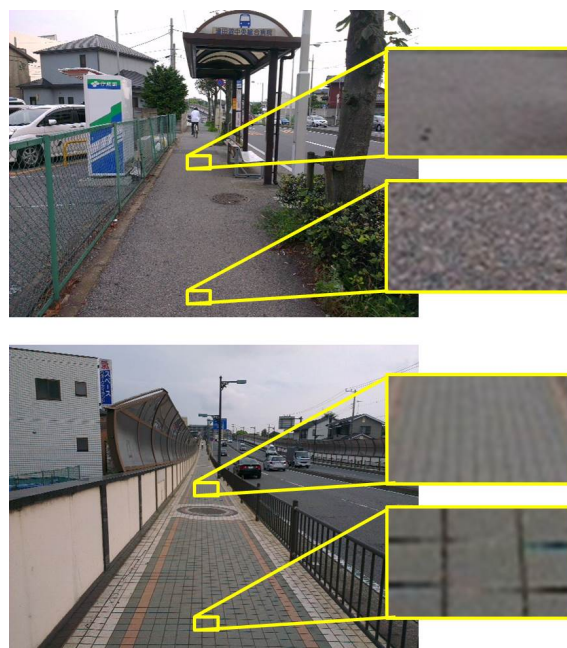


図1: 遠近変化による路面の見えの違い

特徴量を用いて路面検出を行う。テクスチャ特徴量の利点は、局所領域の輝度値変化を手掛かりとするため、色情報に比べて照明条件の影響を受けにくいことである。なお、画像の階調が不足することによる白とび等や、強い日照条件により生じる影への対応については、本研究の範囲外とする。

有力なテクスチャ特徴量としてフィルタバンクが知られている。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ことで、画像テクスチャの分析を行うことが可能である。このフィルタカーネルの集合はフィルタバンクと呼ばれる。以下では、本研究で比較検討に利用したものを概説する。

Leung-Malik (LM) フィルタバンク [2] 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次元の特徴ベクトルが得られる。

Maximum Response (MR) [5] 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次元の特徴ベクトルが得られる。

山崎と稲葉 [7] 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 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、このフィルタの構成を次のように変更して用いる。すなわち、ガボール関数の標準偏差の最小値を 0.2 とし、そこから 0.1 ずつ大きくして、16 のフィルタカーネルを設定する。これにより、より細かなテクスチャにもよく反応するようになる。以下、このフィルタを YI と呼ぶ。

フィルタバンクとは異なる視点の手法として、Local Binary Pattern (LBP) [4] が存在する。LBP は、周囲 8 近傍の画素が注目画素と比べて高い輝度値を持つか否かを調べ、最終的に 8 bit の値を出力する方式である。改良方式が提案されており、テクスチャ解析や、物体認識等への応用も存在する [1][6]。上述した LM や MR8 に比べ、細かなテクスチャに反応しやすい。

3. 遠近変化による見えの違いに対応した画像領域分割

遠近の違いに起因する見えの変化を吸収するために画像の多重解像度表現を用いる。低解像度の画像ではテクスチャの粗さが捨象されるため、路面全体を一つの領域として得ることが可能になる。一方、高解像度では近景のテクスチャの粗さが保持される。これらを合わせることで、同一領域内でテクスチャの粗さの変化がわかり、遠近を知る手がかりとして利用できる。

3.1 多重解像度によるテクスチャ表現

図 1 で示したように、ロボットに搭載したカメラから撮影した画像では、カメラに近い部分では路面のテクスチャが粗く、遠い部分では滑らかに見える傾向がある。ここでは、画像の多重解像度表現を用いて、カメラからの遠近情報を内包したテクスチャ特徴量を定義する。

画像の多重解像度表現は 3 層のピラミッド画像を用いる。原画像を第 1 層の画像として、その各辺の長さを 1/2 に縮小して第 2 層の画像を生成し、その画像に対して同様の処理を行い第 3 層の画像を生成する。その際、元の画像の 4 個の隣接画素を平均して、縮小画像の 1 個の画素にする。これにより、縮小画像においては、テクスチャが平滑化されて滑らかになる。これは遠方領域を撮影したのと類似した結果になる。

このピラミッド画像に対して、テクスチャ特徴量の分布を調べる。まず、あらかじめ、多くの画像から抽出したテクスチャ特徴量に k-means 法を適用して、 K 個の代表ベクトルを求めておく。実装では $K = 20$ とした。これらの代表ベクトルのノルムを求め、全代表ベクトルの中での最大ノルムが 1 となるように全ての代表ベクトルを正規化する。

次に、ピラミッドの各層の画像のテクスチャ特徴量を束ねて、多重解像度のテクスチャ特徴量を生成する。まず、第 n 層の画像の各画素からテクスチャ特徴量を

求め、その特徴ベクトルが K 個の代表ベクトルのどれに最も近いかわかる。そして、その代表ベクトルのノルムをもとに次式により第 n 層での値 v_n を計算する。そして、各層の v_n ($n = 1, \dots, 3$) を並べたベクトルをその画素の多重解像度テクスチャ特徴量とする。

$$v_n = \frac{c}{|f^k| + c} \quad (1)$$

ここで、 $|f^k|$ は、現在の特徴ベクトルが一番近いと判断された代表ベクトルのノルムである。 c は定数であり、実装では 0.3 とした。

今回のようにピラミッド画像が 3 層の場合は、この多重解像度テクスチャ特徴量 v_n を画素の RGB 値に対応づけると、可視化できて便利である。具体的には、第 1 層の画像 (原画像) で得た値 v_1 を B 値、第 2 層の画像 (1/2 画像) で得た v_2 を G 値、第 3 層の画像 (1/4 画像) で得た v_3 を R 値に対応づける。こうして生成した画像をテクスチャ情報画像と呼ぶ。

このようにすると、ピラミッドのすべての層で似たテクスチャ特徴が得られた画素の色は、テクスチャ情報画像では RGB が同じ値となるため、グレースケールになる。縮小画像ではテクスチャは滑らかであるから、これは原画像でもテクスチャが滑らかであることを意味する。遠方領域の空 (晴天時) などはこれに当たる。また、カメラからの離れた位置にある路面もテクスチャが滑らかになるため、同様の振る舞いとなる。一方、ピラミッド化に伴いテクスチャが変化した場合は、RGB の値が同じにはならない。このときの色の傾向はテクスチャ特徴量の種類によって異なるが、例えば YI では、原画像で粗いテクスチャ特徴、縮小画像で滑らかなテクスチャ特徴が得られた画素は、R が大きな値となり、赤みがかかった色になる。これはカメラに近い領域であることを意味する。

すなわち、テクスチャ情報画像は、カメラに近い部分か遠い部分かを識別する手がかりを画素単位で保持していることになる。

3.2 領域拡張法による領域分割

路面を 1 つの領域として抽出したいときに問題となるのは、カメラに近い部分と遠い部分で支配的なテクスチャ特徴量が異なることである。すなわち、カメラに近い部分では粗い特徴が多いが、遠い部分では滑らかな特徴が多い。一方、前述のように、低解像度の画像では粗いテクスチャも滑らかになる。そこで、ピラミッドの低解像度の画像を用いて領域抽出を行えば、路面が 1 つの領域として得られると期待される。ただし、遠近による見えの違いを完全に捨象することは難しいため、領域分割処理には、多少のテクスチャの違いを吸収できる能力が必要になる。

この領域分割には領域拡張法を用いる。領域拡張法は、ある基点を選んだ後、その基点の近傍で基点と性質が近いものを同族として記録していき、領域を拡張していく手法である。最終的に得られる領域に含まれる特徴量が均一となる保証はないが、テクスチャの性質が徐々に変改していく路面領域画像には有効である。また、各画素を一度しか走査しないため、処理が高速である。

テクスチャ特徴量の算出により、画像の各画素に対応する特徴ベクトル $f(i, j)$ が得られたとする。隣接する画素との類似度を下記の式によって算出することで、領域拡張を行うか否かを決定する。

$$\sum_{d=1}^D |f(i, j, d) - f(i + n, j + m, d)| < f_{threshold} \quad (2)$$

ここで、 (i, j) は画像上の現在の座標値を示しており、 $(i + n, j + m)$ はその近傍の座標値を意味する。 d は、各画素に対応する D 次元特徴ベクトルの要素番号である。

低解像度画像では、抽出した領域内の遠近の違いは捨象されている。そこで、この領域をテクスチャ情報画像に投影する。これにより、同一領域内でありながら、カメラから近い部分と遠い部分を区別して表すことができる。

4. 実験

屋外環境で路面を含むカラー画像を 130 枚取得した。画像サイズは VGA (640×480 pixel) である。これらの画像に対して、多重解像度表現による遠近情報の抽出と領域分割についての定性評価を行った。

4.1 遠近変化への対処

3.1 節の方法を用いてテクスチャ情報画像を生成した。図 2 の青枠内に、4 種類の特徴量に対して得られたテクスチャ情報画像を示す。画像 (1) や画像 (3) のようなアスファルト路面において、YI では、路面とカメラとの距離が近づくほど、色味が増している。これは、異なる解像度の画像で異なる特徴ベクトルが得られたことを意味する。LBP についても、他の特徴量と振る舞いは違うものの、例えば画像 (2) の結果などでは遠近に応じた色味の変化を見ることができる。このような漸次的な変化を利用すれば、路面の遠近を推定できる可能性がある。また、路面のように遠近両方を含む領域と、青空やカメラに正対した壁などテクスチャが異なる領域とを区別できる可能性もある。

また、画像 (1) と画像 (4) などと比較すると、アスファルトやタイルなどの素材の違いがテクスチャ情報画像に色として表れており、それらの識別に利用できる可能性がある。特に YI は、細かなテクスチャに反応するように設計されているため、この可能性をよく示している。

4.2 領域拡張法の適用結果

図 2 の緑枠内は、それぞれの特徴量計算結果を入力として領域拡張法を適用した結果を示している。式 (2) の閾値は、テクスチャ特徴量ごとに実験的に決めた。各画像において一つの色が一つの領域を示している。画像 (1) から画像 (4) においては YI の結果が比較的良好であり、遠近による見えの違いによらず路面領域全体を得ることができている。

ただし、画像 (3) のようにアスファルトの質が途中で変わる場合や、画像 (5) のように影があった場合には、どの特徴量も影響を受け、領域が分断される結果となった。これらへの対応は今後の課題である。

5. まとめ

本稿では、一枚の屋外画像から路面領域を検出する手法について述べた。この手法では、まず画像を多重解像度で表現し、それぞれの解像度画像でテクスチャ特徴量を算出する。そして、低解像度画像で領域分割を行うことで、遠近の違いを吸収して路面領域の候補を検出する。一方で、異なる解像度画像の間でのテクスチャ特徴量の変化に着目し、遠近情報を画素単位で保持させたテクスチャ情報画像を生成する。

実画像を用いた実験で、提案手法が有効であることを定性的に確認した。今後は、路面領域分割の定量評価を行い、性能向上を図る。また、路面の遠近変化の推定、路面素材の分類などについても研究を進めていく。

参考文献

- [1] T. Ahonen, A. Hadid, and M. Pietikäinen, “Face Description with Local Binary Patterns: Application to Face Recognition,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, No. 28, Vol. 12, pp. 2037 - 2041, 2006.
- [2] 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.
- [3] D. G. Lowe: “Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision*, 60, 2, pp. 91-110, 2004.
- [4] T. Ojala, M. Pietikäinen, M. and D. Harwood, “A Comparative Study of Texture Measures with Classification Based on Feature Distributions,” *Pattern Recognition* No. 19, Vol. 3, pp. 51 - 59, 1996.
- [5] 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.
- [6] X. Wang, T. X. Han and S. Yan: “An HOG-LBP Human Detector with Partial Occlusion Handling,” in *Proc. of IEEE 12th International Conference on Computer Vision*, pp.32 - 39, 2009.
- [7] K. Yamazaki and M. Inaba: “Clothing Classification Using Image Features Derived from Clothing Fabrics, Wrinkles and Cloth Overlaps,” in *Proc. of IEEE/RSJ International Conference on Robots and Systems*, pp. 2710 - 2717, 2013.

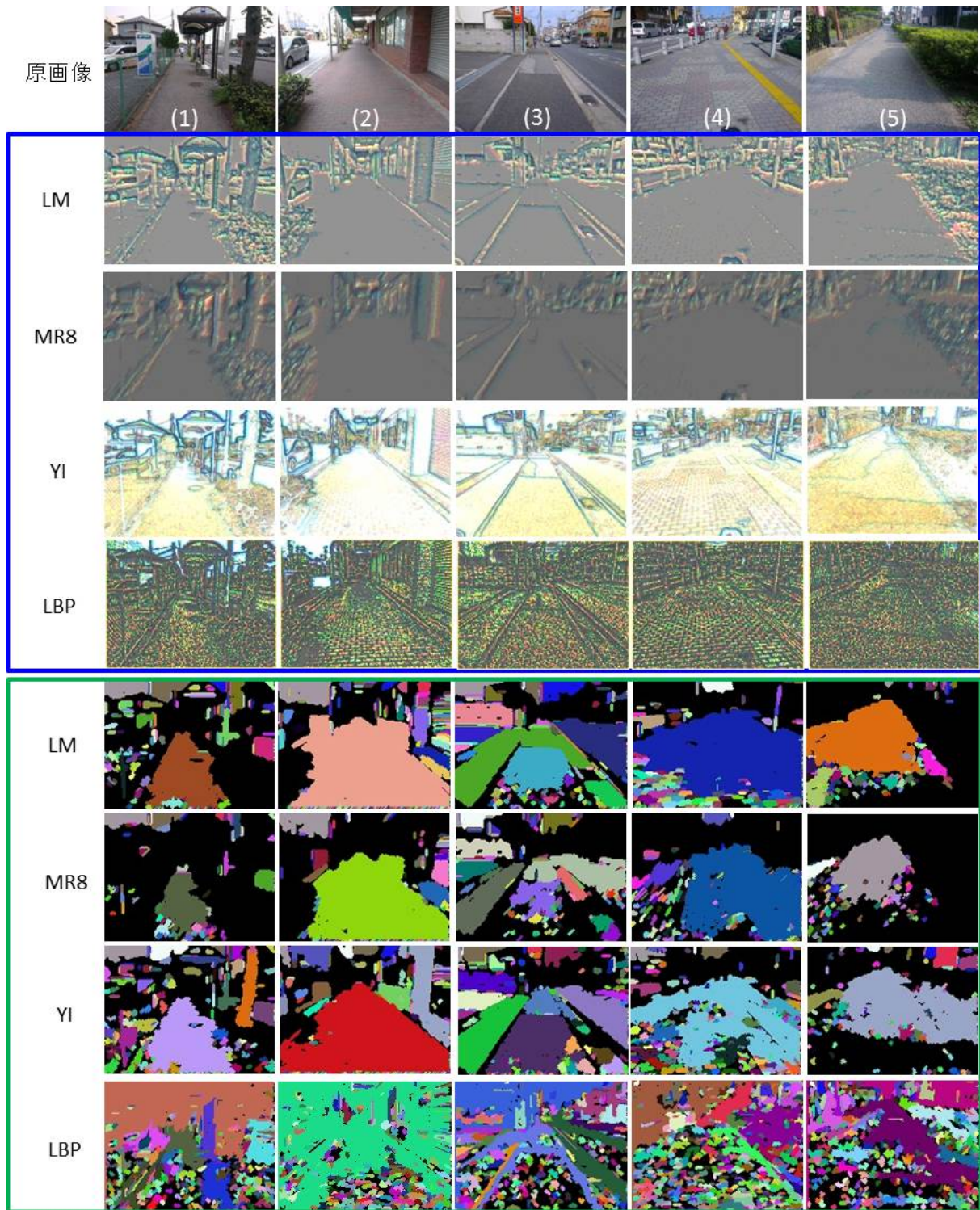


図 2 : 4 種の特徴量を用いた処理結果の例 . 青枠内は特徴量の違いを色づけした結果であり , 緑枠内は領域拡張法を適用した結果である .