高精度な画像マッチング手法の検討

A Study of a High-Accuracy Image Matching Method

伊藤 康一　　高橋 徹　　青木 孝文
東北大学　大学院情報科学研究科

Koichi ITO　　Toru TAKAHASHI　　Takafumi AOKI
Graduate School of Information Sciences, Tohoku University

アブストラクト　本論文では，画像間を密かつ高精度にマッチングするための手法を検討する．画像マッチングは，画像処理・コンピュータビジョン・パターン認識などの分野において，重要な基本処理となっている．近年は，特に，特徴ベースマッチングの研究が盛んに行われており，SIFT (Scale-Invariant Feature Transform) が提案されて以来，さまざまなマッチング手法が研究されている．本論文では，特徴ベースマッチングの特長である画像変形にロバストであることと，領域ベースマッチングの特長である密にマッチングできることを活かしたマッチング手法を提案する．具体的には，特徴ベースマッチングを用いて画像間の大きな変形を補正し，領域ベースマッチングを用いて密に対応付ける．一般に公開されている標準画像を用いて，これまでに提案されている特徴ベースマッチングと性能を比較し，提案手法の有効性を実証する．

1　はじめに

近年，画像処理，コンピュータビジョン，パターン認識などの幅広い分野において，画像マッチング，特に複数の画像間の高精度な対応付けは，重要な基本処理として数多くの研究がなされている [1], [2]．画像マッチングは，特徴ベースマッチングと領域ベースマッチングの 2 つに大別される．コンピュータビジョンやパターン認識の分野では，画像認識や多視点ステレオなどの応用において特徴ベースマッチングがよく使われている．一方で，画像処理などの分野では，映像の動き推定や幾何補正などの応用において領域ベースマッチングがよく使われている．

特徴ベースマッチングは，画像からコーナーなどの特徴点を検出し，その周囲の局所領域に対して局所記述子（特徴量）を定義し，局所記述子の距離に基づいて画像間のマッチングを行う．特徴抽出として，Harris point，DoG (Difference of Gaussian) region，Harris-Affine region，Hessian-Affine region，MSER (Maximally Stable Extermal Regions) などがあり，局所記述子として，SIFT (Scale-Invariant Feature Transform) [3] や GLOH (Gradient Location-Orientation Histgram) [4] が提案されている．また，SIFT を改良したマッチング手法として，PCA (Principal Component Analysis)-SIFT，SURF (Supeeded-Up Robust Features) [5]，ASIFT (Affine-SIFT) [6] などが提案されている．これらの多くは，画像変形にロバストなマッチング手法である．特徴抽出に時間はかかるが，マッチングに時間はかからないため，大量の画像をマッチングするような応用に適している．ただし，特徴を抽出できなかったり，対応づけられない領域が生じるため，対応付けの結果が疎になってしまうことが問題である．

領域ベースマッチングは，基準点を中心とした局所ブロック画像と，入力画像の局所ブロック画像を相違度あるいは類似度の尺度を用いてマッチングを行う．相違度として SAD (Sum of Absolute Differences) や SSD (Sum of Squared Differences)，類似度として NCC (Normalized Cross Correlation) や POC (Phase-Only Correlation) が用いられている．マッチングする際には，入力画像を全探索するのではなく，階層探索することで，大幅に高速化することができる．指定した基準点に対する対応点を探索することができるため，密な基準点を設定することで，画像全体に分布する密な対応点を得ることが可能である．特徴ベースマッチングと比較して計算時間が多いことと，大きな画像変形に対応できないことが問題である．

本論文では，これまでに提案されている特徴ベースマッチングおよび領域ベースマッチングを組み合わせることで，高精度かつ密な対応付けが可能な画像マッチング手法を検討する．具体的には，特徴ベースマッチング（たとえば SIFT など）を用いて画像間のアフィン変形あるいは射影変形を取り除き，領域ベースマッチング（たとえば POC など）を用いて密な対応点を得るような手法を検討する．一般に公開されている標準画像を用いて，これまでに提案されている特徴ベースマッチングと提案手法の性能を比較し，提案手法の有効性を実証する．

2 特徴ベースマッチング

これまでに提案されている特徴ベースマッチングについて特徴抽出 (feature detection) と局所記述子 (local desctiptor) に分けて概説する．

2.1 特徴検出

特徴抽出は，画像中から濃淡値の変化が大きい点や領域を抽出する処理である．現在までに提案されている特徴抽出手法は，edge detector, corner detector, blob detector, region detector の 4 つに分類される [7]．本論文では，特に画像変形にロバストな特徴抽出として知られている Harris-Affine region, Hessian-Affine region, Difference of Gaussians (DoG) region, Maximally Stable Extremal Regions (MSER) の 4 つを用いる．以下では，これらの特徴手法について概説する．

Harris-Affine region [8]

Harris-Affine region は，Harris-Laplace detector で特徴点の位置と拡大縮小率を推定し，2 次モーメント行列に基づく affine adaptation [8] を用いることで，アフィン変形に不変な領域を抽出する．Harris-Affine region で検出される特徴は領域であるが，Harris corner detector を利用しているため，corner detector に分類される．

Hessian-Affine region [8]

Hessian-Affine region は，Hessian-Laplace detector で特徴点の位置と拡大縮小率を推定し，affine adaptation を用いてアフィン変形に不変な領域を抽出する．Harris-Affine region は 2 次モーメント行列を用いて特徴を抽出しているのに対し，Hessian-Affine region は，Hessian 行列を用いて特徴を抽出している．Hessian-Affine region は，blob detection に用いられていることより，blob detector に分類される．

DoG region [3]

DoG は，分散の大きいガウシアンから小さいガウシアンを引くことによって Mexican Hat Wavelet を近似するウェーブレット母関数である．DoG を画像に適用し，その結果の極値を求めることで，拡大縮小に不変な特徴を検出することが可能である．DoG region は，SIFT で用いられている特徴検出であり，blob detection に用いられていることより blob detector に分類される．

MSER [9]

MSER は，周囲よりも明るいもしくは暗い領域であり，かつ，領域を決定する閾値を変化させても安定して抽出される領域である．検出された MSER を楕円領域とすることで，アフィン変形に不変な領域として検出することができる．MSER は，region detector に分類される．

図 1 は，それぞれの検出器を用いて特徴領域を検出した例である．

2.2 局所記述子

局所記述子は，検出した特徴点（または領域）に対して，たとえば，アフィン変形，明るさの変化，ノイズなどにロバストな特徴量（特徴ベクトル）として定義される．本論文では，さまざまな局所記述子のうち，SIFT および GLOH について概説する．また，SIFT を改良した手法である PCA-SIFT, SURF, ASIFT についても概説する．

SIFT [3]

1999 年に Lowe が提案して以来，画像処理やコンピュータビジョンなどの幅広い分野で用いられているマッチング手法である．特徴抽出に DoG を用いることで拡大縮小に不変な特徴点を抽出し，輝度勾配のヒストグラムを用いることで特徴点近傍の回転を求める．そして，特徴点の周辺を 4×4 のブロックに分割し，ブロックごとに 8 方向の勾配ヒストグラムを求め，128 次元の特徴ベクトルとする．

GLOH [4]

SIFT が矩形領域に対する特徴ベクトルであるのに対し，GLOH は，特徴点を中心とする対数極座標系に対して特徴ベクトルを定義することで，SIFT よりも識別性能を向上させた記述子である．半径方向を 3 つに，角度方向を 8 つに分割した領域に対して特徴ベクトルを求める．ただし，特徴点に最も近い領域については，角度方向に分割しないため，合計で 17 の領域に分割する．それぞれの領域について，輝度勾配を求め，それらを 16 方向に分割し，272 次元の特徴量を求める．求めた特徴量に対して主成分分析を適用することで，128 次元の特徴ベクトルとする．

PCA-SIFT [10]

PCA-SIFT は，SIFT で検出した特徴ベクトルに対して主成分分析を適用することで，識別性能を向上させている．39×39 の矩形領域について，水平・垂直方向の輝度勾配を求め，3,042 次元の特徴量とする．求めた特徴量に対して主成分分析を適用することで，36 次元の特徴量とする．ここで，主成分分析に用いる射影行列は，あらかじめ学習画像を用いて算出する必要がある．

SURF [5]

SURF は，SIFT よりも高速な特徴ベースマッチングとして知られている．SIFT との違いは，Hessian 行列と Integral image を用いることで高速な特徴検出を，Haar wavelet を用いることで高速な特徴量抽出を実現している点である．SURF は，単なる高速化手法ではなく，SIFT よりも識別性能が高いと言われている．

ASIFT [6]

ASIFT は，アフィン変形に対するロバスト性をを向上

図 1: 特徴領域を抽出した例

させた画像マッチングである．ASIFT は，さまざまな視点から撮影した画像を生成するために，画像をアフィン変形させ，それらから SIFT 特徴量を抽出し，マッチングし，最も対応づけられた特徴点を出力する．このままでは，計算量が膨大となるため，実際は，低解像度画像でアフィン変形パラメータを推定し，推定したパラメータを用いて高解像度画像で再度マッチングをする．同様な考えを用いた高速なマッチング手法として，文献 [11] がある．

3 領域ベースマッチング

これまでに提案されている領域ベースマッチングについて相違度および類似度の指標と探索手法に分けて概説する．

3.1 相違度および類似度

SAD

SAD は，画像間の相違度を調べる手法であり，次式で定義される．

$$R_{SAD} = \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} |I(i,j) - T(i,j)| \qquad (1)$$

ここで，テンプレートの大きさを $N_1 \times N_2$，テンプレートを $T(i,j)$，対象画像を $I(i,j)$ とする．実際には，探索ウィンドウに対してテンプレートを動かしながらSAD を計算し，SAD の値が最も小さくなった位置を調べる．SAD に対して，等角直線フィッティングを適用することで，サブピクセルレベルのマッチングが可能である [12], [13].

SSD

SSD は，SAD と同様に画像間の相違度を調べる手法であり，次式で定義される．

$$R_{SSD} = \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} (I(i,j) - T(i,j))^2 \qquad (2)$$

SAD と同様にテンプレートを動かしながら最もマッチングする位置を探索する．SSD に対して，パラボラフィッティングを適用することで，サブピクセルレベルのマッチングが可能である [12], [13].

NCC

NCC は，画像間の類似度を調べる手法であり，次式で定義される．

$$R_{NCC} = \frac{\sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} I(i,j)T(i,j)}{\sqrt{\sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} I(i,j)^2 \times \sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} T(i,j)^2}} \qquad (3)$$

NCC もテンプレートを動かしながら最もマッチングする位置を探索する．NCC に対して，パラボラフィッティングを適用することで，サブピクセルレベルのマッチングが可能である [12], [13].

POC

POC は，画像をフーリエ変換して得られる位相情報を用いて画像をマッチングする手法であり，画像間の類似度を計算している．フーリエ変換後の画像を $F(k_1,k_2)$ および $G(k_1,k_2)$ とすると，正規化相互パワースペクトル $R(k_1,k_2)$ は次式で定義される．

$$R(k_1,k_2) = \frac{F(k_1,k_2)\overline{G(k_1,k_2)}}{|F(k_1,k_2)\overline{G(k_1,k_2)}|} \qquad (4)$$

上式を逆フーリエ変換することで，POC 関数が得られる．2 つの画像が類似している場合，POC 関数は，デルタ関数に近いきわめて鋭いピークを有する．この相関ピークの高さは画像の類似度の尺度として有用であり，一方，相関ピークの座標は 2 つの画像の相対的な位置ずれに対応する．連続空間で定義された相関ピークモデルをフィッティングすることで，サブピクセルレベルのマッチングが可能である [14].

3.2 探索手法

特徴ベースマッチングは，特徴量間の距離などを調べることで画像間を対応付けることができる．一方で，領域ベースマッチングは，入力画像に対してテンプレート

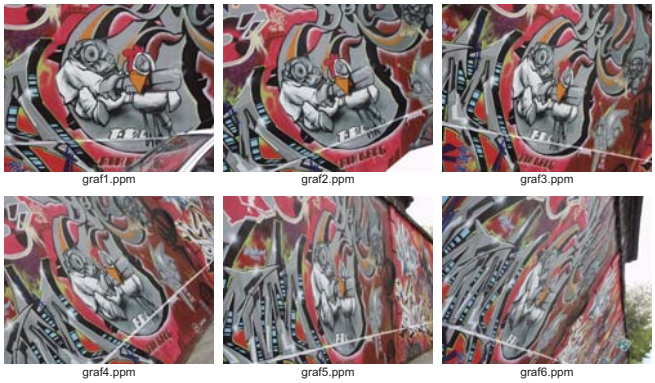


図 2: 実験に用いた画像

を走査させながらマッチングする必要がある．もっともシンプルな探索手法は，画像全域にわたって走査する全探索である．これに対して，処理時間を減少させ，さらに局所解に陥るのを防ぐために階層探索が用いられる．階層探索の詳細については，文献 [15], [16] を参照されたい．本論文では，階層探索を用いて画像をマッチングする．

4 特徴ベースマッチングと領域ベースマッチングの組み合わせ

ここでは，特徴ベースマッチングと領域ベースマッチングを組み合わせることで，高精度かつ密に画像マッチング可能な手法を提案する．

特徴ベースマッチングの画像変形にロバストである特長と領域ベースマッチングの密に対応づけられる特長を融合するために，(i) 特徴ベースマッチングを使って画像間の大きな変形を補正し，(ii) 領域ベースマッチングを使って画像間を密にマッチングする手法を提案する．具体的には，以下のような手順でマッチングする．

Step 1: ASIFT のように，アフィン変形を用いてさまざまな視点から撮影した画像を生成する．生成した画像を SURF でマッチングする．ASIFT では SIFT を用いているが，本論文では，計算時間を抑えるために SURF を用いる．

Step2: 次に，得られた対応関係から，画像間の射影変形パラメータを求め，画像間の大きな変形を補正する．以上の処理により，画像間の変形は，ほぼ平行移動のみとなる．

Step 3: 補正した画像に対して基準点を配置し，基準点に対する対応点を求める．本論文では，POC に基づく対応点探索を用いる [15], [16]．また，基準点は，5 画素間隔で配置する．

5 実験と考察

公開されている標準画像を用いて，これまでに提案されている特徴ベースマッチングと提案手法の性能を評価し，提案手法の有効性を実証する．

標準画像として，文献 [4], [7], [8] で用いられている graffiti[1] を用いる．この画像は，図 2 のように，視点が大きく変化してる．従来法として，SIFT[2], SURF[3], Hariss-Affine[1], Hessian-Affine[1], ASIFT[4] を用いる．それぞれの手法は，脚注より入手可能な実行ファイルを用いる．ただし，Harris-Affine および Hessian-Affine の特徴量として，SIFT および GLOH を用いる．

マッチング精度の評価には，対応付けの精度（誤差）を用いる．本実験で用いる画像は，1 枚目の画像からその他の画像への射影変形行列が与えられているので[1]，基準点を正解の変形行列を用いて投影し，マッチング結果との画像上での距離を用いて対応付けの誤差を求める．

図 3 および 4 に実験結果をまとめたグラフを示す．これらは，横軸を正解との距離（誤差）とし，縦軸を全対応点数に対する正解の割合としてプロットしている．画像変形が小さいペアについては，すべての手法において十分なマッチング精度を有している．一方で，画像変形が大きなペアについては，MSER_SIFT，MSER_GLOH，ASIFT，POC（提案手法）のマッチング精度が高いことがわかる．表 1 は，正解との距離が 1 画素以内であった対応点の数を示している．これより，ASIFT および POC（提案手法）は，画像変形の大きさにかかわらず，密なマッチング結果を示している．

図 5 に，得られた対応点を画像上にプロットした結果を示す．MSER_GLOH の結果は，画像変形が大きすぎるために，十分な対応点が得られていない．ASIFT は，大きな画像変形があっても十分な対応点が得られているが，特徴を抽出できなかった領域については，対応点を得ることができていない．一方で，POC（提案手法）は，十分な数の対応点が得られていることがわかる．

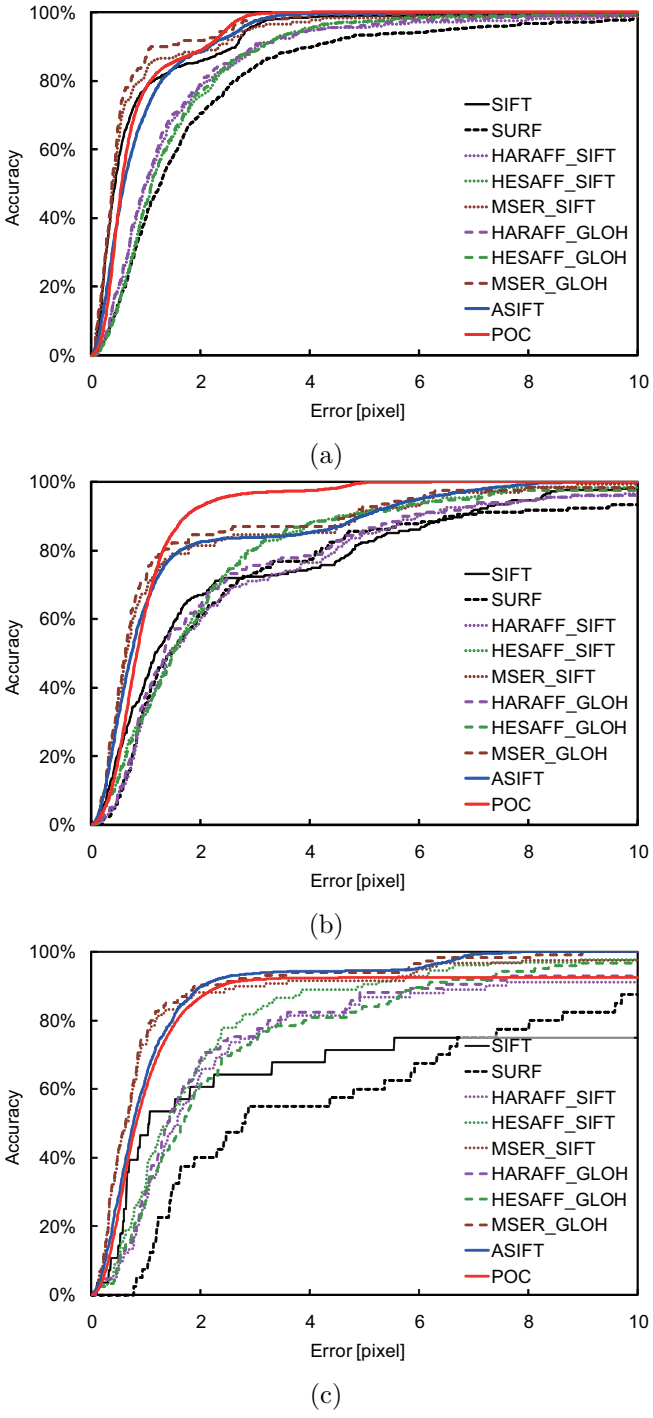
提案手法は，特徴ベースマッチングと比べて，低速である．はじめに，画像を変形させ，各画像ペアを特徴ベースマッチングで対応付けているためである．たとえば，低解像度画像を用いて大きな画像変形を推定したり，文献 [11] のように，あらかじめ学習をして，対応付ける数を減らすことも考えられる．

以上より，提案手法を用いることで，画像変形が大きな画像に対しても，密で高精度なマッチングが可能であることを示した．

[1] http://www.robots.ox.ac.uk/~vgg/research/affine/
[2] http://www.cs.ubc.ca/~lowe/keypoints/
[3] http://www.vision.ee.ethz.ch/~surf/
[4] http://www.ipol.im/pub/algo/my_affine_sift/

表 1: 実験結果

	SIFT	SURF	HARAFF_S	HESAFF_S	MSER_S	HARAFF_G	HESAFF_G	MSER_G	ASIFT	POC
1-2	817	226	186	271	142	179	262	138	2,062	7,967
1-3	104	64	69	97	87	69	93	88	1,670	3,993
1-4	13	3	25	41	87	26	35	86	1,223	3,802
1-5	0	1	5	5	52	3	7	53	674	1,388
1-6	0	2	1	3	30	1	1	30	457	400

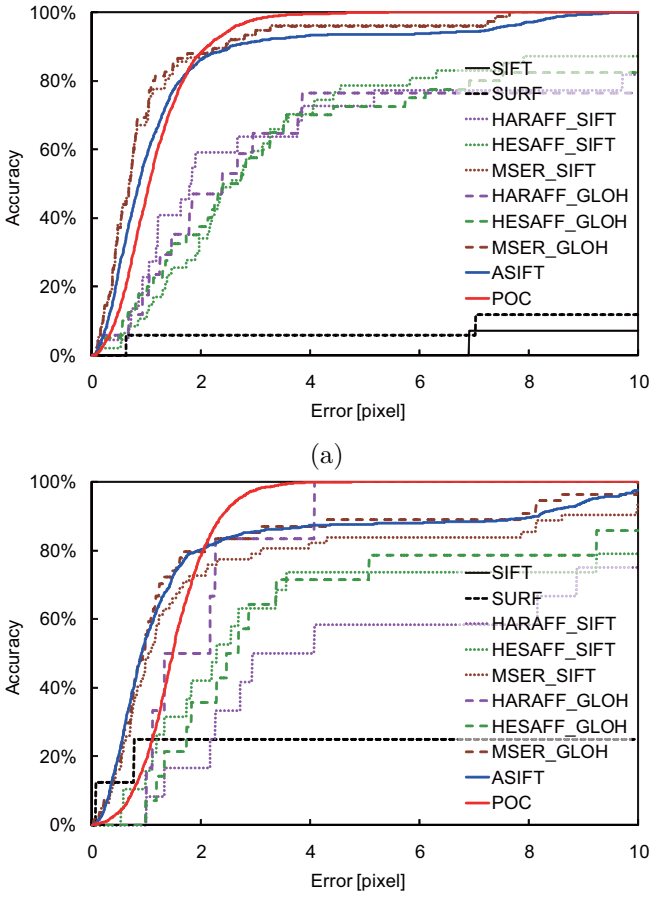


(a)

(b)

(c)

図 3: 実験結果：(a) graf1–2，(b) graf1–3，(c) graf1–4



(a)

(b)

図 4: 実験結果：(a) graf1–5，(b) graf1–6

6 まとめ

本論文では，密かつ高精度な画像マッチング手法を提案した．性能評価実験を通して，現在までに提案されている画像マッチング手法よりも高性能であることを示した．この結果は，たとえば，ワイドベースラインのステレオ画像を密に対応づけることを可能とする．それ以外にも，時間が経過したり，カメラが大きく動いた映像シーケンスでも高精度な動き推定を可能とする．今後は，計算時間の高速化や，さらなる精度の向上を検討する予定である．

参考文献

[1] 奥富正敏（編）：“ディジタル画像処理”，CG-ARTS 協会 (2004).

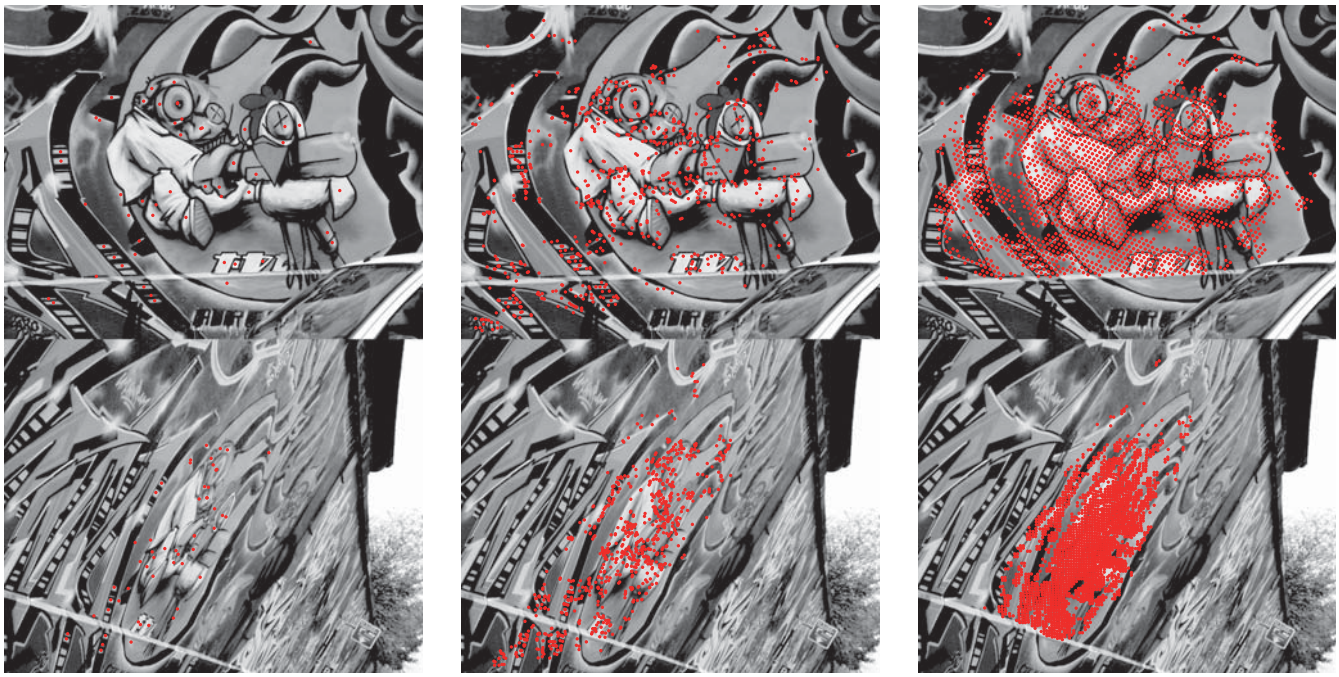


図 5: 得られた対応点の例

[2] R. Szeliski: “Computer Vision: Algorithms and Applications”, Springer (2010).

[3] D. Lowe: “Distinctive image features from scale-invariant keypoints”, Int’l. J. Computer Vision, **60**, 2, pp. 91–110 (2004).

[4] K. Mikolajczyk and C. Schmid: “A performance evaluation of local descriptors”, IEEE Trans. Patt. Anal. Machine Intell., **27**, 10, pp. 1615–1630 (2005).

[5] H. Bay, A. Ess, T. Tuytelaars and L. Gool: “Supeeded-up robust features (SURF)”, Computer Vision and Image Understanding, **110**, pp. 346–359 (2008).

[6] J.-M. Morel and G. Yu: “ASIFT: A new framework for fully affine invariant image comparison”, SIAM J. Imaging Sciences, **2**, 2, pp. 438–469 (2009).

[7] T. Tuytelaars and K. Mikolajczyk: “Local invariant feature detectors: A survey”, Found. Trends. Comput. Graph. Vis., **3**, 3.

[8] K. Mikolajczyk and C. Schmid: “Scale & affine invariant interest point detectors”, Int’l J. Comput. Vision, **60**, 1.

[9] J. Matas, O. Chum, M. Urban and T. Pajdal: “Robust wide baseline stereo from maximally stable extremal regions”, Proc. British Machine Vision Conf., pp. 384–393 (2002).

[10] Y. Ke and R. Sukthankar: “PCA-SIFT: A more distinctive representation for local image descriptors”, Proc. IEEE Comput. Society Conf. Comput. Vision and Pattern Recognition, **2**, (2004).

[11] 西村孝, 清水彰一, 藤吉弘亘：“2 段階の rondomized trees を用いたキーポイントの分類”, 画像の認識・理解シンポジウム, pp. 1412–1419 (2010).

[12] M. Shimizu and M. Okutomi: “Sub-pixel estimation error cancellation on area-based matching”, International Journal of Computer Vision, **63**, 3, pp. 207–224 (2005).

[13] M. Shimizu and M. Okutomi: “Multi-parameter simultaneous estimation on area-based matching”, International Journal of Computer Vision, **67**, 3, pp. 327–342 (2006).

[14] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi and K. Kobayashi: “High-accuracy subpixel image registration based on phase-only correlation”, IEICE Trans. Fundamentals, **E86-A**, 8, pp. 1925–1934 (2003).

[15] K. Takita, M. A. Muquit, T. Aoki and T. Higuchi: “A sub-pixel correspondence search technique for computer vision applications”, IEICE Trans. Fundamentals, **E87-A**, 8, pp. 1913–1923 (2004).

[16] M. A. Muquit, T. Shibahara and T. Aoki: “A high-accuracy passive 3D measurement system using phase-based image matching”, IEICE Trans. Fundamentals, **E89-A**, 3, pp. 686–697 (2006).