ステレオ視とSIFT特徴点追跡を用いた移動ロボットのための環境地図生成

Environmental Mapping for Mobile Robot
by Stereovision and Tracking of SIFT Feature Point

○学　小川 陽子（立命館大）　正　島田 伸敬（立命館大）　正　白井 良明（立命館大）

Yoko OGAWA, Ritsumeikan University
Nobutaka SHIMADA, Ritsumeikan University
Yoshiaki SHIRAI, Ritsumeikan University

This paper presents the method of generating the environmental map using stereo vision and SIFT(Scale Invariant Feature Transform) feature. The purpose of this research is to make the map to estimate the robot position when moving with a monocular camera. The feature points are tracked based on the distance of SIFT feature vector, and their 3D position are estimated with the Kalman filter. Each feature point on the map is composed of the 3D position with its covariance, some SIFT feature vectors, the camera position and the glance vector when this feature point was ever observed. By using this environmental map, the 2D robot position and orientation can be estimated without any stereo cameras or range sensors.

1. 緒言

ロボットが目的地まで自律移動するとき，移動に不確かさがある場合や環境が未知あるいは動的である場合，周囲の状況を知る必要がある．ロボットのナビゲーションにおいて，レーザ距離センサも広く使われているが [1]，視覚では距離情報だけでなく色や模様などの視覚情報を同時に取得できるため，本研究では視覚センサを用いて地図生成を行った．視覚情報には一般に不確かさがあるため，必要な精度・信頼性で情報を得るには，複数の情報を統合することが重要である．今回は，上記の点を考慮して，ステレオカメラによる視覚センサ情報と複数地点での観測結果を統合して地図を作成する方法を述べる．本研究の目標は，上記方法で生成された地図と単眼カメラを用いた移動ロボットの室内自律移動を実現することである．

2. 視覚情報による移動量推定

本研究では Point Gray Research 社の三眼カメラ Digiclops を使用する．ステレオカメラから得られた 3 枚の画像中の特徴点を抽出・マッチングし，さらに時系列で得られた画像間でもマッチングを行い，特徴点を統合し，その 3 次元位置とその点の見えを地図に登録していく．

2.1 SIFT

2 次元画像から得られる特徴点（キーポイント）には SIFT(Scale Invariant Feature Transform) 特徴量 [2] を付与した．SIFT 特徴量は，スケールの変化に不変であり，キーポイント周辺の情報も同時に保存するため，主に 2 次元画像同士の特徴点マッチング（複数画像からのパノラマ画像生成など [3]）に用いられている．1 章でも述べたように，本研究では走行時に単眼カメラのみを用いて移動することを目的としているため，移動時に得られる観測は二次元情報しか持たないと想定している．そこで，2 次元の見えの情報を，ある程度圧縮した上で保存できる方法として，この SIFT 特徴量を用いることにした．それぞれのキーポイントは，画像座標，スケール，オリエンテーション，特徴ベクトルを持つ．これらを用いて対応付けを行う．なお，オリエンテーションとは，キーポイント周辺の画素の輝度勾配方向のヒストグラムがピークになる方向で，すなわちキーポイント周辺の主な勾配方向と考えられる．

2.2 キーポイントのマッチング

まず，カメラから得られた 3 枚の画像からキーポイントを抽出する．以下では，向かって右下に配置されたカメラを右カメラ，その上と左のカメラをそれぞれ上カメラ，左カメラとし，それぞれのカメラから得られた画像をそれぞれ右画像，左画像，上画像と呼ぶことにする．右画像と上画像，右画像と左画像それぞれの画像間でキーポイントマッチングを行い，両方で対応が取れたキーポイントのうち，左右の視差と上下の視差の比が 8 割以内になった点を抽出する．画像間のキーポイントの対応条件は，SIFT の対応付け [2] の条件に加えて，エピポーラライン上の点であること，視差・スケールの差・オリエンテーションの差・特徴ベクトル間の距離が全て閾値以下であることとした．複数のキーポイント同士が対応付いた場合，不安定な対応とし

て削除する．ここで抽出されたキーポイントそれぞれについて，右カメラ座標系における３次元座標をステレオの原理により求め，ステレオ観測誤差分散行列も計算しておく．

2.3 キーポイントの時系列追跡

次に時系列でのキーポイント追跡について述べる．時刻 t における右画像と時刻 $t+1$ における右画像でマッチングを行う．時系列でのマッチングは，画像座標の縦横それぞれの差，スケールの差の割合，視差の差の割合，オリエンテーションの差・特徴ベクトル間の距離が閾値以内という距離の条件を用いてマッチングを行った．複数の特徴点同士が対応付いた場合は 2.2 と同様に対応を削除する．

2.4 移動量算出

地面に対して垂直方向は移動しないものとして，2 次元並進量と回転角度 (x, z, θ) の 3 パラメータのみを求める問題を考える．ステレオ視による三次元情報は視差が小さいほど観測誤差が大きくなるので，対応するキーポイント間のマハラノビス距離の総和を最小にするパラメータを求めることで移動量を算出する．各キーポイントには，正規化した視線ベクトルと観測位置を付与しておく．

3. 環境地図生成

ステレオカメラから得られたキーポイントを地図に登録する．ワールド座標系の原点は観測を開始した位置とする．以下では 2 次元画像から得られた点をキーポイントとし，地図に登録される分散を持つ 3 次元の点を特徴点とする．

3.1 カルマンフィルタ

特徴点の 3 次元位置は，ステレオ視の原理から，視差が小さいほど観測誤差が大きくなる．そこで，観測値に誤差が含まれる場合に物体の状態を精度良く推定することが可能なカルマンフィルタ [4] を用いて特徴点の位置推定を行った．カルマンフィルタは確率過程に基づいたフィルタリング理論の１つであり，以下の式で定義される．

$$\hat{\mathbf{x}}_k = \tilde{\mathbf{x}}_k + P_k C_k' W^{-1} \{\mathbf{y}_k - (C_k \tilde{\mathbf{x}}_k + \bar{\mathbf{w}}_k)\} \quad (1)$$

$$\tilde{\mathbf{x}}_k = A_{k-1} \hat{\mathbf{x}}_{k-1} + B_{k-1} \bar{\mathbf{u}}_{k-1} \quad (2)$$

$$P_k = (M_k^{-1} + C_k' W_k^{-1} C_k)^{-1} \quad (3)$$

$$M_k = A_{k-1} P_{k-1} A_{k-1}' + B_{k-1} U_{k-1} B_{k-1}' \quad (4)$$

ここでは，時点 k における n 次元ベクトル値をとる物体の状態信号 $\hat{\mathbf{x}}_k$ を特徴点の三次元位置とする．$\hat{\mathbf{x}}_k$ は線形の差分式 $\mathbf{x}_{k+1} = A_k \mathbf{x}_k + B_k \mathbf{u}_k$ で支配されているとする．$\hat{\mathbf{y}}$ は観測，C は変換行列であり，$\mathbf{u}$，U，$\mathbf{w}$，W はそれぞれ制御と観測の誤差ベクトル，誤差分散である．移動量の推定が正確であり，特徴点は制御を受けず移動しないと仮定すると，A は単位行列，$\mathbf{u}$ と $\mathbf{w}$ と U は $\mathbf{0}$ となる．これを踏まえて式 (1)~ (4) を書き直すと，

$$\hat{\mathbf{x}}_k = \tilde{\mathbf{x}}_k + P_k C_k' W^{-1} \{\mathbf{y}_k - C_k \tilde{\mathbf{x}}_k\} \quad (5)$$

$$\tilde{\mathbf{x}}_k = \hat{\mathbf{x}}_{k-1} \quad (6)$$

$$P_k = (M_k^{-1} + C_k' W_k^{-1} C_k)^{-1} \quad (7)$$

$$M_k = P_{k-1} \quad (8)$$

となる．このとき，C はカメラ座標系からワールド座標系への座標変換行列であり，具体的には式 9 のようになる．

$$\mathbf{C} = \begin{bmatrix} \cos\theta_k & 0 & -\sin\theta_k & -(x_k \cos\theta - z_k \sin\theta_k) \\ 0 & 1 & 0 & 0 \\ \sin\theta_k & 0 & \cos\theta_k & -(x_k \sin\theta + z_k \cos\theta_k) \end{bmatrix} \quad (9)$$

W はステレオ視の観測誤差である．なお，観測 y の次元数は 3 であるが，C は並進成分を表現するため 3×4 行列となり，状態の次元数は 4 となる．

3.2 地図更新

時系列で対応が取れたキーポイントに関しては，その対応を用いてカルマンフィルタを更新する．このとき，同一特徴点とみなされたキーポイント群は一元化して登録しておく．首振りなどにより，一度見えなくなり再度見えた点に関しては，自己位置推定結果から得られたワールド座標系における 3 次元位置を基に近傍の点を対称にマッチングを行う．マッチングには，対象となった点に登録されているキーポイントのうち，視線ベクトルの差が 20 °以内になるものを用いる．このときも複数の点と対応が付いた場合は対応を削除する．

4. 実験結果

実験は立命館大学クリエーションコア４ F のコンピュータビジョン研究室内にて行った．デスクトップ PC（Pentium4 2.80GHz，メモリ 512MB，OS Windows XP Professional Version 2002 SP2）と前述の三眼ステレオカメラを IEEE1394 で接続し，カメラは三脚に乗せ，手で押しながら計 28 回の観測を行った．プログラムは OpenCV を用いた C++言語で記述した．なお，画像の解像度は 320 × 240 である．

4.1 キーポイント抽出

本研究では，キーポイントを抽出する際，オクターブ数を 1，1 オクターブ内のぼかし画像の数を 7 枚（DoG 空間のキーポイント探索領域を 4 枚）にしている．さらに，処理の高速化のため画像拡大を省略し，代わりにガウシアンフィルタをかける際の標準偏差 σ を本来 1.6 の半分 0.8 に設定している．右画像に 3 つの画像間で対応が付いた点を白で，そのときの視差を黒のラインでプロットし，図 1 に示す．なお，横のライン長は左と右の視差を，縦のラインは上と右の視差を表す．閾値はそれぞれ，視差 30 ピクセル以内，スケールの差± 1 以内，オリエンテーションの差± 20 °以内，特徴ベクトル間の距離 0.05 以内とした．

Fig. 1: Stable feature point

図 1 のとき，得られた特徴点の数は 111 個であった．28 枚全てを実行したとき，1 枚の画像から得られる特徴点数は，平均 123.07 個，標準偏差は 14.14 であった．

4.2 キーポイントの時系列追跡

時系列でのマッチングにおける閾値は，画像座標の縦横それぞれの差 40 ピクセル以内，スケールの差の割合 2 割以内，視差の差の割合 2 割以内，オリエンテーションの差 20 °以内，特徴ベクトル間の距離 0.05 以内とした．27 回のフレーム間マッチングで対応が付いた特徴点数は，平均 59.56 個，標準偏差 9.35 であった．

4.3 移動量算出と環境地図生成

2.4 で述べた方法に基づき移動量を算出し，観測された点のワールド座標系での三次元位置を求め，地図を生成・更新する．今回はコーディングコスト削減のため，全探索（x と z は 10mm ごと 300mm，θ は 1 °ごとに 20 °）で移動量を算出した．図 2~ 4 に，2・10・28 番目の観測を終えた時点で地図に登録されてある特徴点の誤差楕円とカメラの軌跡の俯瞰図，そのときの右画像を示す．地図の原点は画像中央の一番下とし，1 ピクセル 1cm として描いた．ただし，特徴点は 2 回以上観測された点のみを描いてある．カメラ軌跡の終点の一回り大きい円と中心から伸びる線は，そのときのロボットの姿勢を想定して描いたものである．右画像には 3 画像間では対応が付いたが前のフレームと対応が付かなかった点を白，前のフレームとも対応が付いた点を黒で示した．このときの地図上の点数を表 1 に示す．なお，右画像中の黒い点から伸びる線の先の座標は，前のフレームで対応が付いた点の座標である．このときキーポイント抽出とマッチングに平均 4.88[sec]，移動量推定に 7.14[sec] の処理時間を要した．

Table 1: Number of feature point on the map

フレーム数	2	5	10	28
地図上の点数	171	319	580	1676

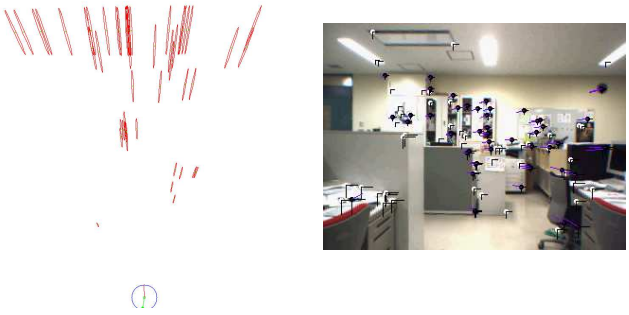

Fig. 2: nvironmental map and right picture No.2

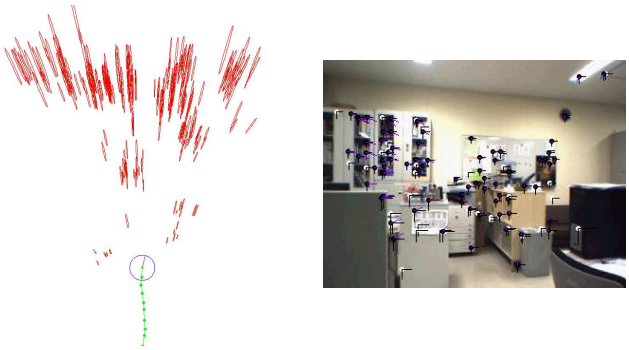

Fig. 3: nvironmental map and right picture No.10

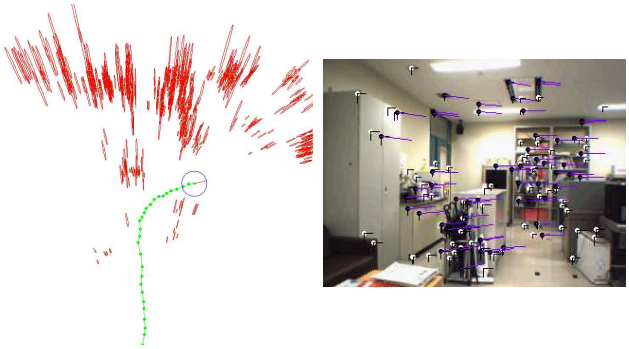

Fig. 4: nvironmental map and right picture No.28

5. 考察

実験では，キーポイント抽出とマッチング，および移動量推定に時間がかかり過ぎているという問題があった．移動量の推定で全探索している部分は，今後最急降下法などの方法を用いて高速化していく予定である．さらに，移動量の推定時，どの程度のキーポイントの対応数でどの程度の精度で移動量推定できるのかといった評価をしていきたい．環境地図生成は，図 2 では特徴点の数も少なく，それぞれが大きな分散で登録されていたが，フレームが進み図 4 に至るに連れ，分散が収縮し点が増えていく様子が確認できた．

5. 今後の展望（自己位置推定）

上記方法で生成された地図から，ある姿勢での推定画像（その姿勢でどのように見えるか）を作成し，入力画像と比較し自己位置を推定していく．現時点では図 5 のように推定画像を生成した．図 5 は $(x, z, \theta) = (10, 1000, -10)$ の姿勢時の推定画像である．この結果とこの座標の近辺で得られた入力画像を比較し自己位置推定を行い精度を検証していく．

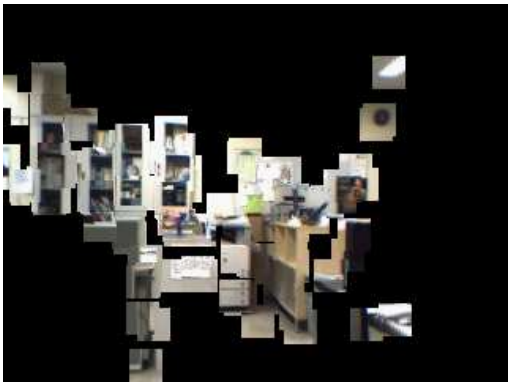

Fig. 5: Estimated image

6. 結言

本研究では SIFT を用いた画像ベースの地図生成方法を提案した．今回の地図を使った推定画像は，今後ロボットの自己位置推定のみならずユーザへの情報提示に役立てられないか検討していきたい．また，環境の変化への対策方法は未だ検討中ではあるが，SIFT の特性を生かして「環境が変化した」というだけでなく,「何がどう変化した」のか（例えば「何か物が置かれた」,「別の場所で見えた」など）を記憶することで，新たな利用法についても考えていきたい．

参考文献

[1] 中本 琢実, 山下 淳, 金子 透: ”レーザレンジファインダ搭載移動ロボットによる動的環境の 3 次元地図生成”, 映像情報メディア学会技術報告, Vol.30, No.36, pp.25-30, 浜松, July 2006.

[2] David G. Lowe: *Distinctive Image Features from Scale-Invariant Keypoints*, publication in the International Journal of Computer Vision, 2004.

[3] Matthew Brown, David G. Lowe: *Recognising panoramas*, International Conference on Computer Vision (ICCV 2003), Nice, France (October 2003), pp. 1218-25.

[4] 有本 卓: ”カルマン・フィルター”, 産業図書株式会社, 1977.