マルチテンプレート生成による実環境下のランドマークシンボル検出

中川　祐†，高橋友和†，目加田慶人‡，井手一郎†，村瀬　洋†
†:名古屋大学大学院情報科学研究科
‡:中京大学生命システム工学部

画像中のランドマーク情報を利用したナビゲーションシステムでは，様々な環境下で撮影された画像からランドマークを表すシンボルを検出する必要がある．撮影条件に頑健な特徴量として SIFT 特徴量が知られており，SIFT 特徴量を用いたシンボル検出手法が提案されている．しかし，極端な見えの変化には対応しきれず，検出精度を低下させるという問題があった．そこで本研究では，この問題に対処するため，対象の見えの変化をモデル化し，複数のテンプレートを生成して用いることによって検出精度の向上を図る．実際に撮影した画像を用いた実験の結果から，テンプレート生成を行わない従来手法では適合率 100%としたときに再現率が最大 54.7%だったのに対し，複数のテンプレートを生成して用いる提案手法では適合率 100%としたときに再現率が最大 71.1%となり，提案手法の有効性が示された．

1. はじめに

近年，情報処理技術の発達により，カーナビゲーション，携帯電話などを用いた歩行者ナビゲーションといったナビゲーション技術への関心が高まっている．そのような技術のひとつとして，ユーザ周辺のランドマーク情報に基づき案内・情報提供をする研究がある．多賀らは視認できるランドマーク及びその方向をユーザが選択することで，ユーザの現在位置を特定するシステムを提案している[1]．山口らはモバイルカメラで撮影された店舗の看板画像について，認識のために有効な特徴量や学習方法を検討している[2][3].

ここで，図 1 のように，携帯電話のカメラで撮影されたランドマークを含む画像を受け取り，画像からランドマーク情報を抽出し解析することで，ユーザに現在の位置情報や周辺案内などを提供するシステムを考える．この場合，システムはランドマークを表すシンボルを画像中から検出する必要がある．しかし，カメラと対象との位置関係による見え方の変化，オクルージョン，輝度変化といった問題から，自由に撮影された画像の中からシンボルを検出するのは一般に困難である．

このような対象の見えの問題に対して頑健な特徴量として， SIFT 特徴量が知られている．SIFT 特徴量を用いた様々な検出･認識技術が報告されている．Ichimura はモータースポーツなどのイベントの映像中に現れる，複数の広告看板を認識する手法を提案している[4]．高木らは車載カメラ映像から道路標識を認識する手法を提案している[5]．どちらの手法もテンプレートと入力画像の SIFT 特徴点を対応付け，対象の認識を行っている．しかし，これらの手法では多少の見えの変化には頑健であるものの，極端に対象の見えが変化した場合に，特徴量が変化するために認識性能が低下するという問題がある．

本研究では，この問題に対処するため，見えの変化をモデル化してテンプレートを複数生成する手法を提案する．これにより，ランドマーク情報に基づく情報提供システムの要素技術として，高精度なランドマークシンボル検出の実現を目指す．

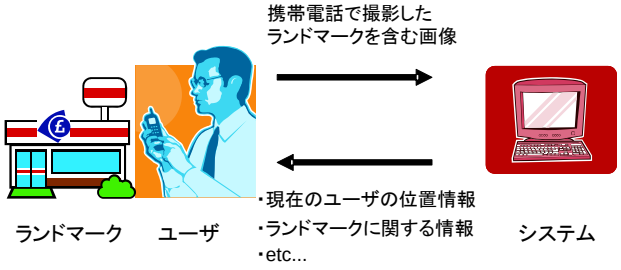


図 1. ランドマークに基づく情報提供システム

2. 提案手法

2.1 手法概要

提案手法の流れを図 2 に示す．提案手法は大きく 2 つの処理にわけられる．1 つ目はテンプレートの生成処理である．与えられた単一のテンプレートに見えの変化モデルを用いた変換を施し，複数のテンプレートを生成する．2 つ目はシンボルの検出処理である．生成された複数のテンプレートと 1 枚の入力画像からそれぞれ SIFT 特徴量を計算し，各テンプレートとシーン画像とで算出された特徴点を対応づける．対応づいた点から射影変換行列を推定し，複数のテンプレートのうち 1 つ以上のテンプレートと入力画像の間である条件を満たす射影変換行列が計算できれば，入力画像中に対象シンボルが存在するとして判断する．各処理の詳細は以下で述べる．

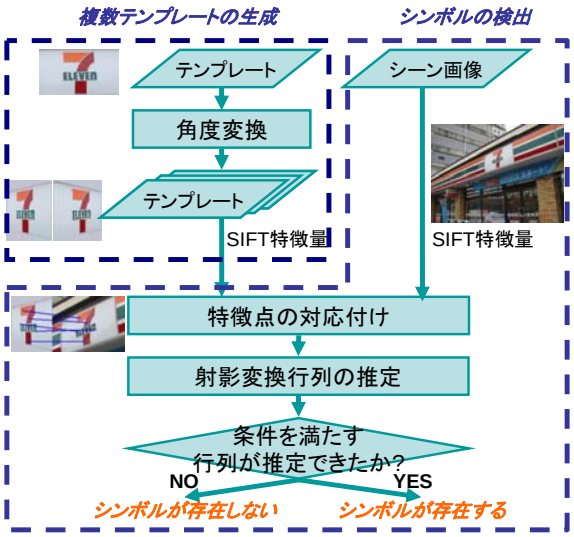


図 2. 処理の流れ

2.2 複数テンプレートの生成

本研究では，店舗看板のように単一の平面上に存在するランドマークシンボルを検出対象とし，テンプレートの生成モデルとして，対象とカメラの位置関係の違いによって生じる見えの変化を扱う．予め正面からランドマークを撮影した画像を，仮想的にシンボルを表す平面として考える．

ある 3 次元座標系 xyz を考える．z 軸上に中心があり xy 平面に並行となるようにランドマークシンボルを表す平面が存在し，座標系の原点には視点 A があるとすると，視軸は z 軸となる．このとき，視軸に垂直に設置された投影面に，シンボルを視点 A に向けて投影したものが基準となるテンプレート画像とする（図 3（a））．ここで，対象とカメラの位置関係の違いによって生じる見えの変化は視点を移動させて再投影することによって表現できる．例えば図 3（b）のように視点を視点 A から視点 B に移動させ，視点とシンボルの中心を結ぶ視軸に垂直な投影面に，シンボルを視点 B に向けて再投影することで，新たなテンプレートが生成される．これはシンボルに対してカメラが右に並行移動した場合の見えの変化のモデルである．

このように，シンボルを表す平面に対して視点を動かして再投影することにより，様々な状況での見えに対応したテンプレートを生成する．

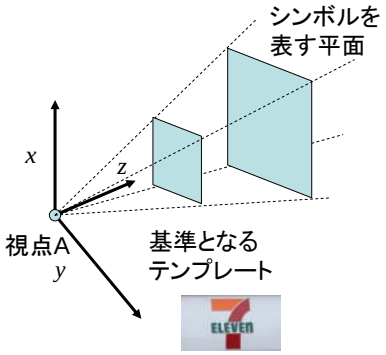
2.3 シンボルの検出

2.3.1 SIFT 特徴量

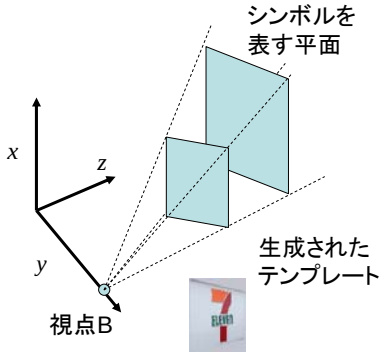
SIFT（Scale Invariant Feature Transform）特徴量は Lowe によって提案された輝度勾配に基づく局所不変特徴量である[6]．SIFT 特徴量は特徴点の位置 $\mathbf{p}$，スケール σ 及び記述子 $\mathbf{d}$ によって構成される．以下に一般的な SIFT 特徴量の計算法について述べる．

エッジ抽出オペレータとして，ガウシアンの 2 階微分である LoG（Laplacian of Gaussian）を近似した DoG（Difference of Gaussian）を用いる．DoG は $G(\cdot)$ をガウス関数，x, y を注目している画素の座標値とすると次式で表される．

$$DoG(x, y, k^n\sigma) = G(x, y, k^{n+1}\sigma) - G(x, y, k^n\sigma) \quad (1)$$



（a）3 次元座標系



（b）視点移動による見えの変化のモデル化

図 3. テンプレート生成

ここで，kは定数パラメータである．また，スケールスペースとしてそれぞれ$n=1,2,3,...,N-1$で計算されたガウシアン画像の階層構造を考える．このスケールスペースにおいて，隣り合うガウシアン画像の差分を取ることで，DoG 画像の階層構造が得られる. DoG 画像の階層構造を 3 次元的に見たときに，ある画素と隣接する 26 近傍の画素とをそれぞれ比較し，その画素が極大となるようなとき，その画素を特徴点の候補とする．この候補点について，DoG 値，主曲率，およびサブピクセルの DoG 値がある範囲にあるものを特徴点とする．特徴点にはスケールσの情報を持たせておく．

得られた特徴点について，その近傍領域で輝度勾配の方向ヒストグラムを計算し，最も頻度が高い方向を探す．そして，特徴点を中心として，画像座標系をその方向に回転した局所座標系を作る．この局所座標系で，特徴点を中心とし，スケールσに比例した大きさの局所領域を考え, その領域を 4×4 のブロックに分割する．各ブロック内で輝度勾配の方向ヒストグラムを計算し，次元がヒストグラムのビンの総数と等しく，ヒストグラムの頻度を値として持つようなベクトルを考える．そして，ブロックごとに計算したベクトルを並べた新たなベクトルを作り，このベクトルのノルムが 1 になるように正規化したものを SIFT 特徴量の記述子$\mathbf{d}$とする．

2.3.2 特徴点の対応付け

まず各テンプレート及び入力画像から SIFT 特徴量を計算し，テンプレートと入力画像の特徴点を対応付ける．t番目のテンプレート画像中のi番目の特徴点$\mathbf{p}_i^t$に対する SIFT 記述子を$\mathbf{d}_i^t$で表現し，同様に入力画像I中のj番目の特徴点，SIFT 記述子を$\mathbf{p}_j^I$，$\mathbf{d}_j^I$で表す．このとき，特徴量間の距離を記述子間のユークリッド距離$d_{ij}=\left\|\mathbf{d}_i-\mathbf{d}_j\right\|$で定義する．距離尺度$d_{ij}$を用いるとき, テンプレート中の各特徴点$\mathbf{p}_i^t$について，入力画像中の最近傍点のインデックスを$NN=\arg\min_j d_{ij}$で表現する．同様に 2 番目に近い入力画像中の特徴点のインデックスを$2\text{-}NN$とすると，次式を満たすようなシーン画像中の点$\mathbf{p}_{NN}^I$をテンプレート上の点$\mathbf{p}_i^t$と対応づける．

$$d_{iNN}\Big/d_{i2\text{-}NN} < match\_threshold \qquad (2)$$

これをテンプレート上の全ての特徴点で計算する．

2.3.3 射影変換行列の推定

対応付けられた特徴点から，入力画像とテンプレートの間に妥当な射影変換行列を求められれば，入力画像中にテンプレートと同じシンボルが存在することがわかり，特徴点の対応からその位置が特定できる．しかし，前節の対応付け結果には，正対応（inlier）だけでなく誤対応（outlier）も含まれる場合があるために，単純に全ての対応を用いただけでは，必ずしも正しい射影変換行列は計算されない．Outlier にロバストな統計量の推定手法として，RANSAC（RANdom SAmple Consensus）[7]が知られている．RANSAC はランダムに抽出されたサンプルから得られた推定値が正しいと仮定した場合の inlier の数を数えるという処理を繰り返し行い，inlier の数が最大になるような推定を最も正しい推定と見なす手法である．本研究ではこの RANSAC に基づいて射影変換行列を推定する．そのアルゴリズムを以下に記す.

(1) 前節の処理で対応づいた点の組からランダムに 4 組を選び，2 次元から 2 次元への射影変換行列$\mathbf{H}$を計算する.
(2) 計算された$\mathbf{H}$を用い，テンプレートT上の特徴点$\mathbf{p}_T$に対応する入力画像I上の特徴点$\mathbf{p}_I$をテンプレート上に変換したときの変換誤差$e=\left\|\mathbf{p}_T-\mathbf{H}^{-1}\mathbf{p}_I\right\|$を求める.
(3) 変換誤差が閾値以下となるテンプレート上の特徴点を inlier とする．テンプレート上の全ての特徴点で(2)の処理を行い，inlier の総数を求める.
(4) (1)~(3)の処理を複数回繰り返し， 最大数の inlier を与えるときの inlier と判定された点の組のみを使って再び射影変換行列$\mathbf{H}'$を計算し，最終的な推定結果とする.

RANSAC の試行回数は次のように決定される．選ばれたサンプル中に inlier しか含まれない確率をP_{In}, 現在の試行回数をkとしたとき，次式が満たされなくなったら試行を打ち切る[8].

$$(1-P_{In})^k > RANSAC\_threshold \qquad (3)$$

ここで，P_{In}は全対応のうち inlier の割合で近似され，あらかじめ与えられる．つまり，選ばれたサンプルに outlier が含まれる場合がk回連続で発生する確率が, 一定値を下回った場合に計算を打ち切る．今までの試行全てに outlier が含まれるのは最悪の

場合であり，この式は最悪の場合が続く回数の上限を決定するものである．

現実に起こり得る見えの変化には限りがあるが，射影変換は自由度の高い変換であるため，現実には起こり得ない射影変換行列が計算されてしまう可能性がある．それを避けるために，計算された射影変換行列が妥当なものであるかを確認する．具体的には，ねじれを持たない，反転しない，ロール角（画像平面上の回転角）が一定範囲内に収まる，変換後の図形が一定以上の面積を持つ，という 4 つの制約を加える．

ねじれは画像頂点の順番が保存されないために発生する．これは射影変換後の画像の辺ベクトルの外積の正負を調べることで検出できる．また，反転は画像頂点の順番が反転した際に発生する．これは射影変換行列の行列式の正負を調べることで検出できる．

3. 実験

3.1 実験目的及び実験条件

提案手法の効果を確認するため，実環境で撮影した画像からランドマークシンボルを検出する実験を行った．我々が街を移動する際に，ランドマークとしては高層ビル，チェーン店舗，駅などをよく利用する．その中で，本実験では特にチェーン店舗及び地下鉄の駅を検出対象ランドマークとして想定し，それらを表す看板・シンボル・マークを検出対象とした．チェーン店舗のような 1 階建ての建築物を撮影する場合に生ずる見えの変化は，一般に水平方向の角度変化が多いと考えられるため，本実験では見えの変化のモデルとして水平方向の角度変化のみを考慮し，テンプレート生成による効果を確認した．実験には次のような条件で撮影されたものを用いた．

- 画像枚数：726 枚
- 対象シンボルの種類：9 種
- 撮影時間帯：午前・午後・夕方
- 撮影角度：ランドマークごとに 7 方向
- カメラ：Canon Powershot S2IS
- 画像サイズ：1296×972 pixels

また，実験の対象となる 9 種のランドマークの詳細及びそれぞれのテンプレートの画像例を表 1 に示す．ここで，テンプレート生成の元となる画像は，実験に用いるデータセットとは別に正面向きに撮影した．この画像からあらかじめ人手によりシンボル部分を切り出したものをテンプレートとして利用する．テンプレートの生成モデルは水平方向の角度変化であり，視点を 15°ごとに回転させ再投影を行う．正面向きのテンプレートだけを用い，テンプレート生成処理を行わずに検出処理を行う手法を従来手法とし，提案手法と精度を比較した．

テンプレート生成の効果を確認するため，水平方向の角度 θ について，正面から撮影した画像を 0°としたとき θ=0°，θ={0°,±30°}，θ={0°,±30°,±45°}，θ={0°,±30°,±45°,±60°}，θ={0°,±30°,±45°,±60°,±75°}，とテンプレート数を変化させて実験を行った．なお検出処理に RANSAC を用いているため，検出試行回数の違いによって結果に差が出ないよう，従来手法では提案手法のテンプレートと同じ回数検出処理を試行した．

表 1．実験対象ランドマークシンボルの詳細

ランドマーク	シンボル画像例	異なり数
サークル K		6
ファミリーマート		4
セブンイレブン		5
ミニストップ		3
ローソン		4
三菱東京 UFJ 銀行		3
KFC		3
ドコモショップ		4
地下鉄		3

3.2 実験結果

検出精度の評価には再現率及び適合率を用いた．図 5 に適合率が 100%になる（誤検出がない）ようパラメータ調整したときの再現率のグラフを示す．再現率について，従来手法では最大 54.7%であったが，提案手法では再現率は最大 71.1%となり，提案手法の効果が確認できた．

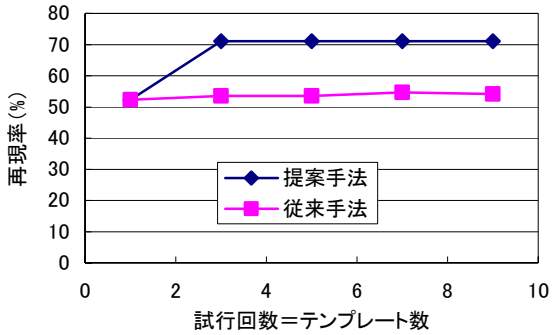


図 5．再現率のグラフ

このときの，ランドマーク毎の検出結果を表 2 に示す．従来手法，提案手法ともに試行回数およびテンプレート数を変えたときに最大となった再現率のみを示した．この結果から良好に検出できるランドマークと検出が難しいランドマークが存在することがわかった．

表 2．ランドマーク毎の最大再現率

ランドマーク	提案手法の最大再現率 (%)	従来手法の最大再現率 (%)
サークル K	5.56	4.76
ファミリーマート	94.1	84.5
セブンイレブン	86.7	66.7
ミニストップ	81.0	57.1
ローソン	63.1	40.5
三菱東京 UFJ 銀行	98.4	77.8
KFC	92.1	74.6
ドコモショップ	79.7	61.9
地下鉄	38.9	24.1

3.3　考察

図 5 から，提案手法はテンプレート数が 3 以上において再現率の向上が見られない．これは本実験で用いたデータセットに，±60゜を超えるような極端な角度変化が含まれていなかったためであると考えられる．もともと SIFT 特徴量はある程度の角度変化に対応できるため，±30゜程度の角度変化によるテンプレート生成で効果が飽和してしまったものと考える．より極端な変化が含まれるデータセットで実験をすることで，提案手法の効果がより詳細に確認できると思われる．

図 6．従来手法による実験結果例
（全対応数：37，inlier と判断された対応数：0）

図 6 に従来手法による実験結果を示す．対応付いたテンプレート上の特徴点と入力画像上の特徴点が線で結ばれて表示されている．37 個の対応が得られたが，RANSAC の結果 inlier と判断された対応はなく，シンボルを検出することができなかった．

図 7 に図 6 と同じ入力画像を提案手法で処理した結果を示す．図 6 と同様に，対応付いたテンプレート上の特徴点と入力画像上の特徴点を線で結んで表示した．図 7（a）は得られた対応全てを線分により図示したものである．また，図 7（b）は inlier だけを線分により図示したものである．提案手法では 38 個の対応が得られ，RANSAC の結果そのうち 15 個が inlier と判断され，シンボルを検出することができた．従来手法と提案手法では得られた対応の総数はほとんど変わらなかったが，シンボルを検出できたのは提案手法だけである．このように，SIFT によるシンボル検出では特徴点の対応の数よりも，より類似した特徴の対応点の組を見つけることが重要である．

表 2 によると，サークル K と地下鉄のランドマークに関して特に検出漏れが多く発生していることがわかる．表 1 の画像例を見ると，この 2 つのランドマークは他に比べてテクスチャが単純だとわかる．SIFT 特徴量はエッジに基づく特徴量であるため，これらのテンプレートでは検出に必要な特徴点が十分に得られなかったことが検出漏れの原因だと考えられる．少ない特徴点をより有効に活用していくことが，この問題を解決する上で重要になると考える．具体的には，検出に有効なテンプレート上の特徴点

(a)　全ての対応を表示した結果　　(b)　inlier のみを表示した結果

図 7. 提案手法（θ =30°）による実験結果例
（得られた対応数：38，inlier と判断された対応数：15）

をあらかじめ学習によって見つけておくという解決法が考えられる．

4. まとめ

本稿では，実環境における見えの変化に頑健なランドマークシンボルの検出手法を提案した．提案手法は，様々な見えの変化をモデル化し，それを用いて複数のテンプレートを自動生成し，更にそれらを SIFT 特徴量を用いたテンプレートマッチングに用いるものである．実験では実画像中からのランドマークシンボル検出に本手法を適用した．その結果，適合率を 100%としたときの再現率は，テンプレート生成を用いない従来手法では最大 54.7%であったのに対し，生成された複数のテンプレートを用いる提案手法では最大 71.1%となり提案手法の効果を確認した．

今後は低解像度化やぶれなどの他の生成モデルの検討，テクスチャが単純なテンプレートを用いた場合の，検出精度の向上を目指す．

謝辞

日頃から熱心に御討論頂く名古屋大学村瀬研究室諸氏に感謝する．本研究の一部は，日本学術振興会科学研究費補助金による．

参考文献

[1]　多賀大泰，高橋直久，“ランドマーク視認状況に基づく歩行者の位置特定システム”，DBSJ Letters Vol.5, no.1, pp.93-96, 2006

[2]　山口高康，青野博，本郷節之，“モバイルカメラで撮影した看板画像の特徴量に関する考察”，信学技報，PRMU2004-105, 2004

[3]　山口高康，青野博，本郷節之，“モバイルカメラで撮影した看板画像の学習・判別手法に関する考察”，信学技報，PRMU2004-106, 2004

[4]　Naoyuki Ichimura, “Recognizing Multiple Billboard Advertisements”, Proc. IEEE Pacific-Rim Symposium on Image and Video Technology 2006, pp.463-473, 2006

[5]　高木雅成，藤吉弘宣，“SIFT 特徴量を用いた交通道路標識認識”，第 13 回画像センシングシンポジウム，LD2-06, 2007

[6]　D.G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, IEEE International Journal on Computer Vision, Vol.60, no.2, pp. 91-110, 2004

[7]　M.A. Fischer, R.C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography”, ACM Graphics and Image Processing, vol.24, no.6, pp.381-395, 1981

[8]　O. Chum, J. Matas, “Matching with PROSAC – Progressive Sample Consensus”, Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2005, Vol.1, pp.220-226, 2005