計測自動制御学会東北支部 第 303 回研究集会 (2016.7.15)
資料番号 303-2

単眼カメラとSfMによる広域環境地図の構築に関する基礎検討

Basic Study on Wide-Range Map Creation using Monocular Camera and SfM

○上田慎也[1†]，間所洋和[1]，佐藤和人[1]，下井信浩[1]

○ Shinya Ueda[1†], Hirokazu Madokoro[1], Kazuhito Sato[1],and Nobuhiro Shimoi[1]

*秋田県立大学 大学院 機械知能システム学専攻

*
Department of Machine Intelligence and Systems Engineering, Akita Prefectural University, Akita, Japan

キーワード： SLAM (Simultaneous Localization and Mapping)，EKF (Extended Kalman Filter)，SfM (Structure from Motion)，Sobel フィルタ (Sobel Filter)

連絡先： 〒 015-0055　秋田県由利本荘市土谷字海老ノ口 84-4 秋田県立大学 システム科学技術研究科 機械知能システム学専攻 脳情報工学研究室
上田慎也，Tel.:0184-27-2000，FAX:0184-27-2180,E-mail: madokoro@akita-pu.ac.jp

1. はじめに

近年，ロボット産業の発展に伴い，人間の生活環境で利用される自律移動ロボットの開発が活発になってきている．一般に，このようなロボットが生活環境内で自律移動するためには，周囲の地図情報を用いてロボットの自己位置を知る必要がある．しかしながら，車輪やクローラにより地上を走行するロボットでは，移動範囲が平面に制限されるため，すれ違う人や他のロボットによって，センシング範囲が遮られてしまう．一方，3 次元空間を自在に移動できる MAV(Micro Air Vehicle) はこのような制約から開放されるため，運動能力と自由度の高さから応用範囲が広い．このため，MAV は日本のみならず，世界的に普及が進んでいる．しかしながら，MAV のフライト時間は搭載するバッテリの容量に依存するため，バッテリ切れによる墜落事故が後を絶たない．また，GPS が捕捉できる屋外と違って，屋内環境の飛行では環境地図の自動構築と自己位置推定が必要不可欠になる．環境地図の構築には SLAM を使用する．試作した MAV の前後左右 4 台のカメラを搭載し，高速かつ広範囲に環境センシングする．また，計測機器の簡便性と MAV の積載重量（ペイロード）を考慮して，LRF やステレオカメラではなく，単眼カメラを利用する．カメラを使用した場合，距離情報だけではなく色や模様などの情報も同時に取得できるため，環境の意味認識や物体認識への応用が期待できる．VisualSLAM の中でも Davison らの MonoSLAM [1)] をベースとするのは，SfM(Structure from Motion) と EKF（Extended Kalman Filter）を組み合わせた基礎的な手法となっているためである．従って，全方位センシング用の専用機体を独自に製

作するとともに，MonoSLAM で課題となっていた，広範囲の地図構築を目的とする．

2. 関連研究

工場等の限定的な環境ではユーザがあらかじめ環境地図を構築し，ロボットに与える手法が取られていた．しかしながら，時々刻々と変化する人間の生活環境において，あらかじめ環境地図を構築することは困難であるため，ロボットが自律的に地図構築と自己位置推定を行う必要がある．環境地図の構築方法として SLAM が活発に研究されている．SLAM ではロボットに搭載された LRF やステレオカメラなどの距離センサから，壁や物体までの距離情報を取得し，グローバル座標系における自身の位置と姿勢を推定して，地図を構築する．従来の SLAM では，高速で正確な距離計測を行える LRF が多用されていたが，装置が高価なことに加えて，一次元の走査情報しか得られないため，広範囲のセンシングには不向きであった．

近年，ステレオカメラや Microsoft 社の Kinect のような深度センサを用いた SLAM が注目されている．カメラを使用するこれらの SLAM は，視覚情報が同時に得られるため，Visual SLAM と呼ばれる．Visual SLAM は，上記の通りロボティクス分野にルーツを持つ技術であるが，単眼カメラに SfM という技術を組み合わせた単眼 SLAM が，近年注目を集めている．

SfM はコンピュータビジョンにルーツを持つ技術で，多視点画像から対象物体や環境全体の 3 次元形状とカメラの運動を推定する．3 次元形状とカメラの運動の両方を同時に推定する初期の方法のひとつに Harris-Pike 法 [2] がある．この手法では，カメラ運動を非線形最適化で求め，特徴点の位置推定のために線形の KF（Kalman Filter）が用いられる．特徴点の位置をユークリッド座標で表現することや，画像座標と視差を使用して表現するなど，カメラ運動の最適化のために特徴点の位置の共分散行列を利用して高精度化が図られており，近年の方法に通じる点が多い．カメラ運動と点の位置を，KF を用いて同時推定する，現代的な最初の方法として Broida らの研究 [3] がある．これは単眼カメラで追跡した特徴点を用いて，EKF によってパラメータを推定する．更にこの手法は，Azarbayejani-Pentland [4] で拡張されて，焦点距離が同時に推定できるようになった．また，Chiuso-Satto らの MFm [5] では特徴点の消失や新たな特徴点の追加に伴う状態変数の削除と追加の方法が検討されている．これは単一の EKF を使用して，地図と特徴点の不確実性について検討しているものの，小型カメラの動作による小グループ物体の追跡に留まっていた．McLanchlan-Murray [6] は同時構造とスパース情報のフィルタフレームワークを使用して，移動カメラから動き情報を復元する VSDF（Variable State Dimension Filter）を導入した．これはバンドル調整の解説 [7] にも記載されており，状態変数を全て最適化する完全バンドル調整と，EKF の間を繋ぐ手法となっている．しかし，VSDF は長期的な追跡やループ閉鎖の実証できなかった．

Visual SLAM に SfM を最初に導入したのが Davison らの MonoSLAM [1, 8] である．Davison は単眼カメラを使用し，カメラの自己位置の推定と環境地図の構築を同時に行うという目標を明示し，それを汎用性の高いツールとともに実現した．MonoSLAM はカメラの姿勢とランドマークの状態変数をとり，標準的な EKF を使用する．その平均と共分散行列をビデオレートで更新をする手法は従来の研究と共通している．特徴点がカメラの運動に伴い一度視野外になった後に，再び視野内に戻ってきた場合，その点を同じ特徴点として認識できるメカニズムに新規性がある．しかし，EKF を使用した SLAM には，次に示す 2 点の課題が残されている．1 点目は，特徴点の増加に伴い計算量が爆発的に増えるため，MonoSLAM は 2 桁程度の特徴点を

扱うのが限界であり，広い範囲の地図構築には適していないことである．2 点目は，EKF は観測モデルを線形近似し，状態変数の分布をガウス分布で近似する方法のため，推定精度を維持するには近似精度が正確性の確保が必須になることである．

これらの課題に対して，SLAM を高速化した FastSLAM [9] が提案されている．Eade-Drummond は FastSLAM に基づいた単眼 SLAM システムを示した [10]．Nister [11] らは，SfM 法に基づいたリアルタイムシステムを利用し，多数の画像の一部に対して最適化を繰り返すことにより，高精度に SLAM を実行できることを示した．また Mouragnon らは，単眼カメラから得た画像の一部にバンドル調整を適用することで精度を維持しつつ，逐次的に 3 次元復元を行う方法を提案した [12, 13]．これらは超音波等の距離センサを使用せず単眼カメラのみでカメラ運動から距離を推定できることから，Visual Odometry と呼ばれる．しかしながら，一度視野外に外れた特徴点を再認識するメカニズムは組み込まれておらず，急速なドリフトをもたらしてしまう．特徴点に Harris コーナを使用し，その追跡を相関ベースで行っているものの，前述の Nister [11] と同様に，特徴点を再認識するメカニズムは組み込まれていない．以上のように，SfM を使った単眼 SLAM の研究は，MonoSLAM の提案から 10 年を経過して様々な手法が提案されているが，いずれの手法も Davison らの考え方 [1] が踏襲されている．加えて，特徴点を再認識するメカニズムに関しては，MonoSLAM に比較優位性がある．ただし，広範囲での地図構築が課題として残されている．

3. SLAM

3.1 SLAM の概要

SLAM の環境計測用センサには，超音波や LRF，ステレオカメラ、Kinect のような深度センサ等が用いられている．特に近年，ステレオカメラや Kinect を用いた Visual SLAM が活発に研究されている．

LRF を使用した SLAM は，距離精度が高く，少ない計算コストで環境地図が生成できる．しかし，センサが非常に高価で重量もかさむため，一般的に平面を移動するロボットに搭載される．また，レーザは指向性が高いため，LRF では水平方向に対して機械的に走査するが，垂直方向の情報は得られない．

ステレオカメラとは，2 台のカメラを用いて，三角測量に基づいて距離情報を取得するカメラシステムである．ステレオカメラは，情報量が多い，視野が広い，処理速度が速いなどの利点がある．しかしながら，ステレオカメラは高価で，重量がかさむ．また，距離情報の信頼性を向上させるためには三角測量のための精度の高い特徴点検出が必要となる一方，この処理の精度を追求すれば，全体の処理速度が落ちる．反対に，処理速度を速めた場合，特徴点抽出の精度が犠牲になるため距離情報の信頼性が落ちる．また，ステレオカメラでの計測精度は，レンズの解像度，基線長によって変わる．

超音波センサは，安価で，有効測定距離が長いが，指向性高く干渉による誤差が生まれやすい．また，照射角度に対する誤差が大きいという問題がある．更に，単点を測るセンサのため，LRF やステレオカメラのように，一度に広範囲の情報を得ることが困難である．

3.2 単眼 SLAM

単眼 SLAM とは Visual SLAM の一種で，複数のカメラを使用せずに，単眼カメラのみで SLAM を実現する．単眼カメラのみでは距離情報が得られないが，SfM を組み合わせることにより，SLAM が実現できる．また，他のセンサと比べ安価であり，カメラの間の同期を取る必要がないため，メカニズムが簡素になる．単眼 SLAM は古典的な SLAM と違い，アクティブビジョンと

してのカメラの積極的な動作を活用して，複数の視点から特徴点を取り，距離情報を算出する．

単眼 SLAM には様々な手法が提案されているが，2 章の関連研究で述べた通り，Davidson らの MonoSLAM [1] は，特徴点がカメラの運動に伴い，一度視野外になった後に，再び戻ってきた場合，その点を同じ特徴点として認識できる点に，他の手法と比較して，絶対的な優位性を持っている．従って，本研究においても，MonoSLAM をベースに，広域環境センシングへの拡張を図る．

4. 特徴点検出

4.1 エッジ検出（Sobel フィルタ）

エッジとは，一般に濃淡が急激に変化している箇所を指し，物体の構造を反映している重要な情報である．本研究では 8bit グレースケール画像を使用し，画像中の濃淡は 0〜255 の数値で表される．デジタル画像は離散データなので，微分は差分で計算することができるが，一般的に画像内には多量のノイズが含まれるため，平滑化と組み合わせることが多い．そこで，平滑化と微分の特徴を組み合わせた Sobel フィルタを用いる．Sobel フィルタとは，1 次微分によりエッジ検出を行うフィルタである．Fig.1 に注目画素に対する 3 × 3 画素の処理対象領域を示す．Sobel フィルタのマスクパターンは 3 × 3 の行列で表される．x 方向と y 方向におけるマスクを Fig.2 に示す．@を処理対象の中心画素として，周囲の数字は中心画素に対する重みを表す．この行列を入力画像の左端上から走査して，たたみ込み演算をする．その結果を共分散行列で表し，固有値を計算する．求めた固有値を λ_1，λ_2 とし，しきい値が λ を上回っている画素を特徴点として抽出する．

4.2 コーナネス

コーナネスとは画像における角らしさを表す．前述の共分散行列を 2 × 2 の正方行列に表したものを A とする．A に正則行列 P が存在すると仮定すると，固有値は次式で表す．

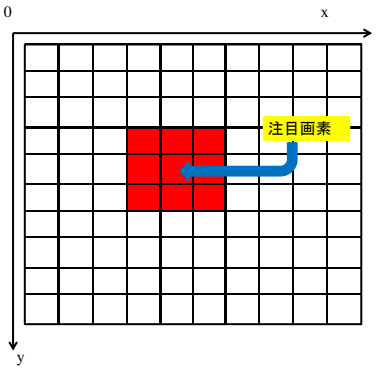


Fig. 1　注目画素に対する 3 × 3 画素の処理対象領域

-1	0	1
-2	@	2
-1	0	1

-1	-2	-1
0	@	0
1	2	1

Fig. 2　Sobel フィルタのマスクパターン

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \tag{1}$$

この行列の対角要素を計算し，特に値が大きい箇所をコーナネスとして抽出する．

5. 地図構築

5.1 SfM(Structure form Motion)

Davison らの MonoSLAM [1] では，SfM の枠組みにより，動画像から，Shi と Tomashi の因子分解法を用いて，被写体の立体形状とカメラ運動を復元している．特徴点の追跡には KLT（Kanede Lucas Tomasi）法を使用する．以下に SfM の処理手順を示す．

1) 動画像の 2 フレーム間で検出された特徴点とそれらが共に示す 3 次元空間の注目点の 3 点を関連付ける (Fig.3)．

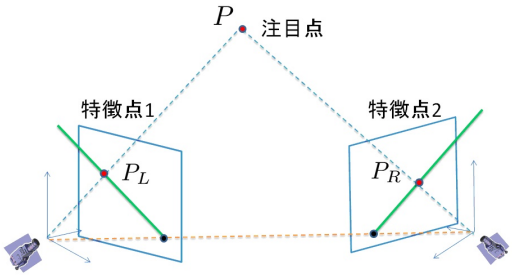


Fig. 3　検出された特徴点と注目点の関連付け

2) 手順 1 で検出した 3 点を結ぶ三角形を求め，特徴点 1 から特徴点 2 を計算するための基礎行列をエピポーラ拘束の関係式により計算する．

3) 2 フレームごとの基礎行列には，再投影誤差が含まれる．この誤差を最小にするように，2 フレーム組を多数用意し基礎行列を Bundle を用いて計算する．この方法で基礎行列を求めることによって， 各フレームでの特徴点の座標とカメラ位置を同時に求めることができる．例を Fig.4 に示す．

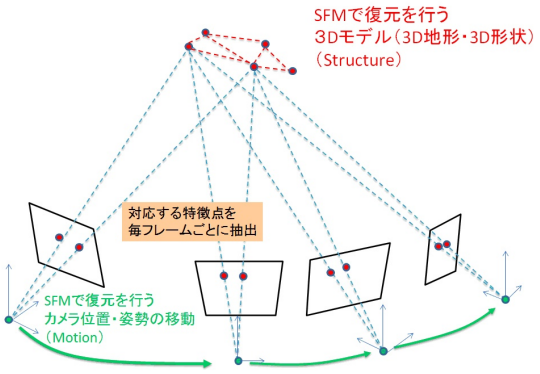


Fig. 4　各フレームの特徴点の座標及びカメラの位置

5.2　EKF

EKF とは，非線形フィルタリングと予測問題へのアプローチとして発表された KF を時系列変化にも対応できるように拡張したフィルタである．本研究では，構築した地図を連続的かつ動的に更新するために EKF を使用する．

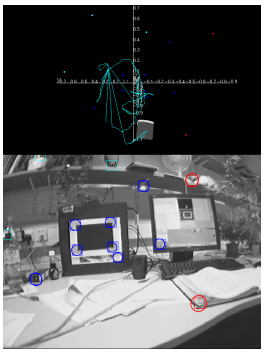
6.　評価実験

6.1　環境構築

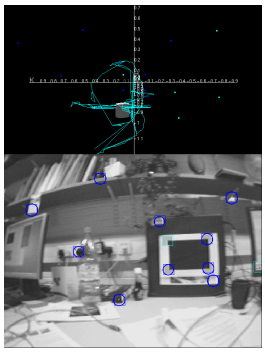
本研究では，単眼 SLAM のオープンソースライブラリである SceneLib 1.0 を使用した．SceneLib は Linux ベースで動作し，Davison らによって C++言語で実装されている．

6.2　ベンチマークデータセットの結果

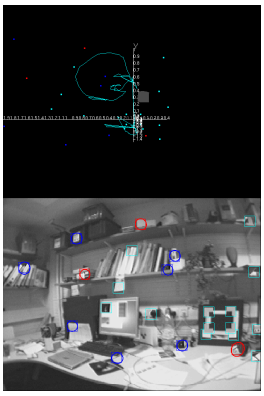
単眼 SLAM の基礎評価実験として，インターネットで公開されている Dvison らのベンチマークデータセットを使用して, 特徴パッチの抽出と環境地図の構築を実施した．このベンチマークデータセットは 999 枚の画像から構成され, 画像の解像度は VGA である．Fig.5 に，300，600，999 枚の結果を示す．赤枠は特徴点として追跡できているパッチ，青枠は特徴点の追跡に失敗しているパッチ，黄枠は一時的に追跡が行えなくなり，保存されているパッチである．ベンチマークデータセットでは，どのフレームでも比較的赤枠が多い．黄枠は少ない結果となっており，999 枚目だけは青枠が比較的多く現れている．カメラの軌跡に関しても，GT(Ground Truth) として，安定して抽出されている．



(a)300枚目



(b)600枚目



(c)999枚目

Fig. 5　ベンチマークデータの結果

6.3　オリジナルデータセットの結果

実環境での評価のために，秋田県立大学本荘キャンパスの学部2号棟3階にある会議室でオリジナルデータを取得した．部屋の広さは縦14.3m，横5.4mである．会議に使用する部屋なので，机，椅子，ホワイトボード以外の物は置かれておらず，シーン構成は極めて簡素である．カメラはGoProHERO+を4つ，Fig.6に示すカメラマウントの前後左右に取り付けをした．

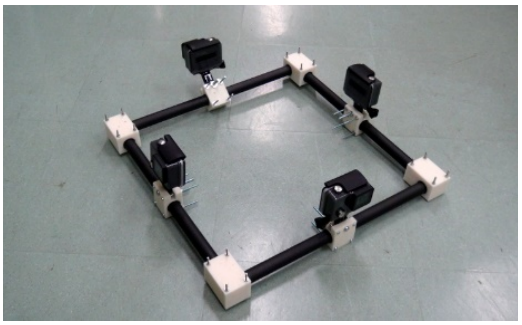


Fig. 6　カメラマウントとカメラ固定材料の組立外観

撮影時のカメラの高さは，人間の身長よりも少し高い2mとし，カメラマウントの前後左右に取り付けた4台のカメラを同時に動かした．ただし，カメラ間の同期は取っていない．撮影時間は8分程度である．撮影した映像を静止画像に切り出す際のフレームレートは10fpsに設定した．得られた静止画像のデータは，約3000枚になった．なお， GoProの最大解像度1980 × 1080pixelで映像データを取得し，MonoSLAMの入力の際，対応する解像度の320 × 240pixelへ変換した．動作モードはベンチマークデータセットによる評価実験と同様に，ディスク上のファイルから入力するオフラインモードとした．特徴点の表示は前節で説明した赤枠，黄枠，青枠で表される．

前節の実験と同様に，300フレーム毎の特徴パッチをFig.7,Fig.8に示す．300枚目では物置棚の角，600枚目では蛍光灯と天井の境界，2400枚目では机のコーナに赤枠が表れている．900，1200，1500，2100，2987枚目のフレームでは赤枠も青枠も表れていない．1800枚目では物置の上の物体，2700枚目では蛍光灯と天井の境界と机の淵に青枠が表れている．

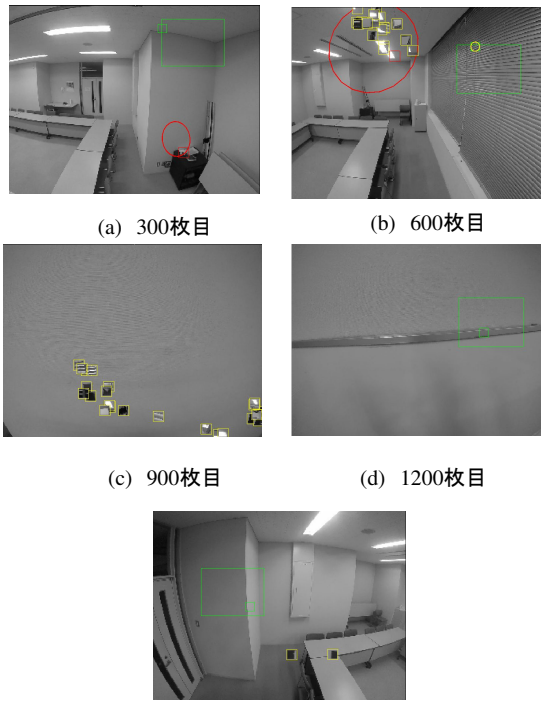


Fig. 7　オリジナルデータセット（300–1500枚目）

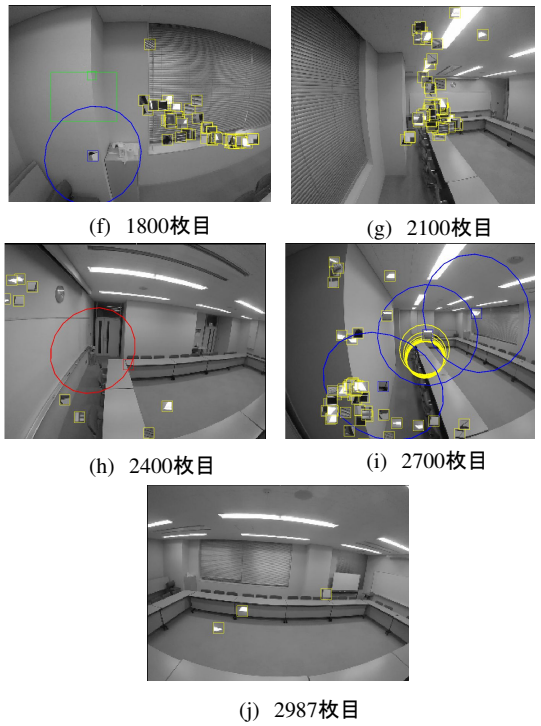


Fig. 8　オリジナルデータセット（1800–2987枚目）

続いて，全実験結果の中から，黄枠以外の特徴パッチが比較的出現しているフレームを任意に選出した．任意に選んだフレームは合計 14 枚で，それぞれ Fig.9，Fig.10，Fig.11 で示す．

(a) では，蛍光灯と天井の境界に青枠が 2 パッチ，ホワイトボードのコーナと机の黒淵に赤枠のパッチが出ている．また探索候補領域内にホワイトボードのコーナが表れている．(b) では，赤枠の 2 パッチと探索候補領域内のコーナが青枠に変わっている．また，黄枠の特徴パッチも青枠に変わっている．(c) では，机と机の境界に 1，蛍光灯と天井の境界に 2，赤枠の特徴パッチが出ている．(d) では，机のキャスタの赤枠が黄枠，蛍光灯と天井の境界の赤枠は継続して赤枠を示している．さらに，新たに机と椅子の境界と，蛍光灯と天井の境界に赤枠が出ている．(e) では，蛍光灯と天井の境界に赤枠が 3，椅子と机の境界に 1 赤枠が出ている．また探索候補領域内に椅子と机の境界が表れている．(f) では，探索候補領域内の黄枠と蛍光灯と天井の境界の赤枠が青枠に変わっている．また，2 つの赤枠は継続して赤枠として表れている．さらに，赤枠の 1 が黄枠に変わっている．(g) では，蛍光灯と天井の境界に 3，椅子と机の境界に 1，ドア上の窓と壁の境界に赤枠 1 が出ている．(h) では，赤枠 2 が黄枠に，ドア上の窓と壁の境界の赤枠が消失している．赤枠 1 が青枠に変わり，赤枠 1 は継続して赤枠として表れている．(i) ではドアノブとドア窓の端が赤枠の特徴パッチとして追跡されている．(j) では，(i) で獲得した赤枠が青枠になり，特徴パッチとして追跡されていない．(k) では．机と机の境界，机のキャスタ，ドア窓のコーナ，物置のコーナーに赤枠，蛍光灯と天井に 1 パッチずつ黄枠と青枠が出ている．(l) では，机と机の境界には赤枠が出ているが，物置のコーナ，ドア窓のコーナ，机のキャスタでは赤枠から青枠に変わっている．また，黄枠の特徴パッチも青枠に変わっている．(m) では，蛍光灯と天井の境界と，机と机の境界に 2，赤

枠の特徴パッチが出ている．(n) では，机と机の境界の特徴パッチは消失しており，蛍光灯と天井の赤枠は青枠に変わっている．

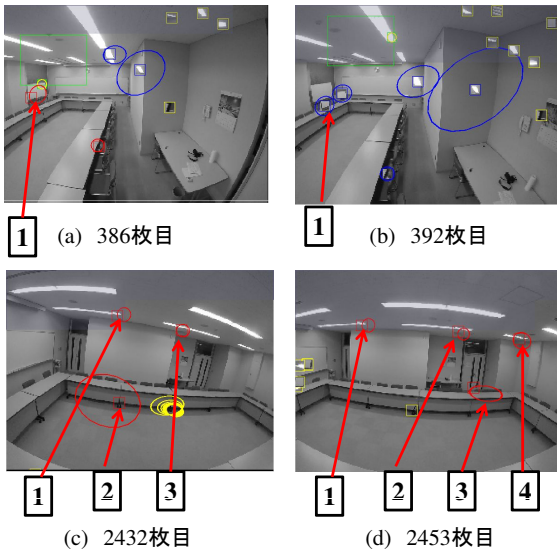


Fig. 9　任意に選定したフレーム（その 1）

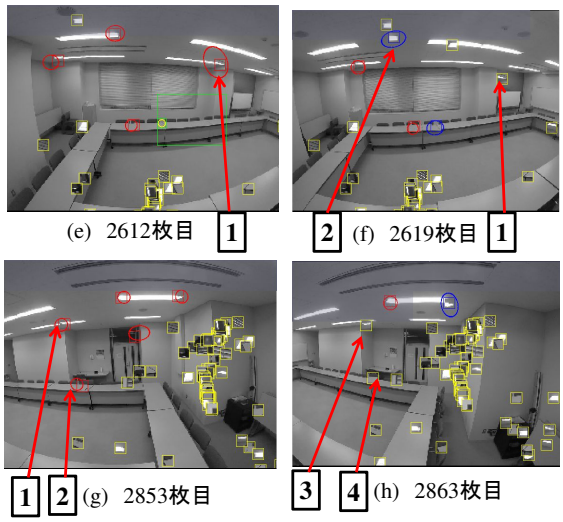


Fig. 10　任意に選定したフレーム（その 2）

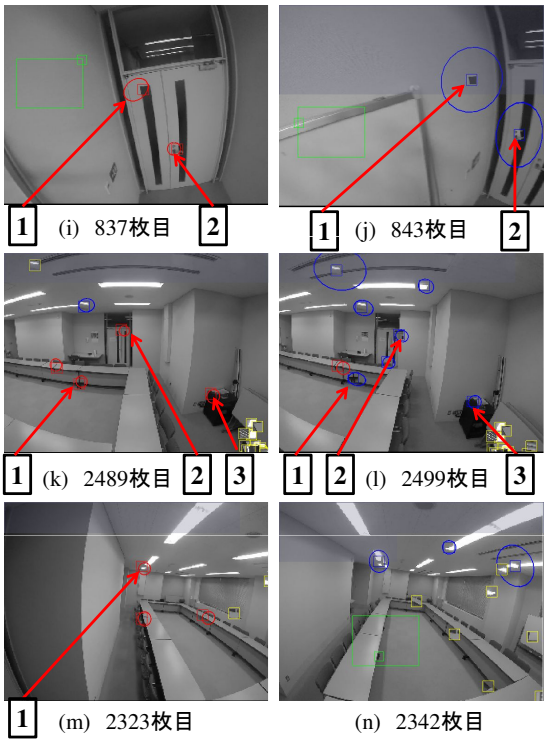


Fig. 11　任意に選定したフレーム（その 3）

7.　結論

7.1　研究成果

本研究では，広域環境地図の構築を目的として，SfM と EKF を組み合わせた単眼 SLAM による環境地図の構築を試みた．本研究を通じて，以下の成果が得られた．画像から特徴点を抽出するアルゴリズムを整理し，ベンチマークデータセットを用いて，特徴点の追跡，待機，及び破棄の基本的動作を確認した．高解像度カメラを用いてオリジナルデータセットを作成し，簡素な空間での特徴パッチの抽出について評価した．また，3 種類の特徴パッチの特性と 3 次元軌跡として構築された環境地図の違いから，SfM における MAV の飛行パターンについて，今後の実験の方向性を得た．

今後の課題として，以下が残されている．簡素な空間でのデータセットだけでなく，ベンチマークデータセットのような複雑な環境，廊下，吹き抜け等の映像を撮影し，地図構築精度の評

価を進めていく．今回の実験ではファイル読み込みによるオフライン処理になったが，オンラインモードでの処理を目指す．MonoSLAMだけではなく，物体認識と組み合わせたVisualSLAM等の他手法を調査し，独自の手法への発展を目指す．

参考文献

1) Davison, A.J., Reid, I.D., Molton, N.D. and Stasse, O. “MonoSLAM :Real-Time Single Camera SLAM,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.29, No.6, pp.1052–1067 2007.

2) Harris, C. and Puke, J.M., “3D positional integration from image sequences,” *Proc.3rd Alvey Vision Conf.*, pp.233–236, 1987.

3) Broida, T.J., Chandrashekhar, S. and Chellappa, R.“Recursive 3-D motion estimation from a monocular image sequence,” *IEEE Transactions on Aeroapace and Electronic Systems.*, vol.26, no.4, pp.639–656, 1990.

4) Azarbayejani, A. and Pentland, A.P. “Recursive Estimation of Motion,Structure,and Focal Length,” *IEEE Trans. Pattern Anal. Mach. Intell*, vol.17 pp.562–575, 1995.

5) Chiuso, A. and Soatto, S. “MFm: 3-D Motion From 2-D Motion Causally Integrated Over Time,” *Proc. ECVV*, pp.735–750, 2000.

6) McLaunchlan, P.F. and Murray, D.W. “A unifying frameeork for structure and motion recovery from image sequences,” *Proc ICCV*, pp.314–320, 1995.

7) B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, “Vision Algorithms: Theory and Practice,” *Lecture Notes in Computer Science*, vol.1883, pp.298–372, 2002.

8) Davison, A.J. “Real-Time Simultaneous Localisation and Mapping with a Single Camera,” *Proceedings of the Ninth IEEE International Conference on Computer Vision*, vol.2, p.1403, 2003.

9) Montemerlo, M, “FastSLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem with Unknown Data Association,” *PhD Thesis, Robotics Institute, Carnegie Mellon University*, 2003.

10) Eate, E. and Drummond, T. “Scalable Monovular SLAM,” *Proc. CVPR*, pp. 469–476, 2006.

11) Niste’r, D., Naroditsky, O. and Bergen, J. “Visual Odometory,” *Proc. CVRR*, pp. 652–659, 2004.

12) Mouragnon, E., Lhuillier, M., Dhome, M., Dekeyser, F. and Sayd, P. “Real Time Localization and 3D Reconstruction”, *Proc. CVPR*, pp.363–370, 2006.

13) Mouragnon, E., Lhuillier, M., Dhome, M., Dekeyser, F. and sayd, P. “Generic and real-time structure from motion using local bundle adjustment”, *Image and Vison Computing*, Vol.27, No.8, pp. 1178–1193, 2009.