TVRSJ Vol.17 No.3 pp.191-200, 2012

基礎論文

視点生成型学習による頑健な平面位置姿勢推定

吉田 拓洋[*1]　斎藤 英雄[*1]　清水 雅芳[*2]　田口 哲典[*2]

Robust Planar Pose Estimation by Viewpoint Generative Learning

Takumi Yoshida[*1], Hideo Saito[*1], Masayoshi Shimizu[*2] and Akinori Taguchi[*2]

Abstract – **We propose a planar pose estimation method that is robust to viewpoint changes. Conventional local features such as SIFT, SURF, etc., have scale and rotation invariance but often fail in keypoint matching when the camera pose significantly changes. To solve this problem we adopt viewpoint generative learning. By generating various patterns as seen from different viewpoints and collecting local features, our system can learn a set of descriptors under various camera poses for each keypoints before actual matching. Experimental results comparing usual local feature matching or patch classification method show both robustness and fastness of learning. Proposed method can achieve a markerless AR system that sets a tracking target on site.**

Keywords : **local features, pose estimation, generative learning, augmented reality**

1　はじめに

近年，実世界の平面上に仮想情報を重畳してみせる拡張現実感は，ARマーカーを使わないマーカーレスが注目されており，その実現に向けて，カメラの位置や姿勢の変化によらず画像中の対象平面を認識する技術が重要である．一般的には，あらかじめ登録された平面パターンと入力画像との間で正確な点対応を得ることで，こうした物体認識や拡張現実感を実現する．正確な点対応を得るにあたっては，画像から局所特徴を抽出することで位置や姿勢変化への頑健性を獲得している．局所特徴抽出は主に2段階で成り立っており，まず画像中から特徴点を検出し，次にその特徴点に対して高次元ベクトルで特徴量を記述する．特徴点同士を正確に対応付けするためには，対象がどのような見え方をしていても同じ特徴点を検出し，また見え方によらず特徴量が近い値に記述されることが求められる．

このような点検出の頑健性や特徴量記述子の不変性を達成しようとして，数多の局所特徴が提案されてきた．HarrisやHessianの検出器を発展させてアフィン変化への対応を施したBaumberg[4]やMikolajczykら[15, 14]の手法は画像の拡大縮小や回転に対する不変性をもつ領域検出の先駆けとなった．LoweのSIFT[13]は特筆すべき不変性をもつ著名な局所特徴の一つであり，Laplacian of Gaussian (LoG)を近似したDifferences of Gaussians (DoG)で特徴点を検出し，128次元の勾配ベクトルを特徴量として記述する．局所特徴による認識の精度を評価したMikolajczykらの研究[16]においてもその拡大縮小変化および回転変化への頑健性は非常に高い評価を示した．そしてSIFTの登場以降，そのアルゴリズムを改良した派生研究が多くなされてきた．PCA-SIFT[9]は主成分分析により特徴量の次元を128から36に削減し，マッチング時の距離計算を高速化させた．GLOH[16]はPCA-SIFTと同様に主成分分析を導入した勾配ベクトルを記述する局所特徴だが，ブロックではなく対数極座標 (Log-Polar) を用いる．Phony SIFT[21]はブロックを小さくし，勾配ベクトルの量子化レベルを低くしたSIFTの軽量版で，携帯デバイスでの実装に取り入れられた．SURF[5]はHessian行列と積分画像を用いることで高速な特徴点検出を，Haar waveletを用いることで高速な特徴量記述を実現した．OpenCVに実装されているSTAR detectorはCenSurE[1]の改良版で，SURF以上の高速化と拡大縮小への強い不変性を達成した．また広角条件の両眼立体視において高密度奥行き情報を推定するために考案されたDAISY[20]や，局所領域の明度の順序と分布を組み合わせて特徴量化したMROGH[7]等はSIFTを上回る精度を実現している部分もあり，今もなお新たな局所特徴は研究され続けている．

しかしながら，これら局所特徴の頑健性は拡大縮小や回転変化に限られており，対象平面が射影的歪みを受けて撮影されるようなカメラの視点変化が生じるときは，特徴量が変化してしまうために点対応がとれなくなってしまう場合が存在することが実験から確認できる．本研究ではこの問題を解決するために視点生成型学習を提案し，大きな視点変化への頑健性獲得を実現する．様々な視点から撮影されたかのようなパター

[*1]慶應義塾大学
[*2](株) 富士通研究所
[*1]Keio University
[*2]FUJITSU LABORATORIES LTD.

ン群を生成するこの学習法は，より多くのパターンから検出される特徴点を抽出して点対応が取れる可能性を広げ，様々なパターンで記述される特徴量を収集しデータベース化して正しいマッチングの増加を促す．それぞれの局所特徴がもつ不変性を生かすことで生成枚数を抑え，高速に学習を終えることができる．

オフラインで長い時間かけて学習したものを拡張現実感に応用する場合においては，ポスターや広告といった対象物体と，重畳表示したいアノテーションやCGオブジェクトといったデータは予め結びついていることになり，その対象物体が目の前に用意されていることを前提としている．一方で，重畳表示したいデータは手元にあって，それを目の前の任意の対象にその場で結びつけたいという需要がある．この場合，その場で対象物体を撮影し直ぐに用いるため，高速に学習を終えることができる本研究の応用が期待される．

2 関連研究

視点変化に対して頑健に位置や姿勢を推定するため，トラッキングのように連続するフレーム間の情報を用いる研究も多い．Wagnerら [21] は，検出した特徴点周りに作成した画像パッチを追跡する手法を提案した．連続するフレーム間では画像パッチの見え方が大きく変わらないことを前提にすることで，画像パッチの変形を線形に予測して次フレームでの対応を高速にかつ頑健に行い，携帯デバイスでの拡張現実感をも実現した．Itoら [8] は，追跡する対象平面がカメラの視線方向に対して大きく傾いたり遠くに離れる際に生じる平面の見えの実効的な解像度低下が，追跡性能に悪影響を与えることを指摘した．そこで画像の標本化過程をモデル化し，追跡アルゴリズムにこれを組み込むことで頑健な平面トラッキングを実現した．連続するフレーム間の情報を用いるこれらの手法では，対象を正確に追跡し続けている間は有効であるが，一度対象が画角外に出てしまったり追跡に失敗してしまったときは，再びカメラ位置を初期状態に戻してトラッキングし直さなければならない．我々の手法は，参照元の正面画像とカメラから現在入力されている画像の2枚間で点対応を求めるフレーム間独立の処理なので，動画が入力であっても上述の問題は発生せず，再初期化の手間がかからないという利点がある．

また，SIFTを発展させた手法の一つであるASIFT [17] は，アフィン変化に対する頑健性を獲得しマッチング精度を大きく向上させた．カメラが対象を中心として半球状を移動すると仮定したアフィンカメラモデルをつくり，マッチングさせたい2枚の画像それぞれに対して，その仮想視点上から撮影した画像を多数生成した．そして生成した画像群同士を多対多でSIFTを用いてマッチングさせることで，1対1の対応では得られないような点対応を大量に取得できるようになった．しかしASIFTは1枚処理を行うごとに，低速なSIFTを幾度も用いてしまうため，処理にSIFTの何倍もの時間を要してしまうのが欠点といえる（文献 [17] では2.25倍と述べられている）．

一方で，検出した特徴点周りに作成した画像パッチを決定木に学習させることで，高速かつ頑健なマッチングを実現する手法が提案されてきた．決定木のアルゴリズムにRandomized Treesを用いたもの [10]，Random Fernsを用いたもの [18]，Fernsを軽量化したPhony Ferns[21]，Randomized Treesを2段階に発展させたもの [22] などがある．特徴量同士の類似度計算をユークリッド距離で算出する手法に対し，これらはランダムに選択した画像パッチ内画素の明度の大小を比較するだけで決定木をたどって点対応が求まるため，実時間処理をも可能にする高速性を備えている．本論文で提案する手法においても，画像を様々変形させることで見えの変化に頑健な特徴点を抽出する方法を取り入れているが，大きく異なる部分が2つある．

1) 特徴点同士を対応させるためには，まず2枚の画像上で同じ特徴点が検出されなければならない．Randomized Trees等では正面画像で検出された特徴点が，生成した画像上でも検出されるかどうかを検証している．しかしこの手法では，正面画像上は特徴点として検出されないが視点が変わった時には検出されるような特徴点を扱うことができない．筆者らの提案手法ではそういった特徴点も含めて見えの変化に頑健なものを抽出することができるようになっている．

2) Randomized Trees等は特徴量ではなく画像パッチで識別を行うので，大きな視点変化に対応するために多くの枚数を変形して画像パッチを生成する必要があり，学習に数分以上を要してしまう．一方で筆者らの提案手法では，局所特徴が備える拡大縮小や回転への不変性を生かした上で，射影的歪みへの頑健性を向上させることができる．その効率的な学習は生成枚数を抑え，30秒以下で終わらせることも可能である．

3 視点生成型学習

局所特徴量の計算では拡大縮小や回転に対して不変性をもつようなアルゴリズムが組まれているが，視点が大きく変化すると記述される特徴量が変化してしまうため，特徴点同士の対応が取れなくなってしまう場合が存在する．そこで我々は様々な視点における特徴量を収集することで，拡大縮小や回転に加え，更なる視点変化への頑健性を獲得する．ここで提案手法の学習には村瀬が提唱した生成型学習 [23] の枠組みを利用する．生成型学習による画像認識は，入力パターン

の多様な変動や変形に対処するための手法として，学習パターンを変動や変形により様々に生成し，これと入力パターンとを照合する．実際のシーンをサンプルとして収集し用いるのではなく，実際に起こりうるようなシーンを仮想的に生成し用いる学習法であり，収集できるパターンが少ない場合に特に有効といえる．よって視点生成型学習では，対象となる平面の正面画像1枚を入力に用いて，様々に角度を変えて撮影したかのようなパターン群を生成して局所特徴を抽出，収集した後にクラスタリングによって少数の特徴で代表させデータベース化する．学習を終えると，正面画像と実際に撮影されたカメラからの入力画像はそれぞれ，データベース中の特徴量と抽出された特徴量を比較することで点対応を得る．

また，2章で述べた Randomized Trees 等 [10, 18, 21, 22] は学習時に変形パターンを生成するため，生成型学習の枠組みに含まれていると言える．生成型学習の困難性について村瀬は文献 [23] 内で「生成型学習では，いかに少数の学習パターンから入力パターンに含まれるような変形や変動した学習パターンを生成するかが重要な鍵である」と述べているが，提案手法の特徴量群クラスタリングは，Randomized Trees 等が行っているような全ての見え方の特徴量のヒストグラムに基づく決定木を作る方法に比べて，学習に要する時間を大幅に短縮することが可能となる．これは Randomized Trees 等が特徴点ごととはいえ数千単位の数の変形パターンを生成する必要があるのに対して，数十枚程度のパターン生成で済むことが大きく起因している．よって，Randomized Trees 等が数分以上の時間をかけて学習する必要があるのに対して，提案手法の視点生成型学習では追跡したい対象のパターンを撮影した直後数秒から数十秒でオフライン処理を終えることができ，それに基づいた追跡による拡張現実感が直ぐに実行できるという利点がある．

3.1 視点変化パターン群の生成

はじめに対象となる平面の正面画像を1枚入力し仮想的に様々な視点から見たかのようなパターン群を生成する．このパターン群はアフィン変換ではなく，射影変換で生成を行う．Randomized Trees 等のような画像パッチ内の明度値を比較して分岐を行う決定木を作りヒストグラムに基づいて判定を行う手法では，局所領域の幾何学変化に対して近似を行っても有効に機能するため，2次元的な回転しか扱えないアフィン変換でも十分であった．提案手法では抽出した特徴量をそのままデータベースとして用いるため，カメラの視点変化をより反映させることができる射影変換を用いることが重要となる．これにより学習後オンラインで実際に撮影した画像に近いものを得られ，抽出・収集される特徴量が有効に機能しやすくなる．射影変換行列 $\mathbf{P}$ はカメラの内部パラメータ $\mathbf{A}$ 及び外部パラメータである回転行列 $\mathbf{R}$ と並進ベクトル t より $\mathbf{P} = \mathbf{A}(\mathbf{R}|t)$ で算出される．内部パラメータは既知とし，変換に用いる回転行列 $\mathbf{R}$ は式1及び図1に示すように3つの角度を要素にもつ．

$$\mathbf{R} = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \quad (1)$$

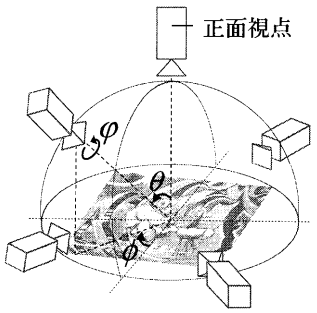


図1 視点変化パターン生成時のカメラモデル
Fig. 1 A camera model for pattern generation of various viewpoints

角度 ϕ と θ はそれぞれ半球状を移動するカメラ位置を決定する経度と緯度にあたり，ψ はカメラの光軸周りの回転角である．各角度は $\phi \in [-75°, 75°], \theta \in [-75°, 75°]$ の範囲内で値を選択するものとするが，本実験で用いる局所特徴は回転不変性を備えているため，光軸周りの ψ は $0°$ に固定し回転を行わない．並進ベクトルはカメラと対象平面との距離を d として $t = (0, 0, d)^T$ と表されるが，本実験で用いる局所特徴は拡大縮小への不変性も備えているため固定値として焦点距離を用いる．

3.2 見えの変化に頑健な特徴点の抽出

次に生成された各パターンから局所特徴の検出器により特徴点を検出する．様々なパターンから同じ位置に検出される特徴点は，見えの変化に頑健な，高い可検出性をもつ特徴点（以下 stable keypoint）であり，stable keypoint の抽出が点対応の増加につながる．パターンが正面画像から平面射影変換行列 $\mathbf{H}_i$ で生成できるとすれば，パターン上で検出された特徴点 p_i は，正面画像上の p'_i に逆行列を用いた式 $p'_i = \mathbf{H}^{-1} p_i$ によって投影される．投影点 p'_i の最近傍特徴点を探索することでその点が stable keypoint であるかを調べることができる．もしも投影点 p'_i と最近傍特徴点とのユークリッド距離がしきい値以下であれば，2枚の画像で同じ特徴点が検出されたものとみなし，その特徴点に対して投票を行う．しきい値以上だった場合は，正面画像では検出されなかったが視点変化によって新

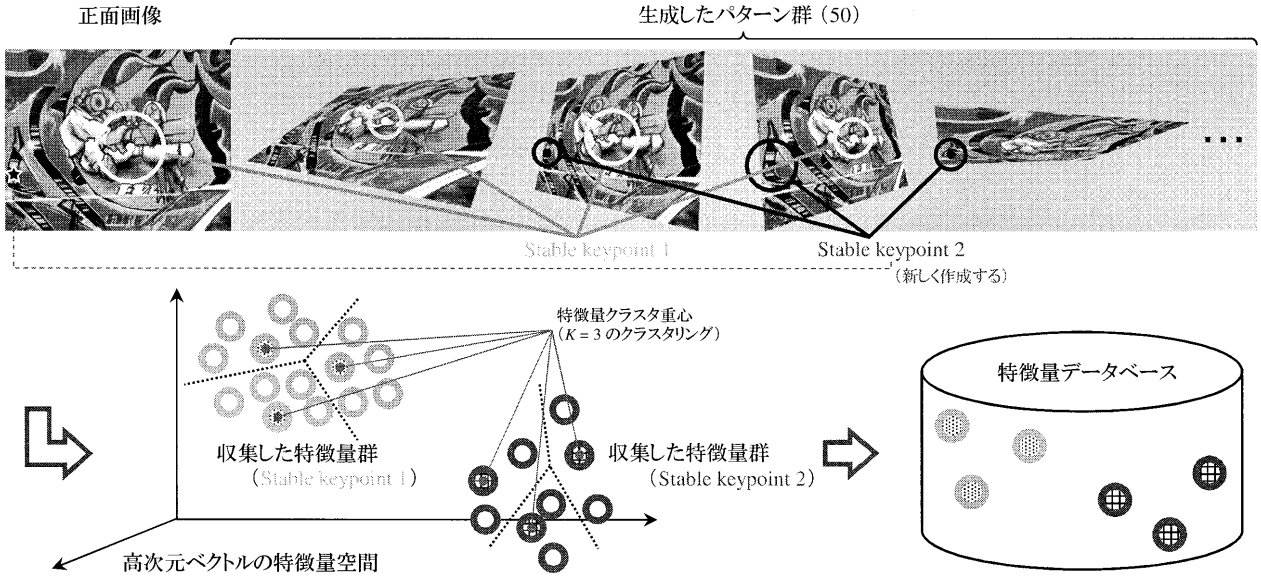


図2 生成されたパターン群から収集した特徴量をクラスタリングしデータベース化
Fig. 2 The database creation via clustering of the descriptors extracted from the generated patterns

たに検出された特徴点だと考えられ，投影点を新たな特徴点として正面画像上に作成する．

図2に示すように，stable keypoint 1の場合は正面画像上と視点変化パターン上ともに特徴点が検出されるが，stable keypoint 2の場合は視点変化パターンからは検出されるのに対し正面画像からは検出されないので，正面画像上に新しく特徴点として作成する．生成した全パターン，検出された全特徴点に対して投影や投票を終えたとき，正面画像上の特徴点は得票数を元に順位付けすることができる状態にあり，この中から上位のみをstable keypointとして抽出する．抽出する数は精度や速度に影響を及ぼすパラメータであるため，実験で評価し決定する．

ただし，生成したパターン上で同一の箇所に検出されない特徴点ばかりが検出されるような状況下ではstable keypointを決定することができない．この状況が発生する条件は，画像から検出される特徴点が限りなく少ないことが代表として挙げられるが，特徴点の数や検出箇所の広がりは検出器のパラメータだけでなく対象のテクスチャに大きく依存するため，stable keypointを決定するための制約条件を明確に定義することはできない．よって，本研究ではstable keypointが決定できる状況下で学習が行われることとする．

3.3 特徴量データベースの作成

特徴点検出と同時に特徴量を記述すれば，stable keypointの抽出を終えたとき，各stable keypointは検出されたパターンの枚数と同じ数だけ記述された特徴量を収集している．収集された全特徴量をデータベースに登録してしまうと，マッチング時における探索に膨大な時間を要してしまうため，クラスタリングによってクラスタ重心となる特徴量のみをデータベースに登録する．特徴量空間で等距離性を保つように重心となる特徴量を選択することで，注目特徴点の視点変化による特徴量変化を許容しつつ，他特徴点の特徴量を許容しにくくしている．クラスタリング手法の一つであるk-means法は，データの中からk個の重心を選択し，最近傍のクラスタに所属するデータ量が均等になるまで重心を繰り返し選択し直していく．我々がクラスタリングに用いるk-means++[3]は標準のk-means法から収束速度と精度を共に向上させた手法である．完全にランダムな初期重心を選ぶため局所解に陥ってしまいがちなk-means法の弱点を克服するために初期配置を考慮した手法で，既に置かれた重心の近くには別の初期重心が配置される確率を減らし，なるべく等距離上に重心を配置できるように改良されたものである．k-means++は高速に収束するため，クラスタリングにかかる時間が学習時間に影響を与えないことも利点に挙げられる．本論文で行った全ての実験で反復回数の上限を10回に設定しているが，10回未満で収束することを確認している．

図2にはstable keypointが抽出されてから特徴量のクラスタ重心がデータベースに登録されるまでの流れを示した．クラスタリングは各stable keypointが収集した特徴量ごとに行うため，クラスタ重心はstable keypointとk対1で完全に紐づいており，マッチング時の識別を容易にしている．stable keypointの数をN，設定したクラスタ数をKとすればデータベース内には$N \times K$個の特徴量が登録されることになる．

4 マッチングと位置姿勢推定

学習を終えると，実際に撮影された入力画像から対象平面を検出し位置姿勢を推定する．まず入力画像から特徴点を検出し，特徴量を記述する．次に入力画像から得た特徴量をデータベース内の特徴量と比較し，最もユークリッド距離が近い組を対応させるが，最近傍だけでなく 2 番目に近い特徴量についても探索を行っておく．これは Mikolajczyk ら [14] の最近傍比を用いて誤対応を除く手法で，2 番目に距離が近い特徴量と比べて最近傍の特徴量がより近接している場合を正対応とする．特徴量 D_A に対し最近傍の特徴量 D_B と 2 番目の特徴量 D_C が以下の関係式を満たすとき，正対応とみなしてマッチングが行われるのである．

$$\frac{\| D_A - D_B \|}{\| D_A - D_C \|} < t \tag{2}$$

しきい値 t を大きくすれば対応数は増えるが，誤対応も増加する．逆に小さくすれば対応数は減るが，誤対応も減少するという関係にあり，本論文中で行った全実験では経験的に 0.6 と決定した．(2) の条件式よりデータベース内の特徴量に対応すれば，正面画像上の stable keypoint と一意に対応がとれ，マッチングが完了する．このようにして入力画像と正面画像の間で点対応が求まった後，ロバスト推定法の RANSAC で更に誤対応を除去した上で，誤差を最小化するレーベンバーグ・マルカート法を用いて平面射影変換行列を算出することにより，位置姿勢を推定することができる．

5 評価実験

ここでは提案手法の性能を評価するために行った実験について示す．Mikolajczyk らは幾何学または光学的変化を施した様々な画像を対象とした実験から，局所特徴によるマッチングの精度を求めた [16]．彼らはその実験に用いたデータセットを提供しており，我々はその内の視点変化シーケンスである Graffiti と Wall の 2 種類のデータセットを使って各種実験を行った．各データセットは正面画像 1 枚とテスト画像 5 枚の計 6 枚で構成されており，テスト画像は 1 番から 5 番まで順に，20°, 30°, 40°, 50°, 60° の角度だけ正面から傾けて撮影されている．また，各テスト画像への真値の平面射影変換行列が与えられており，それを用いて精度の評価を行う．一般的に，特徴点および特徴領域を真値の平面射影変換行列で投影したものを対応したもう片方のものと比較し，その距離や領域の重なりを元に評価を行うが，提案手法では 1 つの stable keypoint が複数の特徴量を所持しているようなデータベースを作成しているため，マッチングの精度のみ評価する．本研究は局所特徴自体ではなく，それを用いた位置姿勢推定の頑健性を高めることが目的であるため，推定時に必要とするマッチングが評価されることの意義は大きい．評価の基準とするマッチング精度を次式に示す．

$$\text{マッチング精度} = \frac{\text{正しいマッチングの数}}{\text{全マッチングの数}} \tag{3}$$

マッチングがとれた特徴点のうち正面画像上の点を m_A，テスト画像上の点を m_B とする．真値の平面射影変換行列 $\mathbf{H}$ を用いれば m_A はテスト画像上の位置 m'_A に投影される: $m'_A = \mathbf{H}m_A$．ここで m'_A と m_B の距離 d が近ければ同一の特徴点がマッチングされているとわかり，そのしきい値は τ とする: $d(m'_A, m_B) < \tau$．本論文の全ての実験において $\tau = 2$ とし，上式を満たしたマッチングを正しいマッチングとする．

更に我々は推定された平面の位置姿勢をコーナーの再投影誤差によって評価する手法 [12] を用いる．マッチング精度は点対応の正確性を測る上で重要な評価であるが，位置姿勢推定はその精度の高さだけでは正確とは言えない．例えば，正しいマッチングの数が多くても，一部の領域に偏っていた場合は平面全体の推定結果は大きく誤ってしまうことがある．この手法では，正面画像の 4 隅の点がテスト画像上ではどの位置にあるかを推定した結果で評価が行われる．真値の平面射影変換行列で投影したコーナーを p_j，マッチングから算出された平面射影変換行列で投影したコーナーを q_j とすれば，再投影誤差 err は次式で表される．

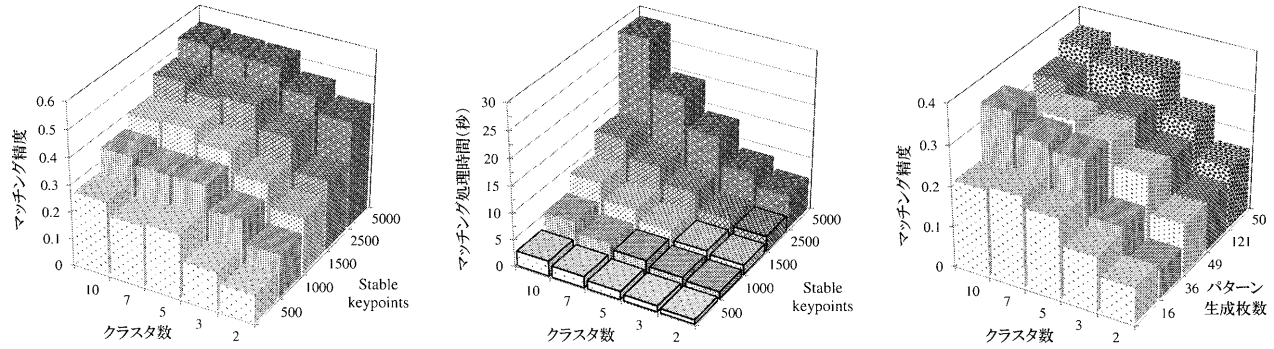
$$err = \sqrt{\frac{1}{4}\sum_{j=1}^{4} ||p_j - q_j||^2} \tag{4}$$

err が 10 画素を超えるとき推定は失敗したものとして扱い，10 画素以下であればその数値が低いほど正確に位置姿勢が推定されたことを示す．

これら 2 つの評価を用いて，はじめに視点生成型学習で用いるパラメータが速度や精度にどのように影響するかを確認する．次に様々な局所特徴を視点生成型学習に適用させ，アルゴリズムによらずその精度を向上させたことを示す．最後に ASIFT や Randomized Trees 等との比較を行うことで提案手法の有用性を確認する．また本実験では，CPU : Intel Core i3 3.07 GHz, GPU : GeForce 310 589MHz, メモリ: 2.92 GB の環境下で処理時間を測定した．

5.1 最適なパラメータの選定

視点生成型学習では大きく 3 つのパラメータを扱う．1 つ目は生成するパターンの枚数であり，これを W とする．3.1 章で述べたように回転角の範囲は $\phi, \theta \in [-75°, 75°]$ と定めた．ここで回転角のインターバルを $a°$ とすれば，$a°$ 刻みで回転角を決定することでパター



(a) stable keypoint 数とクラスタ数の変化によるマッチング精度の変化 (b) stable keypoint 数とクラスタ数の変化によるマッチング処理時間の変化 (c) クラスタ数とパターン生成枚数の変化によるマッチング精度の変化

図 3 視点生成型学習に用いる最適なパラメータの選択
Fig. 3 The choice of most effective parameter for viewpoint generative learning

ンを生成できる．生成する枚数は $W = (\frac{150}{a} + 1)^2$ と表され，表 1 のような組み合わせをとる．

2 つ目は stable keypoint を抽出する数 N である．対象平面のテクスチャや用いる局所特徴によって，検出できる特徴点数は大きく変わるため，本実験では Graffiti のデータセットから SIFT[13] を検出したときの結果で評価を行う．SIFT は Graffiti の正面画像から 5000 点の特徴点を検出するので，N を 500 点から 5000 点の間で変化させた．

3 つ目は k-means++[3] でのクラスタ数 K である．評価はマッチング精度，マッチング処理時間，学習時間の項目で行うが，マッチング精度では視点変化が大きいときの精度をより重視するために，1 番と 2 番の画像は使わず，以降の 3 枚に対してマッチング精度を測定し，重み付け平均で算出したマッチング精度で評価する．このとき重みとして撮影角の sin 値（$\sin 40°$, $\sin 50°$, $\sin 60°$）を用いる．マッチング処理時間は，データベースを探索しはじめてから，平面射影変換行列を算出し終えるまでの時間とし，特徴点検出や特徴量記述の処理時間は含めない．学習時間は正面画像を入力してからデータベースを作成し終えるまでの時間を計測した．

まず，stable keypoint の数 N とクラスタ数 K との関係性を図 3(a) と図 3(b) から考察する．N が増加すれば，マッチング精度の向上に伴ってマッチング処理時間も増大している．K の増加も同様の結果を招いているが，K が 5 以上の場合においては処理時間の増大に対して精度がほとんど向上していないことがわかる．また，この 2 変数のみを変更した場合は学習時間に変化はなく一定であった．

一方で，表 1 が示す通り，生成するパターンの枚数 W は学習時間に大きく影響する．ただし，図 3(c) によれば 36 枚以上のときは W が増加しても精度の向上が見られない．よって本実験においては 30 秒で学習を終える枚数が精度と速度のトレードオフ点となることが確認された．表 1 の‘V’は $W = 50$ でマッチング精度を最大にするようにパターン生成時の撮影角度を手動で指定したものであり，ϕ と θ は $\pm 75°, \pm 60°, \pm 15°, 0°$ の角度をとる．ここで決定したパターン生成角度は以下の実験でも同様のものを使用する．

表 1 パターン生成枚数の変化によるマッチング精度及び学習時間の変化
Table 1 The precision and learning time by changing the number of patterns

インターバル角度	50°	30°	25°	15°	V
パターン生成枚数	16	36	49	121	50
マッチング精度	0.17	0.23	0.24	0.24	0.28
学習時間 (秒)	9.9	21.5	29.4	77.7	29.9

3.3 章で述べたように，データベース内に登録された特徴量の数は $N \times K$ である．$N \times K$ が 5000 以下になるように N と K が設定されているとき，図 3(b) 中では枠線で囲まれている部分に該当するが，2 枚の画像を単純に SIFT でマッチングさせる手法と同じもしくはそれ以上に高速な処理時間となる．よって処理時間を低下させずに精度を向上させるパラメータは，クラスタ数 K を 5 とし，パターン生成枚数 W は 50，stable keypoint の数 N は正面画像で検出される特徴点数の 1/5 が最適であるとし，以下の実験に用いる．

5.2 様々な局所特徴量への適用

視点生成型学習は収集した特徴量がクラスタリングできれば良いので，高次元ベクトルで記述された特徴量であればアルゴリズムによらず適用可能である．よって次の実験では様々な局所特徴を用いて学習を行ったときの性能を評価する．各局所特徴の実装元は以下の通りである．SIFT[13] は SiftGPU[1]．SURF[5] と CenSurE (STAR)[1] は OpenCV[2]．CenSurE と組み合わせ

[1] http://www.cs.unc.edu/~ccwu/siftgpu/
[2] http://opencv.willowgarage.com/

て用いる特徴量記述の M-SURF (Modified-SURF) は OpenSURF[3]．Harris/Hessian - Laplacian[4] 並びに Harris/Hessian - Affine[15] の両特徴点検出器は，特徴量として GLOH[16] と MROGH[7] を用いる．検出器と GLOH はデータセットと同じ提供元[4]で，MROGH は著者の Fan ら自身が公開しているもの[5]を利用した．本実験では，2 枚の画像から単純に局所特徴を用いて点対応を取得し位置姿勢推定を行う手法と提案する学習を取り入れた手法で比較を行う．このとき，精度評価の指標として用いるマッチング精度（式 3）は検出される特徴点の数に依存するため，局所特徴の抽出パラメータに共通の値を用いることで同じ特徴点数を検出し公平性を図る．ところが，提案手法ではマッチング時の処理時間増大を防ぐため，stable keypoint の決定時に正面画像上の特徴点を 1/5 に絞り込んでしまう．よって，マッチング精度だけではなく正しいマッチングの数も同時に評価することで，点数の違いによる不公平性を避けた．

図 4 中で，左側は棒グラフに正しいマッチングの数を，折れ線グラフにマッチング精度を示し，右側は再投影誤差を示している．つまり左側では数値が高いほど精度が高く，右側では数値が低いほど推定結果が正しい．視点生成型学習を取り入れたものは Viewpoint Generative Learning の頭文字を取って VGL と記載しているが，多くの場合において VGL は単純なマッチングを上回る結果となっている．

単純に局所特徴を用いて対応付けた方は，4 番や 5 番のテスト画像に対してはそもそも再投影誤差が出力されていない（グラフが途切れている）ことがわかる．これは 5 章の定義式 4 に基づいて再投影誤差を算出した結果 10 ピクセル以上となったため，位置姿勢推定が失敗しているとみなされている．一方，VGL では，SURF を除き 4 番並びに 5 番のテスト画像に対しても再投影誤差は 10 ピクセル以下に収まり．位置姿勢推定を行うことができている．よって VGL はマッチングの精度を向上させただけでなく頑健な位置姿勢推定を実現していることがわかる．

5.3 ASIFT 及び Random Ferns との比較

図 5，図 6，表 2 は ASIFT 及び Random Ferns で画像パッチを学習する手法との比較結果を示す．ASIFT と Ferns は視点変化への頑健性をある程度備えているため，撮影角度が 80° を超える Adam と Magazine を実験対象に加えた．このデータセットと ASIFT は著者の Morel らが提供しているもの[6]，Ferns は OpenCV に実装されているものを利用した．また，Ferns は学習時間とマッチング精度が生成する画像枚数に依存しているため，Graffiti の 4 番目のテスト画像で再投影誤差が 10 画素以下に推定可能な範囲となる枚数を検証し，最小の 2500 枚を FernsS，最大の 5000 枚を FernsL として比較対象とした．

ASIFT の結果ではあらゆるデータセットにおいて多量の正しいマッチングを得ているが，比率であるマッチング精度や再投影誤差をみると，SIFT を提案手法 (VGL) に利用した SIFT(VGL) が上回っているものも確認できる．このとき Graffiti データセットにおいて，ASIFT は 1 枚あたり 47 秒も処理時間がかかっているのに対し，SIFT(VGL) は 4 秒であった．文献 [17] では ASIFT の処理時間は SIFT の 2.25 倍であると述べていたが，本実験環境では 10 倍以上の差がついていることがわかった．よって動画が入力となるような場合では ASIFT は実用には向かず，精度と処理時間の両方の面で提案手法 (VGL) が有効であると言える．

一方，Ferns は処理時間が 1 秒未満であり実時間処理が行えるものの，FernsS では 7 分，FernsL では 10 分半も学習に時間を要する上に，5 番目のテスト画像や Adam，Magazine といった大きく視点が変わる画像に対してはマッチング精度も低い．提案手法 (VGL) は 30 秒以下で学習をすませ，そうした画像に対しても位置姿勢推定を行える程度のマッチング精度がある．画像パッチを決定木に学習させる手法と比較しても，提案手法 (VGL) は有効な学習法であると言える．

提案手法を拡張現実感環境へ応用する際は，画像サイズを小さくしたり，特徴点数を減らしたりするなど，最も処理時間を要する局所特徴抽出部分の計算量を減らすことで高速処理を実現する．Graffiti では画像サイズ 800 × 640 ピクセルで約 5000 点を検出していたため 4 秒の処理時間を要したが，画像を 320 × 256 ピクセルにリサイズし約 500 点を検出するように変更した場合は約 0.01 秒で処理が可能となる．

6 おわりに

我々は視点変化に対して頑健な平面の位置姿勢推定法を提案した．様々な視点から撮影したかのようなパターン群を生成し，収集した局所特徴量をクラスタリングすることで，事前に特徴点が違った見え方をする際に記述される特徴量の違いを学習することができた．この視点生成型学習は，より可検出性の高い特徴点を抽出し，各局所特徴の不変性を生かした上で射影歪みへの頑健性を獲得し，30 秒以内に完了する高速処理を可能にした．学習のフレームワークは高次元ベクトルで記述する特徴量記述においては全てに適用可能であり，将来的に新たな局所特徴が誕生したとき，その

[3] http://www.chrisevansdev.com/
[4] http://www.robots.ox.ac.uk/~vgg/research/affine/
[5] http://vision.ia.ac.cn/Students/bfan/mrogh.htm
[6] http://www.cmap.polytechnique.fr/~yu/research/ASIFT/

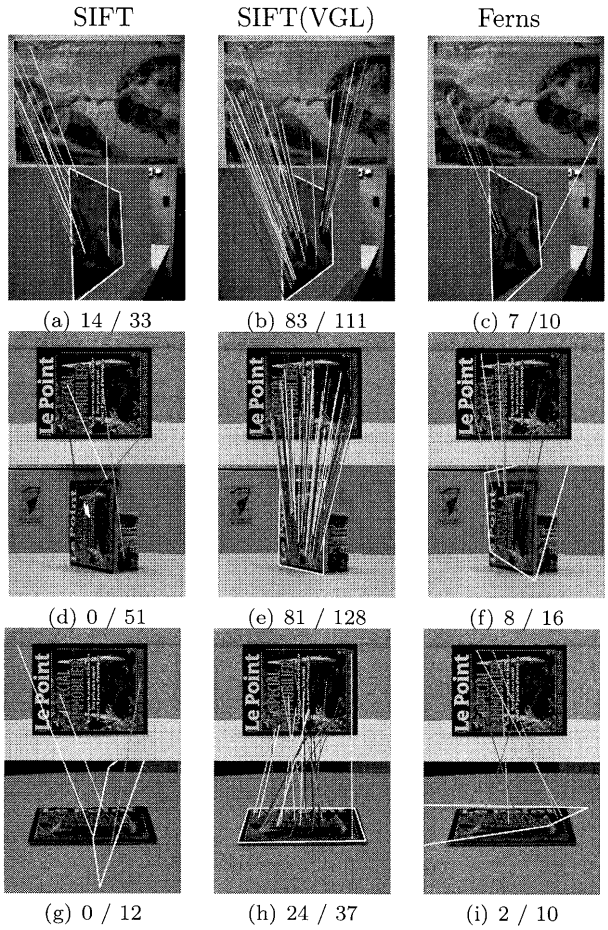


図 5 Adam, Magazine のマッチング比較
Fig. 5 The comparison of matching on the dataset of Adam and Magazine

表 2 マッチング精度の比較
Table 2 The comparison of precision with ASIFT and Ferns

	SIFT(VGL)	ASIFT	FernsS	FernsL
G1	562/801 (0.70)	2388/3561 (0.67)	440/671 (0.66)	190/333 (0.57)
G2	435/709 (0.61)	1085/2459 (0.44)	176/357 (0.49)	119/216 (0.55)
G3	239/494 (0.48)	904/1711 (0.53)	24/42 (0.57)	48/109 (0.44)
G4	96/274 (0.35)	465/953 (0.49)	6/12 (0.50)	16/32 (0.50)
G5	53/252 (0.21)	249/504 (0.49)	0/8 (0)	0/9 (0)
W1	1937/2434 (0.80)	8329/11108 (0.75)	607/685 (0.89)	726/839 (0.87)
W2	1842/1953 (0.94)	6266/7059 (0.89)	482/492 (0.98)	674/693 (0.97)
W3	1023/1378 (0.74)	2215/3717 (0.60)	177/249 (0.71)	294/388 (0.76)
W4	449/687 (0.65)	981/1978 (0.50)	58/96 (0.60)	120/189 (0.63)
W5	109/199 (0.55)	279/696 (0.40)	0/8 (0)	0/7 (0)
A	83/111 (0.75)	80/149 (0.54)	1/10 (0.10)	7/10 (0.70)
M1	81/128 (0.63)	245/437 (0.56)	2/11 (0.18)	8/16 (0.50)
M2	24/37 (0.65)	125/235 (0.53)	0/13 (0)	2/10 (0.20)

弱点を補うためのパターンを生成することで，あらゆる視点変化への対応を取り入れることができると考える．また，提案手法は並列処理や GPGPU との親和性が比較的高いことが考えられ，今後こうした演算を取り入れていくことで更なる高速化が期待できる．

近年は，より高速な特徴量記述とマッチングを達成すべく，バイナリ値で特徴量を持ち，ハミング距離で対応計算を行う局所特徴も研究されてきている．BRIEF [6]，ORB [19]，BRISK [11]，CARD [2] などが該当するが，これらは高速化を重視した分，従来の局所特徴に比べて，変化に対して脆弱な部分が見られる．今後はこうしたバイナリ特徴量へも視点生成型学習が適用できるかを検討していきたい．

参考文献

[1] Motilal Agrawal, Kurt Konolige, and Morten Rufus Blas. Censure: Center surround extremas for realtime feature detection and matching. *ECCV*, Vol. 5305, pp. 102–115, 2008.
[2] Mitsuru Ambai and Yuichi Yoshida. Card: Compact and real-time descriptors. *ICCV*, 2011.
[3] D. Arthur and S. Vassilvitskii. k-means++: The advantages of careful seeding. *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 1027–1035, 2007.
[4] Adam Baumberg. Reliable feature matching across widely separated views. *CVPR*, pp. 774–781, 2000.
[5] Herbert Bay, Tinne Tuytelaars, Van Gool, and L. Surf: Speeded up robust features. *ECCV*, Vol. 3951, pp. 404–417, 2006.
[6] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. Brief: Binary robust independent elementary features. *ECCV*, Vol. 6314, pp. 778–792, 2010.
[7] Bin Fan, Fuchao Wu, and Zhanyi Hu. Aggregating gradient distributions into intensity orders: A novel local image descriptor. *CVPR*, pp. 2377–2384, 2011.
[8] Eisuke Ito, Takayuki Okatani, and Koichiro Deguchi. Accurate and robust planar tracking based on a model of image sampling and reconstruction process. *ISMAR*, 2011.
[9] Yan Ke and Rahul Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. *CVPR*, 2004.
[10] V. Lepetit and P. Fua. Keypoint recognition using randomized trees. *TPAMI*, Vol. 28, No. 9, pp. 1465–1479, 2006.
[11] Stefan Leutenegger, Margarita Chli, and Roland Y Siegwart. Brisk: Binary robust invariant scalable keypoints. *ICCV*, 2011.
[12] Sebastian Lieberknecht, Selim Benhimane, Peter Meier, and Nassir Navab. A dataset and evaluation methodology for template-based tracking algorithms. *ISMAR*, pp. 145–151, 2009.

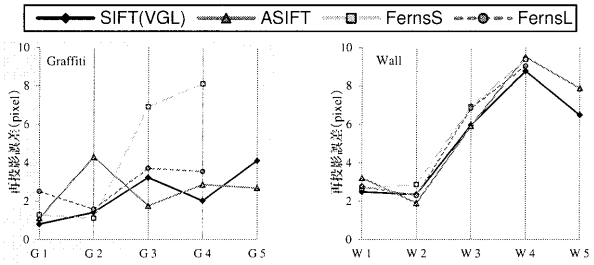


図 6　再投影誤差の比較
Fig. 6　The comparison of reprojection error

[13] David G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, Vol. 60, pp. 91–110, 2004.
[14] K. Mikolajczyk, T. Tuytelaars, Cordelia Schmid, Andrew Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *IJCV*, Vol. 65, pp. 43–72, 2005.
[15] Krystian Mikolajczyk and Cordelia Schmid. Scale & affine invariant interest point detectors. *IJCV*, Vol. 60, pp. 63–86, 2004.
[16] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *TPAMI*, Vol. 27, pp. 1615–1630, 2005.
[17] J. M. Morel and G. Yu. Asift: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, Vol. 2, No. 2, pp. 438–469, 2009.
[18] M. Ozuysal, M. Calonder, V Lepetit, and P Fua. Fast keypoint recognition using random ferns. *TPAMI*, Vol. 32, No. 3, pp. 448–461, 2009.
[19] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: an efficient alternative to sift or surf. *ICCV*, 2011.
[20] Engin Tola, Vincent Lepetit, and Pascal Fua. Daisy: an efficient dense descriptor applied to wide-baseline stereo. *TPAMI*, Vol. 32, pp. 815–830, 2010.
[21] Daniel Wagner, Gerhard Reitmayr, Alessandro Mulloni, Tom Drummond, and Dieter Schmalstieg. Real-time detection and tracking for augmented reality on mobile phones. *TVCG*, Vol. 16, pp. 355–368, 2010.
[22] 西村孝, 清水彰一, 藤吉弘亘. 2 段階の randomized trees を用いたキーポイントの分類. *MIRU*, 2010.
[23] 村瀬洋. 画像認識のための生成型学習. 情報処理学会論文誌 (CVIM), 2005.

(2012 年 2 月 27 日受付)

[著者紹介]

吉田 拓洋　(学生会員)

2011 年慶應義塾大学理工学部情報工学科卒業. 現在, 同大学大学院理工学研究科修士課程に在学中. コンピュータビジョン, 拡張現実感に関する研究に従事.

斎藤 英雄　(正会員)

1987 年慶應義塾大学理工学部電気工学科卒業. 1992 年同大学院理工学研究科博士課程電気工学専攻修了. 同年同大学理工学部助手, 2006 年より同大学理工学部情報工学科教授. 1997 年から 99 年までカーネギーメロン大学ロボティクス研究所訪問研究員. コンピュータビジョンや, それを仮想現実感や 3 次元映像メディア処理等に応用する研究に従事.

清水 雅芳

1988 年慶應義塾大学理工学部電気工学科卒業. 1990 年同大学院理工学研究科修士課程電気工学専攻修了. 同年 (株) 富士通研究所入社. コンピュータビジョン, 高画質化などの画像処理に関する研究に従事. 現在, 同社メディア処理システム研究所イメージコンピューティング研究部部長. 1998 年～1999 年米国ロチェスター工科大学客員研究員.

田口 哲典

2000 年電気通信大電気通信学部電子工学科卒業. 2005 年東京大学大学院工学系研究科先端学際工学専攻博士課程修了. 博士 (工学). 現在, (株) 富士通研究所に勤務. 主にコンピュータビジョンに関する研究に従事.

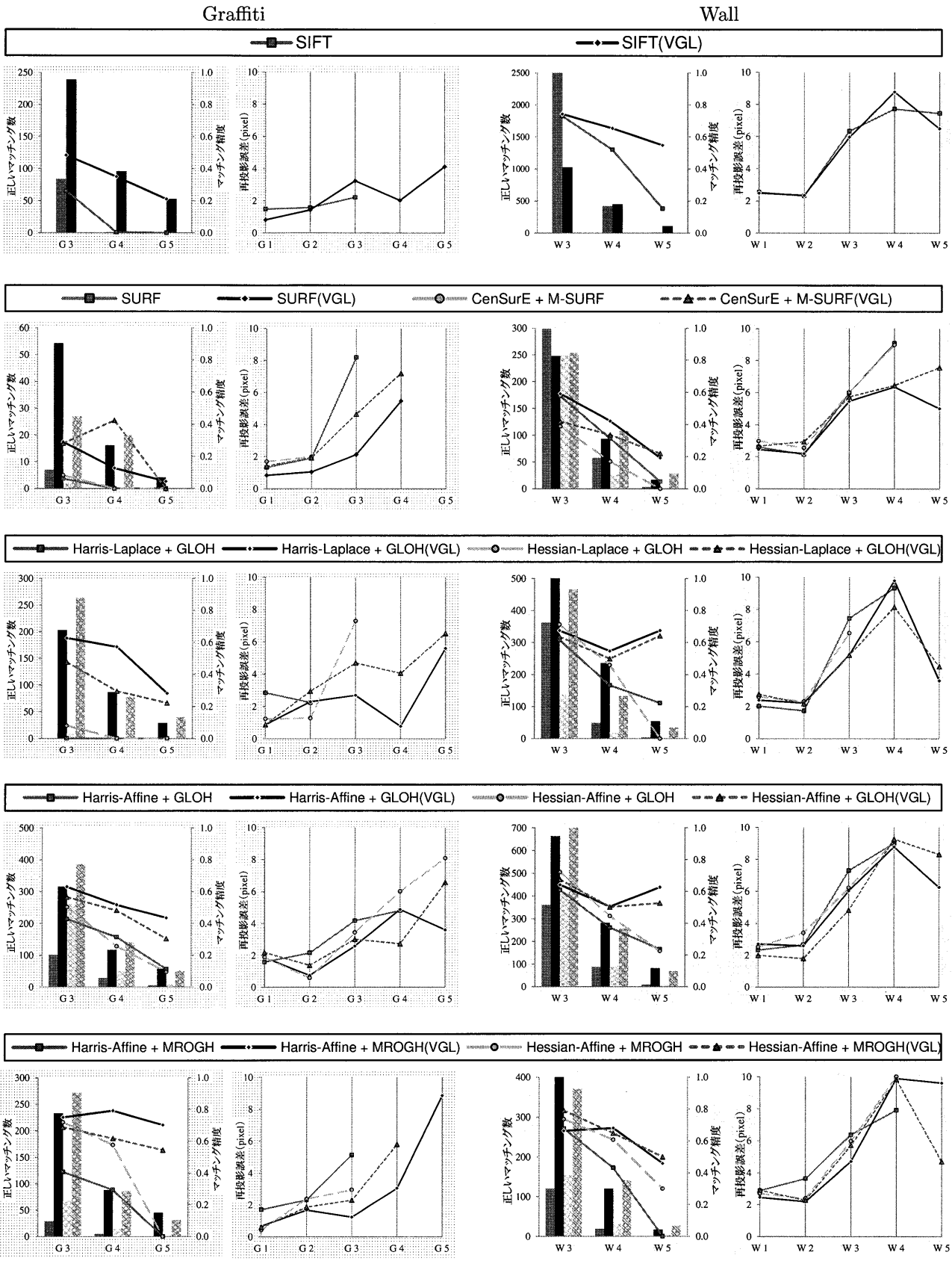


図 4 様々な局所特徴量を用いた精度比較
Fig. 4 The comparison of precision by using some local features