テンソル分解による視点合成フィルタ群のコンパクト化

河合康平 †長谷川昂宏 †山下隆義 †藤吉弘亘 †

†中部大学

E-mail: koh@mprg.cs.chubu.ac.jp

Abstract

キーポイントマッチングは、特定物体認識やパノラマ 画像生成等の様々なタスクで用いられる. ASIFT は射 影変化に不変なマッチングを行うために、入力画像を アフィン変換して複数の視点合成画像を生成し、生成し た全ての画像から SIFT 特徴量を記述する. 視点合成 画像は、オンライン処理でアフィン変換を行うため、計 算コストが高いという問題がある. この問題を解決す るために、提案手法は ORB 記述子をフィルタとして表 現し、このフィルタにアフィン変換を施して特徴量を記 述する. このとき、アフィン変換した視点合成フィルタ 19,200 枚の畳み込みを行うことは非効率である. そこ で、視点合成フィルタ群を4つの軸によるテンソルを構 築する. このテンソルを Touker モデルへ分解する際の 次元削減を軸ごとに行い、コンパクトかつ精度を維持し た特徴量記述を実現する.

1 はじめに

キーポイントマッチングは,特定物体認識やパノラマ 画像生成等の様々なタスクで用いられる.キーポイン トマッチングを用いたアプリケーションは,異なる視点 で撮影された画像の対応付けを行わなければならない. Affine SIFT(ASIFT)[1]は,入力画像をアフィン変換し て複数の視点合成画像を生成し,生成した全ての画像か ら Scale-Invariant Feature Transform(SIFT)[2]を用い て特徴量を記述する (図 1(a)). このとき,オンラインで アフィン変換を行うため,計算コストが非常に高いとい う問題がある.

そこで、本手法では特徴量記述子を特徴記述フィルタ の畳み込みで表現し、このフィルタにアフィン変換を施 して視点合成を行い、視点合成フィルタを生成する.視 点合成フィルタを特徴点周辺画像に畳み込みを行うこ とで視点合成に基づいた特徴量を記述できる(図1(b)). しかし、精度の良いマッチングを行うためにアフィン変 換を密に行う必要があり、19,200枚の視点合成フィル タが必要となる.この生成された全てのフィルタを畳 み込むことは計算コストが高くなり非効率である.そ



図1 特徴記述フィルタの概要

こで,視点合成フィルタ群を因子分解法を用いてコンパ クト化する.これまでに、アフィン領域抽出の為の異方 性 Laplacian-of- Gaussian(LoG) フィルタを特異値分 解(SVD)を用いてコンパクト化する手法が提案されて いる [3]. しかし, SVD では1つの軸にのみ次元削減を 行うため、最適な次元削減が行われているとは限らない. そこで、本研究ではテンソル分解を用いて視点合成フィ ルタのコンパクト化する手法を提案する、テンソルは、視 点合成フィルタをベクトル化したフィルタ軸, サンプリン グペア数となるサンプリングペア軸,アフィン変換で用 いる2つのアフィンパラメータ軸の4つの軸(モード)で 4階テンソルを構築する. テンソル分解には the higherorder singular value decomposition[4](HOSVD) を用 いて Touker モデル [5] ヘ分解を行う. このとき, 各モー ドに特異値分解を行うため,モードごとに次元削減が可 能である.これにより、特徴量記述の精度に影響が高い モードの次元を多く採用し、影響の低いモードの次元を より削減することで最適な次元削減を実現し, SVD よ りもコンパクトなサイズで特徴量記述の近似が可能と なることに示す.

パッチ画像の輝度や勾配は,特徴量記述の有効な情報 であり,これらを特徴量として記述する手法が提案され ている [11, 12, 13, 6, 7]. さらに,視点変化に対応した 特徴量記述として,視点合成に基づく手法が提案されて いる [1, 10].本節では,従来手法として輝度ベースの特 徴記述子,視点合成ベースの特徴記述子,SVDを用いた 視点合成フィルタ群のコンパクト化について説明する.

2.1 輝度ベースの特徴記述子

輝度に基づく特徴記述方法は、パッチ画像内の2点の輝 度差を用いる。Binary robust independent elementary features (BRIEF) [11] は、256 個または512 個の無作 為にサンプリングされたピクセルペアの輝度差により 特徴量を記述する.また、Oriented FAST and Rotated BRIEF(ORB)[12] は分散が最大となるサンプリングペ アの組み合わせを greedy 探索する. Discriminative-BRIEF(D-BRIEF)[13] は、パッチ画像 I と重み W の内 積を特徴とする.これらの輝度ベースの特徴量は、輝度 差を2 値化して特徴量を表現するため、距離計算の高速 化だけでなく、メモリを節約する効果がある.

2.2 視点合成ベースの特徴記述子

視点変化が強い画像間のマッチングを行うために,視 点合成に基づいた特徴記述が提案されている.ASIFT[1] は入力画像に対して複数の視点合成変換を行い,SIFT 特徴量[2]を記述する.これにより,視点合成に基づい た特徴量を記述する.本研究では計算コストの高い入力 画像の視点合成変換をフィルタの視点合成変換に置き 換える.Affine Subspace Representation(ASR)[10] で は入力画像のパッチ画像にアフィン変換を施す.生成 されたアフィン変換パッチ画像に学習用パッチ画像か ら生成した PCA 射影行列を掛けることで特徴量を記述 する.また,同じキーポイントの特徴量を部分空間に投 影することで ASIFT よりロバストな特徴量を表現して いる.

2.3 線形フィルタ群のコンパクト化

アフィン領域抽出の為の異方性 LoG フィルタを特異 値分解を用いてコンパクト 化する手法が提案されてい る [3]. 異方性 LoG フィルタは x, y 方向のスケールと 回転角の 3 つパラメータで生成される線形フィルタで ある. この全てのフィルタ群をパッチ画像に畳み込むの は非効率的である.そこで, SVD を用いて線形フィル タ群のコンパクト化を行う.列ベクトルをフィルタサ イズ,行ベクトルをフィルタ群の数とした行列 L を生 成して以下の式のように分解する.

$$\boldsymbol{L} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}} \tag{1}$$

分解後の特異値行列 **S**より基底数を決定し,次元削減 を行う.次元削減後の行列をそれぞれ **U**', **S**', **V**' とし, パッチ画像をベクトル化したものを *I* として以下の式 で LoG フィルタの応答値を計算する (図 2).

$$\boldsymbol{F} = \boldsymbol{I}\boldsymbol{U}'\boldsymbol{S}'\boldsymbol{V}'^{\mathrm{T}}$$
(2)

パッチ画像 *I* は畳み込み回数を削減するため, 先に *U'*と計算する. しかし, SVD では行列 *V* のサイズが大きくなり, 処理コストが思うように削減されない問題がある.



図 2 LoG フィルタの応答値の計算

3 提案手法

提案手法では, ORB 特徴量を線形フィルタの畳み込 みに置き換えて用いる. このフィルタに対してオフラ イン処理でアフィン変換による視点合成を行うことで 視点合成に基づいた特徴量記述が可能となる. また, 視 点合成フィルタ群をテンソル分解によりコンパクト化 することで畳み込み処理回数や全体のサイズを削減し, 計算コストを大幅に削減する.

3.1 ORB 特徴記述子の線形フィルタ表現

提案手法では,特徴量記述子を線形フィルタの畳み込みで表現する.記述子には単純かつ性能が良い ORB のサンプリングパターン (図 3)を用いる.線形フィルタは ORB のサンプリングペアに "+1"と "-1",それ以外には "0"を代入して生成する.線形フィルタをベクトル化したものを $\omega^{d=1,2,\dots,256}$,特徴点周辺画像をベクトル化したものをIとする.このとき,d番目のピクセルペアの輝度差 f^d は式(3)で計算される.

$$f^d = (\boldsymbol{\omega}^d)^{\mathrm{T}} \boldsymbol{I} \tag{3}$$





3.2 視点合成に基づく特徴量記述

視点変化が生じた画像間のホモグラフィ行列 **H** は画 像の局所領域においてアフィン行列 **H**_i で近似可能で ある.

$$\boldsymbol{H}_{i} = \lambda \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$
(4)

 $\lambda > 0$ はスケールパラメータ, $\psi \in [0, 2\pi)$ はカメラ軸 の回転パラメータ, t > 1は傾きパラメータであり緯度 に相当する. $\phi \in [0, \pi)$ は経度の回転パラメータに相当 する. このとき, $\lambda \ge \psi$ は SIFT のスケールとオリエン テーションで代用することができるため, 以下の式のよ うに置き換えることが可能である.

$$\boldsymbol{H}_{i} = \begin{bmatrix} t & 0\\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi\\ \sin\phi & \cos\phi \end{bmatrix}$$
(5)

このアフィン行列 H_i を用いて特徴量記述子である線 形フィルタをアフィン変換する.アフィン変換により 生成した視点合成フィルタを $\omega^d(t,\phi)^{\mathbf{T}}$ とすると視点合 成に基づく特徴量 $f^d(t,\phi)$ は式 (6) で計算される.パ ラメータは $\lambda \geq \psi$ を除外した 2 つのパラメータ t ={1.0, 1.2, 1.4, · · · , 2.0}, $\phi =$ {0°, 12°, 24°, · · · , 168°} を 用いてアフィン変換する.

$$f^{d}(t,\phi) = \boldsymbol{\omega}^{d}(t,\phi)^{\mathrm{T}}\boldsymbol{I}$$
(6)

特徴次元は d = 256 とする.

3.3 テンソル分解による視点合成フィルタ群のコンパ クト化

視点合成フィルタは線形フィルタをアフィン変換す ることで 19200 枚のフィルタの畳み込みを行うことに なる.そのため、テンソル分解を用いてコンパクト化を 行う.テンソル分解のモデルには Touker モデルを使用 し、分解法は HOSVD を用いる.式 (7) のように視点合 成フィルタ群を分解する (図 4).

$$\mathbf{W} = \mathbf{G} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \times_3 \boldsymbol{U}_3 \times_4 \boldsymbol{U}_4 \tag{7}$$

ここで, テンソル **W** $\in \mathbb{R}^{I_1 \times I_2 \times I_3 \times I_4}$ のモード-1 は視 点合成フィルタ $\omega^d(t, \phi)$ で構成される. I_1 はフィルタ サイズ (49 × 49 = 2401), I_2 は t のパラメータ数, I_3 は ϕ のパラメータ数, I_4 はサンプリングペア数 d である.

3.4 特徴量の記述

テンソル **G** $\in \mathbb{R}^{i_1 \times i_2 \times i_3 \times i_4}$ はコアテンソル, 行列 $U_1 \in \mathbb{R}^{I_1 \times i_1}$ はフィルタサイズの行列, 行列 $U_2 \in \mathbb{R}^{I_2 \times i_2}$ はtのパラメータの行列, 行列 $U_3 \in \mathbb{R}^{I_3 \times i_3}$ は ϕ のパラメータの行列, 行列 $U_4 \in \mathbb{R}^{I_4 \times i_4}$ は ϕ のサンプ リングペアの行列である. パッチ画像 I の畳み込み回数 は U_1 のサイズに依存する. しかし, U_2 , U_3 , U_4 のサ イズを小さくすることでコアテンソルが小さくなり提 案手法のコンパクト化に繋がる. U_1 とパッチ画像をベ



図 4 4 階テンソルの構築と Touker モデルへの 分解

クトル化した行列 I は予め計算して行列 $IU_1 \in \mathbb{R}^{1 \times i_1}$ とする.

$$\boldsymbol{I}\boldsymbol{U}_1 = \boldsymbol{I} \times \boldsymbol{U}_1 \tag{8}$$

次に特徴量を以下の式で計算する.

$$\mathbf{F} = \mathbf{G} \times_1 \boldsymbol{I} \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \times_3 \boldsymbol{U}_3 \times_4 \boldsymbol{U}_4$$
(9)

ここで, **F** の各要素が式 (6) の $f^d(t, \phi)$ となり, 特徴量 が記述される (図 5).



4 評価実験

提案手法の有効性を示すために3つの評価実験を行う.

4.1 実験概要

1 つ目にパラメータごとに次元数を変化させる有 効性を示す. 2 つ目に近似誤差との比較実験を行 う. 3 つ目にコンパクト化の評価を行列サイズを用 いて行う. また, 2 つ目と 3 つ目は SVD との比較 を行う. 実験に使用するデータセットは Region Detector Evaluation Datasets[14] と Affine Convariant Regions Datasets[15] から Underground(viewpoint), Grace(viewpoint), Posters(viewpoint), Boat(zoom and rotation), bikes(blur) and Leuven(light) を用い る. また, 評価指標にはキーポイントマッチングの精度 である matching rate(correct matches/all matches) を 用いる.

4.2 次元数を変化させた際の性能評価

この実験は次元を下げた場合にマッチング精度が下 がりやすいモードを調査する.実験はモード-2,3,4に ついて行う.この3つのモード内,2つのモードを特異 値の寄与率90%までの削減で固定する.固定したモー ド以外の次元を削減する.図6より一番精度の低下が大 きいモードはサンプリングペアである.また,パラメー タtはほとんどマッチング精度に影響が出ないことが 確認できた.



図 6 単一のモードを次元削減した際のマッチン グ精度の変化.

4.3 近似誤差を用いた比較実験

SVD との比較実験を行う. 比較には近似誤差を用いる. 近似誤差の算出にはフロベニウスノルム ||・||_Fを用いて以下の式より計算する.

近似誤差 =
$$||\mathbf{W} - \mathbf{W}'||_{\mathrm{F}}$$
 (10)

ここで、W は元テンソル、W' はテンソル分解後に復 元したテンソルとする. SVD とテンソル分解で同等の 近似誤差のマッチング精度を比較する. 図7の結果よ り SVD より精度が良い場合があることがわかる. こ

表1 図7で用いたテンソルの基底数

	ϕ	t	sampling pair	coretensor size
proposed1	2	2	85	340
proposed2	12	3	13	468

れは、同等の誤差でも次元削減のパターンによって精度 に大きな影響が出るためである.次に、近似誤差が約 20 でのマッチング率の比較を行う.proposed1 はサン プリングペアの次元のみを多く残したパターンである. proposed2 はパラメータ t とパラメータ φ の次元を多 く残したパターンである (表 1).図 9 より、proposed1, 2 を比較すると proposed1 の精度が良く、図 6 の傾向 通りであると言える.また、コアテンソルのサイズも proposed1 の方が小さくなっている.これらにより、テ ンソル分解で適切な次元削減を選択することの有効性 がわかる.また、propsed1 は SVD よりも精度が良いこ とも確認できる.図 8 はマッチング結果である.視点変 化が起こった画像間で 250 点以上の正解点を対応付け られている.



図 7 近似誤差による SVD との比較

4.4 コンパクト化の評価

次にコンパクト化の評価を行う. 評価方法として マッチング率と分解後のサイズを用いる. 分解後のサ イズとして特異値分解のサイズを式 (2)のU', S', V'のサイズの和とする. また, テンソル分解を式 (7)の G, U_1, U_2, U_3, U_4 のサイズの和とする. 図 10 は同等 のサイズでのマッチング精度の比較である. 結果より 同等のサイズでテンソル分解の方が精度が良い場合が あることがわかる. また, 表 2 は同等のマッチング精度 でのサイズの比較を行う. 結果より提案手法は SVDの 約 1/2 のサイズで同等のマッチング精度を実現可能で ある. これらにより, 提案手法は SVD より省メモリで 良い精度のマッチングを実現できる.



図8 マッチング結果



図 10 サイズによる SVD との比較

	表 2	同等の精度でのコンパク	ト化の比較
--	-----	-------------	-------

	平均マッチング率	サイズ
提案手法	70.18%	340113
SVD	70.96%	765972

4.5 処理コストの評価

最後に処理コストの評価を行う.今回は畳み込み処 理の回数を同一するために式(2)の行列U'と式(7)の 行列U1のサイズを同じに設定する.表3は精度が同 等の際の計算コストと処理時間の比較である.このと き,計算コストを畳み込み以外のfor文の回数はSVDの 約12%であり,計算コストが削減されている.また,処 理時間はSVDの約19%まで削減されている.また,図 11はパッチ画像に畳み込みする行列以外の行列とテン ソルのサイズを実際の比率で可視化し,比較した結果で ある.提案手法はSVDよりも小さいサイズになってい ることが直感的にわかる.これらにより,SVDで計算 コスト削減の問題となっていた行列Uと行列SV^Tの 積算を提案手法では大きく軽減し,処理時間を抑えるこ とができる.

表3 処理コストの比較

	平均マッチング率	for 文の回数	処理速度
提案手法	67.191[%]	1,120,320[回]	29.00[s]
SVD	66.785[%]	6,835,200[回]	156.206[s]
		19,200	
SVD 375-	S	${^{\prime\prime}}m{V}{^{\prime\mathrm{T}}}$	5
	686	τ	
提案手法 375-	G	/4 J ₃ J ₂	

図 11 サイズの可視化

5 おわりに

本論文では、テンソル分解を用いたアフィン線形フィ ルタの次元削減を提案した. SVD では元の特徴フィル タ行列との誤差が大きく精度が下がってしまう場面で も,提案手法ではモードごとの精度低下を調査すること で,同等の近似誤差で高い精度を維持できることを確認 した. また、提案手法は SVD よりも約 1/2 のサイズで 同等の精度を実現した.また,畳み込み処理回数を統一 した際に提案手法は SVD の約 12%まで計算コストを 削減でき、処理時間では SVD の処理時間の約 19%まで 削減できた. これらより, テンソル分解を用いた特徴記 述フィルタの削減では各モードで適切な次元削減を行 うことでよりコンパクトな行列及びテンソルで高い精 度を維持でき、かつ低い計算コストを実現できることを 確認した. 今後の研究では、様々なテンソル分解手法を 用いて研究を行う.また, ORB 以外の記述子でも提案 手法の有効性が示せるか実験を行う.



図 9 近似誤差が約 20 におけるマッチング率の比較

参考文献

- Jean-Michel Morel and Guoshen Yu, "ASIFT: A new framework for fully affine invariant image comparison," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 438–469, 2009.
- [2] David G Lowe, "Distinctive image features from scale-invariant keypoints," *International journal* of computer vision, vol. 60, no. 2, pp. 91–110, 2004.
- [3] Takahiro Hasegawa, Mitsuru Ambai, Kohta Ishikawa, Gou Koutaki, Yuji Yamauchi, Takayoshi Yamashita, and Hironobu Fujiyoshi, "Multiple-hypothesis affine region estimation with anisotropic log filters," in *The IEEE International Conference on Computer Vision* (*ICCV*). IEEE, 2015, pp. 585–593.
- [4] Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle, "A multilinear singular value decom-

position," SIAM journal on Matrix Analysis and Applications, vol. 21, no. 4, pp. 1253–1278, 2000.

- [5] Ledyard R Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.
- [6] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "SURF: Speeded up robust features," in European conference on computer vision(ECCV), 2006, pp. 404–417.
- [7] Engin Tola, Vincent Lepetit, and Pascal Fua, "DAISY: An efficient dense descriptor applied to wide-baseline stereo," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 5, pp. 815–830, 2010.
- [8] Kwang Moo Yi, Eduard Trulls, Vincent Lepetit, and Pascal Fua, "LIFT: Learned invariant feature transform," in *European conference on computer* vision (ECCV), 2016, pp. 467–483.
- [9] Vassileios Balntas, Edgar Riba, Daniel Ponsa,

and Krystian Mikolajczyk, "Learning local feature descriptors with triplets and shallow convolutional neural networks.," in *British machine* vision conference(BMVC), 2016, p. 3.

- [10] Zhenhua Wang, Bin Fan, and Fuchao Wu, "Affine subspace representation for feature description," in *European conference on computer* vision (ECCV), 2014, pp. 94–108.
- [11] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua, "BRIEF: Binary robust independent elementary features," in *European conference on computer vision (ECCV)*, 2010, pp. 778–792.
- [12] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski, "ORB: An efficient alternative to sift or surf," in *The IEEE international* conference on Computer Vision (ICCV). IEEE, 2011, pp. 2564–2571.
- [13] Tomasz Trzcinski and Vincent Lepetit, "Efficient discriminative projections for compact binary descriptors," in *European conference on computer* vision(ECCV), 2012, pp. 228–242.
- [14] Cordes Kai, Rosenhahn Bodo, and Ostermann Jorn, "Increasing the accuracy of feature evaluation benchmarks using differential evolution," in Symposium on differential evolution, 2011, pp. 1–8.
- [15] Krystian Mikolajczyk and Cordelia Schmid, "A performance evaluation of local descriptors," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.