CNN を用いた視点合成ベースの特徴量抽出に関する研究

TP18012 森康彰　　指導教授：山下隆義

1. はじめに

キーポイントマッチングは，異なる画像間の同じ点を対応づけするタスクであり，画像検索やパノラマ画像の生成，自己位置推定等に用いられる．キーポイントマッチングでは，画像間の視点や照明，回転等の変化に不変な特徴量を抽出することが重要である．Affine Subspace Representation(ASR) [2] は，ASIFT [1] の視点合成によって得られる複数視点の画像から局所特徴量を記述し，アフィン部分空間に投影することで，ロバストな特徴量を得る．この手法は，人が設計した視点条件に対応した画像変換を行うため，特徴点画像に対する最適な視点合成とは言えない．そこで本研究では，1 枚の特徴点画像に対して複数の変換パラメータを推定する Multiple affine-transformers を提案する．各変換パラメータは，独立した複数の CNN が学習によって獲得するため，入力画像に対する最適な視点合成を実現できる．

2. Affine Subspace Representation

ASR は，視点変化が生じた画像ペアのマッチングを実現するため，図 1 のように入力画像に対して複数のアフィン変換画像を用意し，特徴量を抽出する．その後，抽出した特徴量群をアフィン部分空間に投影することで，視点変化に頑健な特徴量の記述を実現している．アフィン変換パラメータである $t \geq 1$ 及び $\phi \in [0, \pi)$ は事前に人が設計したものを使用する．ここで，t 及び ϕ は小刻みに標本化しているため，計算量が増加するという問題がある．

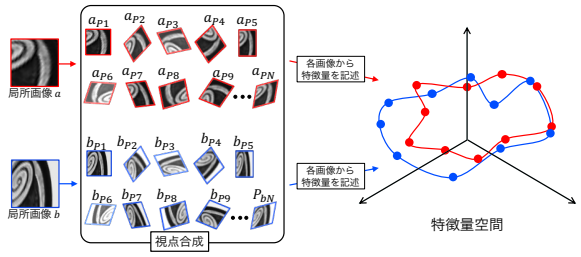


図 1：視点合成ベースの特徴量抽出

3. 提案手法

視点合成は，複数のアフィン変換パラメータを用意する必要があるため，単一の CNN では実現することが難しい．そこで，提案する Multiple affine-transformers は，図 2 に示すように重みを共有しない複数の CNN にてアフィン変換パラメータを推定し，視点合成画像を得る．

3.1 Multiple affine-transformers

Multiple affine-transformers の各 CNN はそれぞれ異なるアフィン変換パラメータを推定する．推定したパラメータは，式 (1) に示す線形アフィン行列 $\mathbf{H}_A$ に代入する．

$$\mathbf{H}_A = \lambda \begin{bmatrix} \cos(\psi) & -\sin(\psi) \\ \sin(\psi) & \cos(\psi) \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix} \quad (1)$$

ここで，スケールパラメータ λ は，特徴点検出時のスケールに置き換えることが可能なため $\lambda = 1$ とする．同様に，視点軸に対応する面内回転パラメータ ψ は視点合成後に推定するオリエンテーションに置き換える．

3.2 学習方法

Multiple affine-transformers の出力であるアフィンパラメータは，教師データが存在しない．そのため，同じ特徴点を異なる撮影条件にて撮影した局所画像ペアを使用して学習を行う．損失関数には式 (2) を用いる．これは，Triplet loss をベースとした損失関数であり，Positive pair に設定した特徴量間のユークリッド距離を小さく，Negative pair

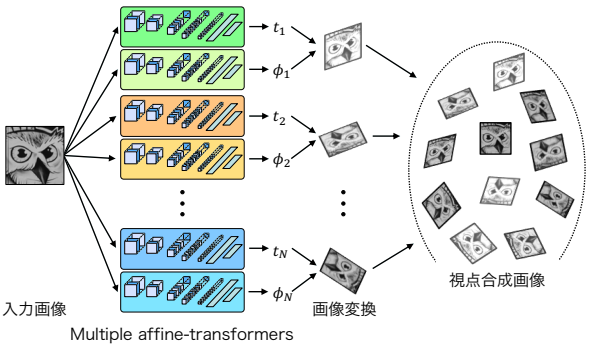


図 2：提案手法の概要

に設定した特徴量間のユークリッド距離を大きくするようにネットワークを学習する．

$$\mathcal{L}_{train} = \frac{1}{2N}\left\{\sum_{i=1}^{N} \max\left(0, E\left(d_i^q, d_{j_{min}}^p\right) - E\left(d_i^q, d_{i_{sec}}^q\right)\right) + \sum_{j=1}^{N} \max\left(0, E\left(d_j^p, d_{i_{min}}^q\right) - E\left(d_j^p, d_{j_{sec}}^p\right)\right)\right\} \quad (2)$$

ここで，$E(\cdot)$ は，ユークリッド距離を示す．式 (2) 中の $d_{j_{min}}^p$ は基準とする Query sample である d_i^q から最も距離が近い Positive sample である Positive pair を示し，$d_{j_{sec}}^p$ は最も距離が近い別の Query sample である Negative pair を示す．これにより，図 3 のように Positive pair 間のユークリッド距離が小さくなるように，Negative pair 間のユークリッド距離が大きくなるように学習を行う．この損失関数を使用することで，Positive sample との距離が小さくなるとともに視点合成画像同士の類似度が低くなり様々なアフィン変換画像を得ることが可能になる．

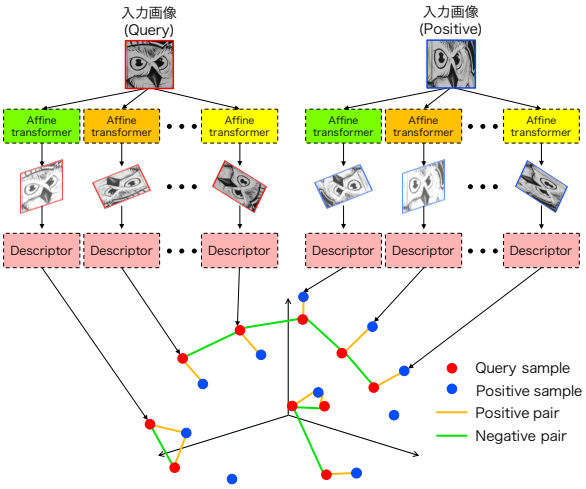
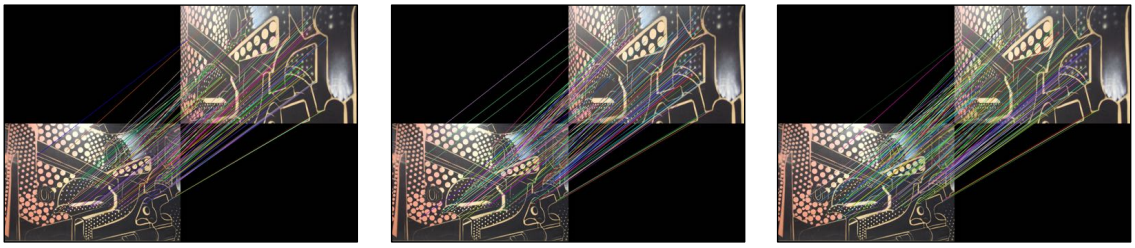


図 3：提案手法の学習

3.3 特徴量の記述

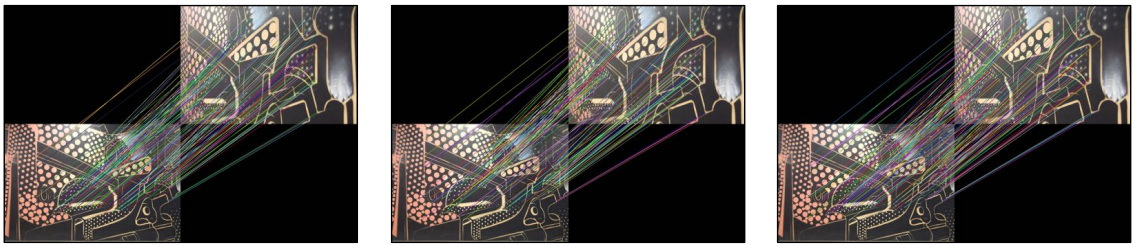
提案手法は，1 つの特徴点画像に対して複数の視点合成画像を得る．実際にキーポイントマッチングを行う際には，各画像に対してオリエンテーション推定を行い，式 (1) の ψ にあたるパラメータを求め，正規化を行う．その後，特徴量抽出を行うことで視点合成数と同じ数の多視点特徴量を得る．この多視点特徴量に対して，ASR と同様に PCA を適用し，部分空間特徴量に変換する．これにより，1 つの特徴点に対して視点変化にロバストな 1 つの特徴量を記



DeepDesc
正解マッチング数 : 103

ASR + DeepDesc
正解マッチング数 : 124

Ours + DeepDesc
正解マッチング数 : 150



HardNet
正解マッチング数 : 154

ASR + HardNet
正解マッチング数 : 157

Ours + HardNet
正解マッチング数 : 187

図 4 : キーポイントマッチング結果例

表 1 : 各特徴量記述子における Matching score

特徴抽出	視点合成	Matching score [%]	AUC
DeepDesc	-	65.5	0.873
DeepDesc	ASR	78.2	0.938
DeepDesc	Ours	**78.9**	**0.944**
HardNet	-	79.4	0.968
HardNet	ASR	84.2	0.965
HardNet	Ours	**85.7**	**0.984**

述することができる．そして，特徴量間の距離が最も近い特徴点ペアを対応点とする．

4. 評価実験

提案手法の有効性を検証するために評価実験を行う．精度比較には，Matching score(*#correct matches*/*#matches*) 及び Recall-Precision curve における area under curve(AUC) を用いる．CNN の学習及び評価には，視点変化及び照明変化が生じた画像ペアが用意されている HPatches dataset を用いる．評価には，学習に使用していないデータのうち視点変化が生じている 20 シーン 100 ペアを用いる．

本実験では，Multiple affine-transformers の最適化手法に MomentumSGD を使用し，学習率を 0.001，Momentum 係数を 0.9 とする．比較手法には，CNN を用いた特徴量抽出法である DeepDesc，HardNet を用い，各手法に対して提案手法及び ASR の視点合成を適用した際の精度を評価する．本実験では，ASR 及び提案手法の視点合成数を 30 とし，得られる多視点特徴量をアフィン部分空間に投影して得られる部分空間特徴量を用いて精度の評価を行う．

4.1 評価結果

各特徴量記述子における Matching score を表 1 に示す．結果より，CNN を用いた特徴抽出において，視点合成を適用することにより精度が向上することが確認できる．その中でも提案手法の視点合成は，ASR の視点合成と比較して精度が向上した．これは，CNN が学習によって特徴量記述子に対する最適な視点合成を学習できているためであると考えられる．

図 4 は，画像間に大きな視点変化が生じている画像ペアに対するマッチング例である．提案手法は，どちらの特徴量記述子においても，ASR と比較して正解マッチング数が 20 以上増加している．また，HardNet の結果において，視点合成を適用していない場合及び ASR の視点合成を適用した場合を比較すると精度があまり向上していないことがわかる．また，HardNet では，ASR を適用した際の正解マッチング数にあまり変化がないことから，ASR の視点合成は HardNet に有効な視点合成であるとはいえない．

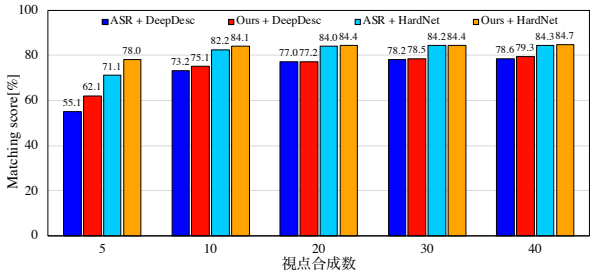


図 5 : 各視点合成数における Matching score

しかし，提案手法では正解マッチング数が 33 点増えていることから，学習により HardNet に有効な視点合成が実現できたといえる．

4.2 視点合成数の違いによる精度の比較

視点合成の数を変えた際の Matching score を図 5 に示す．図 5 より，提案手法は実験した全ての視点合成数において，ASR と比較して精度が向上していることがわかる．このことから，提案手法は設定された視点合成数における最適な視点合成を学習することができていると考えられる．さらに，HardNet において提案手法の視点合成数 20 の場合の精度は ASR の視点合成数 30 の精度と比較して高精度であり，視点合成枚数の削減が可能であることがわかる．

5. おわりに

本研究では，Multiple affine-transformers による視点合成ベースの特徴量抽出手法を提案した．視点合成パラメータを独立した CNN を用いて学習により決定することで，ASR の視点合成を使用した場合と比較して精度が向上した．今後は，Multiple affine-transformers を適宜選択する機構を導入し，視点合成数のさらなる削減を目指す．

参考文献

[1] J. Morel, *et al.*, “ASIFT: A New Framework for Fully Affine Invariant Image Comparison”, SJIS, Vol. 2, No. 2, pp. 438–469, 2009.

[2] Z. Wang, *et al.*, “Affine Subspace Representation for Feature Description”, ECCV, 2014.

研究業績

[1] 森康彰 等, “Coarse-to-Fine 構造の導入によるキーポイントのオリエンテーション推定の高精度化”, 東海支部連合大会, 2018.

[2] Y. Mori, *et al.*, “Coarse-to-Fine Deep Orientation Estimator for Local Image Matching”, ACPR, 2019.