Vol.2017-CG-168 No.27
Vol.2017-DCC-17 No.27
Vol.2017-CVIM-209 No.27
2017/11/9

ニューラルネットワークに基づく パノラマ画像を用いたカメラ位置姿勢推定

花崎 厚年[1]　内山 英昭[1]　島田 敬士[1]　谷口 倫一郎[1]

概要：本稿では，パノラマ画像に対する入力画像のカメラの位置姿勢 (平面射影行列) を推定する手法を提案する．提案手法では，DNN に基づくクラス識別を用いた入力画像のパノラマ画像上での位置推定と，入力画像とパノラマ画像間の画像の位置合わせを組み合わせることで，高精度な平面射影行列の推定を実現する．初めに，パノラマ画像を矩形領域に分割して構築した視点クラスを定義し，各視点クラスの画像領域に対して様々な平面射影変換に基づく学習用画像群を生成する．次に，生成した画像群と視点クラスの対応を DNN に訓練させ，入力を画像，出力を視点クラスとする DNN を構築する．入力画像に対する平面射影行列推定では，初めに訓練させた DNN を用いて入力画像の視点クラスを推定する．次に，視点クラスの位置を初期値とし，最適化に基づく入力画像とパノラマ画像間の画像の位置合わせを行うことで，平面射影行列を高精度に推定する．実験では，各視点クラスの画像群生成方法と視点クラスの推定精度の関係を定量的に評価し，本手法の制約を考察する．

キーワード：カメラの位置姿勢推定, パノラマ画像, DNN

HANASAKI ATSUTOSHI[1]　UCHIYAMA HIDEAKI[1]　SHIMADA ATSUSHI[1]　TANIGUCHI RIN-ICHIRO[1]

1. はじめに

デバイスの位置姿勢推定技術は，仮想現実感 (VR) や拡張現実感 (AR) を実現するための重要な要素である [12, 16]．VR におけるヘッドマウントディスプレイの位置姿勢算出では，慣性センサから得られる加速度と角速度を用いることで 3 自由度の位置姿勢を算出し，頭部の移動に応じた空間描画を実現している [26]．AR においては，カメラを用いて 3 次元位置と 3 次元姿勢の計 6 自由度の位置姿勢を算出することで，幾何学整合性の保たれた AR 表示を行うことができる [9]．屋外環境のようにカメラと物体の距離が遠い場合，その遠景を平面に近似し，平面射影行列に基づく AR 表示も行われている [13, 25]．この場合，平面射影行列を算出するための参照画像として，複数のカメラや魚眼カメラ，全方位カメラを用いて作成されたパノラマ画像を利用する枠組みが提案されている [14, 28]．この枠組みでは，AR で表示する情報をあらかじめパノラマ画像上の画素と関連付け，一般的なカメラを用いて撮影された入力画像に対する AR 表示を行っている．しかし，この枠組みでは，特徴点マッチングを利用して入力画像とパノラマ画像の対応付けを行うため [3, 15, 21]，パノラマの解像度が高くなるにつれて計算コストが高くなるとともに，算出されるカメラの位置姿勢の精度が低下することが問題であった．

本稿では，ディープニューラルネットワーク (DNN) を利用し，パノラマ画像上に対する入力画像の位置姿勢（平面射影行列）を推定する手法を提案する．DNN を用いることで，特徴点マッチングを行わず，パノラマ画像の解像度に関係なく，一定の処理速度で推定を可能とする手法を構築できる．提案手法の基本的なアイデアは，パノラマ画像からあらゆる平面射影行列の画像を合成して DNN に訓練し，入力画像の平面射影行列を DNN を用いて推定する，といった枠組みである．しかし，平面射影行列の算出を回帰問題として解く場合，学習時の誤差が大きいまま収束する，といった傾向が予備実験によって確認できた．そこで，回帰問題として解くのではなく，DNN によるクラス識別と最適化に基づく画像の位置合わせを組み合わせ，高精度なカメラ位置姿勢を算出する手法を提案する．初めに，パノラマ画像を矩形領域に分割し，各領域を 1 クラスとした視点クラスを定義する．次に，各視点クラスの画像とその

[1] 九州大学
Kyushu University

Vol.2017-CG-168 No.27
Vol.2017-DCC-17 No.27
Vol.2017-CVIM-209 No.27
2017/11/9

クラスの対応を DNN に訓練させることで，入力画像が与えられた場合に，その視点クラスを推定する DNN を構築する．各視点クラスは格子状に配置された矩形領域であるため，推定された視点クラスの位置は入力画像に対する大まかな位置姿勢である．そこで，入力画像とパノラマ画像間で最適化に基づく画像の位置合わせ行うことで，パノラマ画像に対する入力画像の平面射影行列を高精度に推定する．実験では，視点クラスを生成するときのパラメータに関して定量的に評価し，本手法の制約を述べる．

2. 関連研究

複数の参照画像をデータベースとして利用したカメラの位置姿勢推定には，機械学習に基づく画像検索技術が用いられる [22]．初めに，3 次元空間と関連付けられた各参照画像を Bag of Visual Words などの特徴ベクトルに変換することで，特徴空間を構築する．次に，位置姿勢を推定する入力画像も同様に特徴ベクトルへと変換し，類似する特徴ベクトルを特徴空間から検索することで，対応する参照画像を算出する [5]．この算出は，一般に最近傍探索の手法を利用して実現される [18]．最後に，入力画像と検索された参照画像の間で特徴点マッチングを行うことでカメラの姿勢姿勢を高精度で算出する [3, 15, 21]．

前述したカメラの位置姿勢推定問題に対し，DNN を用いた手法も提案されている [10, 17, 27]．これらの手法に共通することは，画像をカメラの位置姿勢へと射影する関数を学習する，といった枠組みである．すなわち，学習用データセットとして，カメラの位置姿勢推定を行う 3 次元空間の多数の画像とその真値となるカメラの位置姿勢の組を用意し，DNN を訓練する [10]．これにより，学習を行った 3 次元空間において，学習に用いられた視点と異なる視点の画像が入力された場合にも，そのカメラの位置姿勢を推定することが可能となる．従来より行われてきた人による特徴量の設計と比較し，DNN では学習用データセットから識別に適した特徴を自動的に学習するといった点が特徴として挙げられる．すなわち，DNN に基づく手法は，人が設計した特徴では認識できないようなテクスチャの場合にも，それを識別するための特徴を自動的に学習し，高精度な認識を実現している [2, 7, 19]．例えば，異なる季節に撮影された画像間の対応付けも可能としている [1]．

3. 提案手法の概要

図 1 と図 2 に，提案手法の概要と手法の流れを示す．提案手法では，パノラマ画像を参照画像とし，入力画像のパノラマ画像に対する平面射影行列を算出することを目的する．次章以降で述べるように，カメラの位置姿勢を推定する際，入力画像のパノラマ画像上での大まかな位置を DNN によるクラス識別を用いて算出する．このために，パノラマ画像を矩形領域に分割し，各領域を 1 クラスとした視点クラスを定義した画像データベースを構築する．次に，DNN で算出した位置を初期値とし，入力画像とパノラマ画像間で最適化に基づく画像の位置合わせ [8] を行うことで，高精度な平面射影行列を算出する．

参照画像として利用するパノラマ画像は，一般的なカメラで撮影した画像列から合成したり，広い画角のカメラを用いることで取得できる．今回は，一般的なカメラを回転体の上に設置して撮影した画像群を，Image Composite Editor[*1] を用いてパノラマ画像を作成した．なお，カメラの位置姿勢推定を行う入力画像は，一般的なカメラにより撮影した画像とする．

提案手法は，DNN による回帰，つまり画像を入力とし，パノラマ画像に対する平面射影行列を出力する，といった手法ではなく，クラス識別を用いた手法である．これは，回帰の学習を行った際，誤差が大きいまま収束するといった傾向が見られたからである．予備実験として，Detone らと同様に DNN の出力を平面射影行列となるような学習を行った [7]．この実験では，パノラマ画像上の 4 点をランダムに選択して画像を切り出して入力とし，4 点の座標を出力させるように，DNN に訓練させた．しかし，表 1 に示すように，損失関数の値が収束した場合においても，訓練画像に対する平均誤差の値が画像の幅の 10% ほどとなった．この結果より高精度な平面射影行列の推定を行うために，次章以降で説明するように，カメラの位置姿勢をクラス識別問題として解く手法を提案する．

表 1: 予備実験の結果

解像度 (pixel)	訓練画像に対する誤差 (pixel)
8192×700	70.3
4096×350	48.4

4. 視点クラスの構築

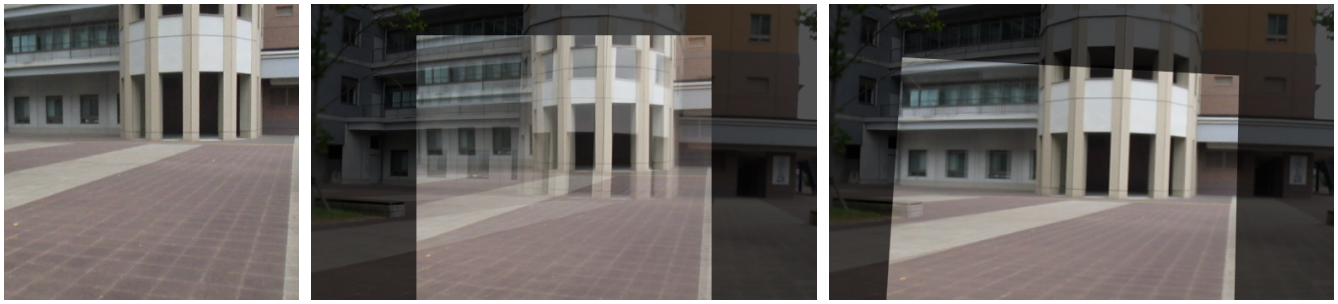
機械学習によるクラス識別手法を用いてカメラの位置姿勢推定を解くために，視点クラスを定義し，クラス識別を用いて入力画像が含まれる視点クラスを算出する．パノラマ画像を重複のある矩形領域に分割し，各領域を 1 クラスとした視点クラスを設定する．さらに，各視点クラスには，対応する領域周辺の画像に対して様々な平面射影変換を適用した画像を合成することで，視点変化に頑健な認識を実現するためのデータを生成する．

視点クラスの矩形領域の設定方法を図 3 に示す．視点クラスに属する画像は $(w+2r) \times (h+2r)$ の領域から様々な平面射影変換を用いて合成する．このとき，$w \times h$ は DNN の学習に用いられる画像サイズであり，r により視点クラスのサイズが決定する．様々な平面射影変換の画像を合成

[*1] https://www.microsoft.com/en-us/research/product/computational-photography-applications/image-composite-editor/

Vol.2017-CG-168 No.27
Vol.2017-DCC-17 No.27
Vol.2017-CVIM-209 No.27
2017/11/9

(a) 参照画像



(b) 入力画像　(c) DNN による視点クラス推定　(d) 画像の位置合わせによる高精度化

図 1: 提案手法の概要

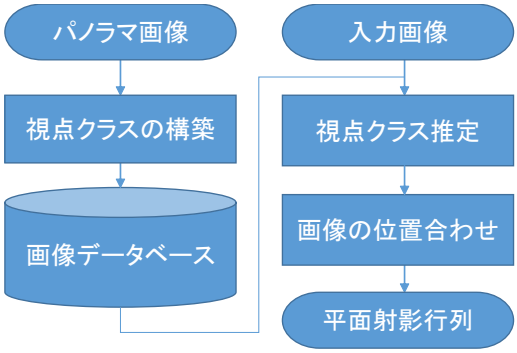


図 2: 提案手法の処理の流れ

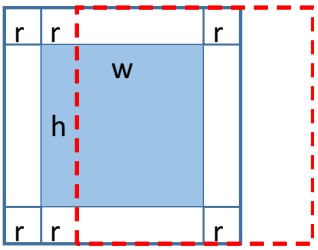


図 3: 各視点クラス領域の設定方法

する際，図 4 で示すように，画像の各 4 頂点の $2r \times 2r$ の範囲からランダムに一点選択する．次に，選択された領域を平面射影行列を用いて $w \times h$ に変換する．各視点クラスの間隔は，上下または左右に $2r$ である．つまり，r の値によってパノラマ画像上の視点クラス数が決定される．視点クラスの数 c は以下の式で計算される．

$$c = \frac{w_p - w}{2r} \times \frac{h_p - h}{2r} \tag{1}$$

ここで，w_p, h_p はそれぞれパノラマ画像の幅および高さを表す．r の違いによる視点クラスの違いの例を図 5 に示す．7 章に述べる実験においては，$w = 224, h = 224$ とし，$w_p = 4096, h_p = 350$ とした．これにより，$r = 10, 20, 30$ のときの視点クラス数は，以下のように計算される．

$$128(r = 30) = \frac{4096 - 224}{2 \times 30} \times \frac{350 - 224}{2 \times 30}, \tag{2}$$

$$288(r = 20) = \frac{4096 - 224}{2 \times 20} \times \frac{350 - 224}{2 \times 20}, \tag{3}$$

$$1158(r = 10) = \frac{4096 - 224}{2 \times 10} \times \frac{350 - 224}{2 \times 10}. \tag{4}$$

5. DNN を用いた視点クラス推定

前述した視点クラスとその領域から合成された画像群の対応を DNN を用いて学習する．近年，様々な DNN モデルが提案されている中，提案手法では屋外環境におけるカメラの位置姿勢推定を目標としているため，日時により異なる屋外環境に対しての頑健な認識を実現している DNN モデルを参考とした．従来手法において，DNN の上層の出力を特徴量として用いた場合，下層と比べて F 値が高いことや [23]，複数の畳み込み層の出力を統合して全結合層の入力とすることで高精度な認識を実現できること [1]，が報告されている．提案手法では，これらの DNN モデルに基づくネットワーク構造を構築する．

図 6 に提案手法で用いる DNN モデルを示す．このモデルは [1] を基にしており，上層の出力を統合することで環境の変化に頑健な認識の実現を目指している．各ユニットは畳み込み層，ReLU およびプーリング層から構築される．中間の 3 つのユニットは同じサイズの畳み込み層とプーリング層から構築され，出力は統合されて全結合層に入力される．最終層はソフトマックス層であり，各視点クラスの確率を出力とする．図 7 に，入力画像に対して得られる上位 k 個の視点クラスの出力結果の例を示す．なお，入力の RGB 画像のサイズは，224×224 とする．

DNN の訓練に用いる最適化手法は Adam [11] とした．損失関数にはクロスエントロピーを使用し，初期訓練率を 1.0×10^{-4} に設定する．訓練する際は，用意した訓練画像

Vol.2017-CG-168 No.27
Vol.2017-DCC-17 No.27
Vol.2017-CVIM-209 No.27
2017/11/9

図 4: 各視点クラス領域から合成した学習用画像

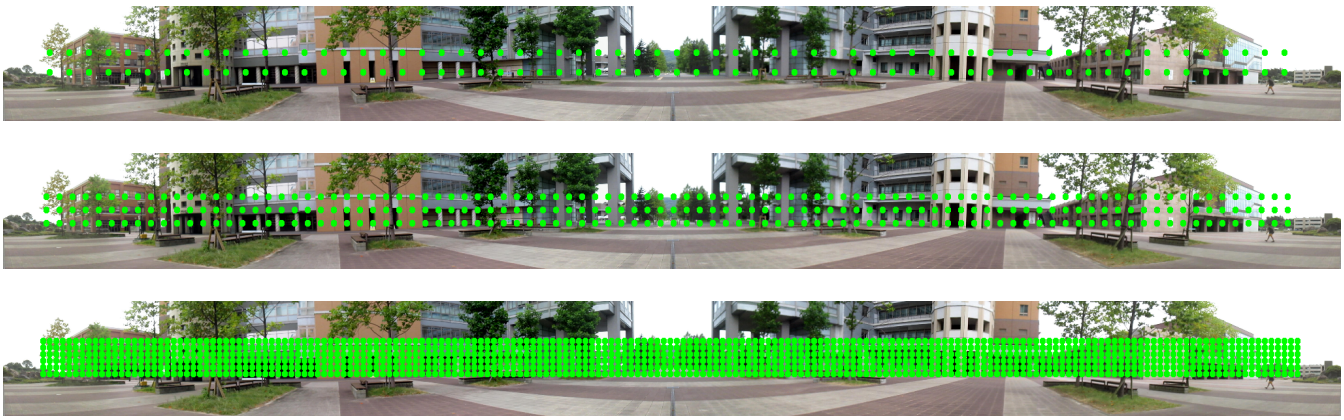


図 5: 視点クラスの中心座標の可視化（上から $r = 10, 20, 30$）

から視点クラスごとに 10 枚ランダムに選択し，選んだ画像群で一度に訓練する．訓練に用いる画像にはあらかじめ RGB の各チャンネルに対して正規化する．画像を全て訓練するまで前述の処理を繰り返す．

6. 画像の位置合わせを用いた高精度化

DNN は入力画像のパノラマ画像上における視点クラスを確率付きで出力する．しかし，図 7 に示すように，最も確率の高い視点クラスが常にパノラマ画像上の正しい位置になるとは限らない．そこで，上位 k 個の出力に対し，入力画像とパノラマ画像間で最適化に基づく画像の位置合わせを適用し，最も位置合わせの類似度が高い結果を出力する．

上位 k 個の位置に対し，Evangelidis らが提案した色強度に基づく画像の位置合わせを適用する [8]．この位置合わせ手法では，平面射影行列を変換パラメータとした最適化を行っている．初めに，各視点クラスの位置に対し，パノラマ画像から視点クラスの周辺領域を選択する．次に，図 1 に示すように，パノラマ画像から選択された画像と入力画像の間で画像の位置合わせを行う．最後に，位置合わせの結果の相関が最も高いものを選択する．これにより，適切な視点クラスを選択するとともに，平面射影行列を高精度に推定できる．

7. 実験

7.1 評価項目

本稿では，以下の 3 つの変数に関する精度評価を行う．

表 2: 評価用 DNN モデルの名前

		r		
		10	20	30
N	100	m1	m4	m7
	300	m2	m5	m8
	500	m3	m6	m9

- 各視点クラスごとに切り出す訓練画像の枚数 N
- 式 (1) の視点クラスの数を決定する r
- 画像の位置合わせの候補数を決定する k

N, r は上位 k の順位に影響を与えるため，表 2 に示すように，$r = 10, 20, 30$ 及び $N = 100, 300, 500$ のそれぞれの場合について，DNN モデルを訓練した．

7.2 評価用データセット

視点の変化への頑強性を評価するため，図 8 に示すように，複数の地点から撮影を行った．各地点間の距離は約 2m である．各地点において，回転体に載せたカメラを用いて 30 枚ずつ画像を撮影した．図 8 の中心地点で撮影された画像群から訓練用のパノラマ画像を作成し，それ以外の地点で撮影された画像をテストに利用した．テスト用画像の視点クラスの真値は手動で作成した．

7.3 結果

DNN による視点クラス推定の評価方法として，Szegedy らのように，上位 k 個の出力に真値が含まれる確率を算出した [24]．図 9 に表 2 で示した DNN モデルの精度を示す．また，表 3 に上位 10 位までの結果を利用した場合の精度

Vol.2017-CG-168 No.27
Vol.2017-DCC-17 No.27
Vol.2017-CVIM-209 No.27
2017/11/9

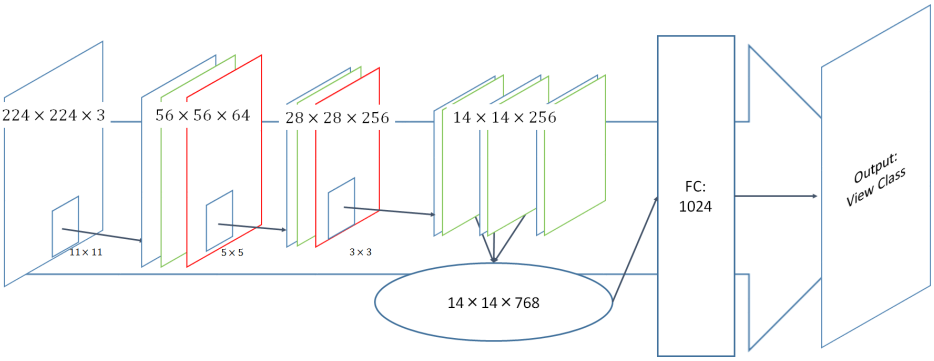


図 6: DNN モデル

図 7: DNN による入力画像の視点クラス推定の順位

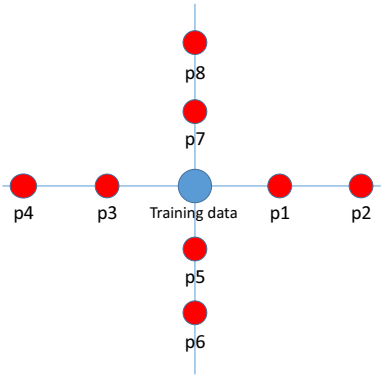


図 8: 訓練画像およびテスト画像の撮影地点

を示す．

r に注目した場合，$r = 10$ のときの m1, m2, m3 の精度が他より低いことが分かる．これは，視点クラスの密度が高すぎるために入力画像が真値に近いクラスに推定されることが原因だと考えられる．画像検索の観点では，真値となるクラスを推定する必要があるため，精度が低くなっている．しかし，真値に近いクラスを推定できており，推定したクラスを初期位置として画像の位置合わせを行うため，視点クラスの誤差の影響は低減される．$r = 30$ の場合の精度が最も高いのは，訓練画像が多様な変化やスケールを含むために見かけの変化に頑健になったからと考えられる．

N に注目した場合，精度に対して大きな差を生じさせていないことが分かる．これは，各視点クラスに含まれる画像が，視点クラスを識別するために十分存在するためだと考えられる．そのため，理論的にこの変数の値を決定できれば，精度の向上や訓練時間の削減につながる．

撮影した画像の地点に注目すると，p1 に関して精度が低くなっていることが確認できる．この結果は，p1 のテスト画像が訓練データベースの画像に対して下側を撮影してしまっており，テスト画像の一部が訓練できていないことに起因する．こういった場合にも正しく位置を推定するためには，Kendall ら [10] のように，テスト画像から複数の画像を切り出してその結果を平均することで推定することが考えられる．

7.4 失敗例

DNN は入力画像のパノラマ画像上における位置を近似的に推定するために使用し，画像の位置合わせにより高精度な平面射影行列の算出を実現した．しかし，図 10 に示すように，障害物により位置合わせが失敗する場合があった．建物の壁，木およびベンチが映っている場合には，撮影地点の変化により配置が変化し，画像の見かけも大きく変化する．DNN を訓練する際，大きな変化も学習させたため，視点クラスは推定することができた．一方で，射影変換を用いる画像の位置合わせにおいては配置の変化が原因となり画像の位置合わせに失敗する場合が存在する．これらは，他の位置合わせ手法を用いることで，この問題を解決することができる可能性がある [4, 6, 20]．

8. 結論

本稿では，DNN によるクラス識別と画像の位置合わせを組み合わせ，入力画像のパノラマ画像に対する平面射影行列を高精度に推定する方法を提案した．実験を通して，DNN を用いて視点クラスを識別できることを示し，視点クラスを画像の位置合わせを行う際の初期値として利用する

Vol.2017-CG-168 No.27
Vol.2017-DCC-17 No.27
Vol.2017-CVIM-209 No.27
2017/11/9

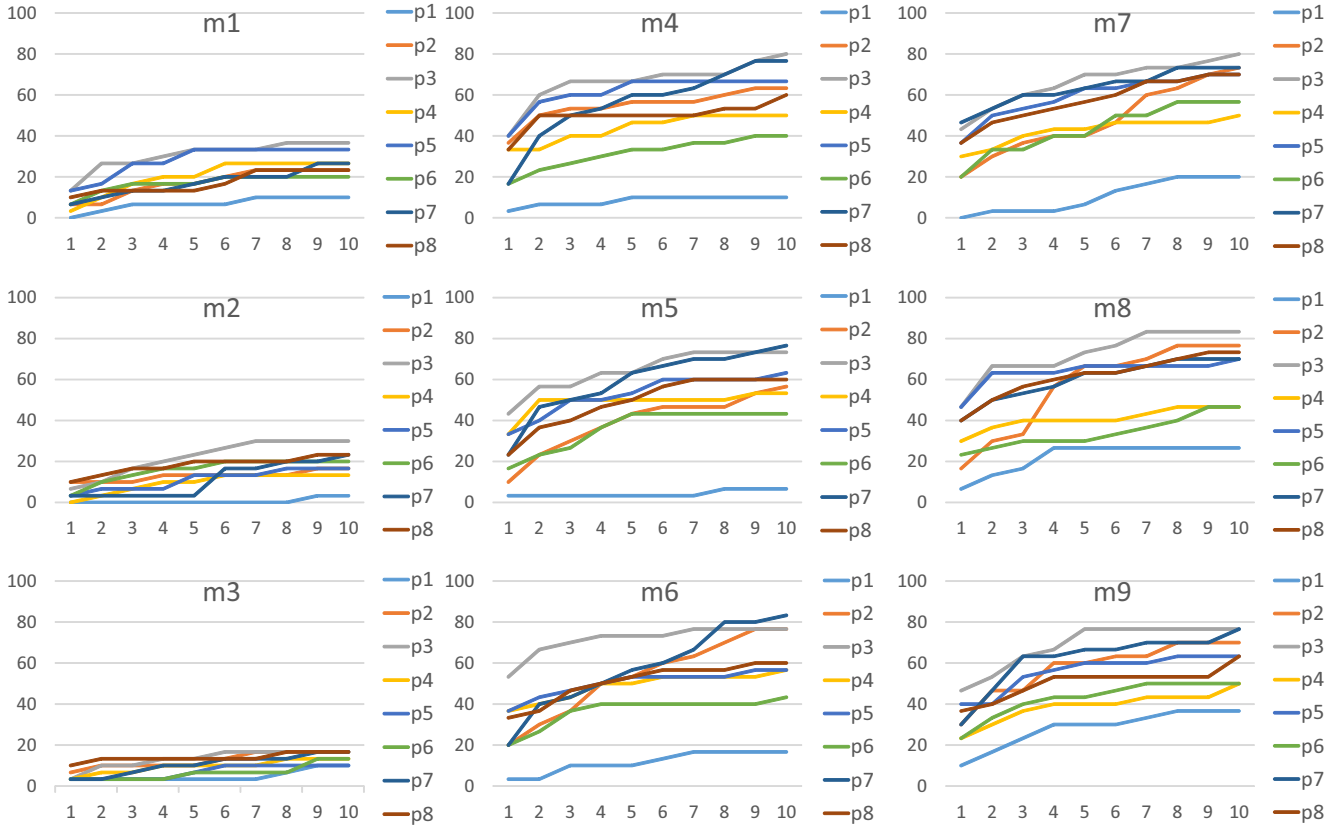


図 9: k に対する各 DNN モデルの精度.

表 3: 上位 10 位までの各モデルと地点に対する精度 (%)

		DNN モデル								
		m1	m2	m3	m4	m5	m6	m7	m8	m9
地点	p1	10.0	3.3	10.0	10.0	6.6	16.6	20.0	26.6	**36.6**
	p2	23.3	16.6	16.6	63.3	56.6	**76.6**	73.3	**76.6**	70.0
	p3	36.6	30.0	16.6	80.0	73.3	76.6	80.0	**83.3**	76.6
	p4	26.6	13.3	16.6	50.0	53.3	**56.6**	50.0	46.6	50.0
	p5	33.3	16.6	10.0	66.6	63.3	56.6	**70.0**	**70.0**	63.3
	p6	20.0	20.0	13.3	40.0	43.3	43.3	**56.6**	46.6	50.0
	p7	26.6	23.3	16.6	76.6	76.6	**83.3**	73.3	70.0	76.6
	p8	23.3	23.3	16.6	60.0	60.0	60.0	70.0	73.3	**76.6**
平均精度		25.0	18.3	13.7	55.8	54.1	58.7	61.7	61.6	60.8
平均精度		19.0			56.2			61.4		

ことで，高精度な平面射影行列を算出できることを示した．

本研究の前提として，対象とした環境は平面に近似できると仮定し，平面射影行列をカメラの位置姿勢として利用した．しかし，この仮定は近景では成り立たない場合も多く，実際に画像の位置合わせが失敗する場合もあった．これらの解決法として，パノラマ画像を参照画像として利用するのではなく，対象空間の 3 次元モデルからレンダリングされた画像群を参照画像することが考えられる．今後の方針として，データベースの画像と入力画像の見かけの違いや，日時による環境の変化に対して頑健になるように手法を改善していく．

参考文献

[1] R. Arroyo, P. F. Alcantarilla, L. M. Bergasa, and E. Romera. Fusion and binarization of cnn features for robust topological localization across seasons. In *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, pages 4656–4663. IEEE, 2016.

[2] Q. Bateux, E. Marchand, J. Leitner, F. Chaumette, and P. Corke. Visual servoing from deep neural networks. In *RSS workshop on New Frontiers for Deep Learning in Robotics*, Boston, Ma, July 2017.

[3] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.

[4] S. Benhimane and E. Malis. Real-time image-based tracking of planes using efficient second-order minimization. In

(a) 入力画像　(b) DNN による視点クラス推定　(c) 画像の位置合わせによる結果

図 10: 失敗例.

Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on, volume 1, pages 943–948. IEEE, 2004.

[5] M. Cummins and P. Newman. Fab-map: Probabilistic localization and mapping in the space of appearance. *The International Journal of Robotics Research*, 27(6):647–665, 2008.

[6] A. Dame and E. Marchand. Accurate real-time tracking using mutual information. In *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*, pages 47–56. IEEE, 2010.

[7] D. DeTone, T. Malisiewicz, and A. Rabinovich. Deep image homography estimation. *arXiv preprint arXiv:1606.03798*, 2016.

[8] G. D. Evangelidis and E. Z. Psarakis. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1858–1865, 2008.

[9] H. Kato and M. Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Augmented Reality, 1999.(IWAR'99) Proceedings. 2nd IEEE and ACM International Workshop on*, pages 85–94. IEEE, 1999.

[10] A. Kendall, M. Grimes, and R. Cipolla. Posenet: A convolutional network for real-time 6-dof camera relocalization. In *Proceedings of the IEEE international conference on computer vision*, pages 2938–2946, 2015.

[11] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[12] D. Koller, G. Klinker, E. Rose, D. Breen, R. Whitaker, and M. Tuceryan. Real-time vision-based camera tracking for augmented reality applications. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 87–94. ACM, 1997.

[13] T. Langlotz, C. Degendorfer, A. Mulloni, G. Schall, G. Reitmayr, and D. Schmalstieg. Robust detection and tracking of annotations for outdoor augmented reality browsing. *Computers & graphics*, 35(4):831–840, 2011.

[14] T. Langlotz, D. Wagner, A. Mulloni, and D. Schmalstieg. Online creation of panoramic augmented reality annotations on mobile phones. *IEEE pervasive computing*, 11(2):56–63, 2012.

[15] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[16] E. Marchand, H. Uchiyama, and F. Spindler. Pose estimation for augmented reality: a hands-on survey. *IEEE transactions on visualization and computer graphics*, 22(12):2633–2651, 2016.

[17] D. Massiceti, A. Krull, E. Brachmann, C. Rother, and P. H. Torr. Random forests versus neural networks—what's best for camera localization? In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pages 5118–5125. IEEE, 2017.

[18] M. Muja and D. G. Lowe. Scalable nearest neighbor algorithms for high dimensional data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(11):2227–2240, 2014.

[19] Y. Nakajima and H. Saito. Robust camera pose estimation by viewpoint classification using deep learning. *Computational Visual Media*, 3(2):189–198, 2017.

[20] R. Richa, R. Sznitman, R. Taylor, and G. Hager. Visual tracking using the sum of conditional variance. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference On*, pages 2953–2958. IEEE, 2011.

[21] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: An efficient alternative to sift or surf. In *Computer Vision (ICCV), 2011 IEEE international conference on*, pages 2564–2571. IEEE, 2011.

[22] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. In *null*, page 1470. IEEE, 2003.

[23] N. Sünderhauf, S. Shirazi, F. Dayoub, B. Upcroft, and M. Milford. On the performance of convnet features for place recognition. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 4297–4304. IEEE, 2015.

[24] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[25] D. Wagner, A. Mulloni, T. Langlotz, and D. Schmalstieg. Real-time panoramic mapping and tracking on mobile phones. In *Virtual Reality Conference (VR), 2010 IEEE*, pages 211–218. IEEE, 2010.

[26] G. Welch and E. Foxlin. Motion tracking: No silver bullet, but a respectable arsenal. *IEEE Computer graphics and Applications*, 22(6):24–38, 2002.

[27] J. Wu, L. Ma, and X. Hu. Delving deeper into convolutional neural networks for camera relocalization. In *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pages 5644–5651. IEEE, 2017.

[28] N. Yazawa, H. Uchiyama, H. Saito, M. Servieres, and G. Moreau. Image based view localization system retrieving from a panorama database by surf. In *Proc. IAPR Conf. on Machine Vision Applications (MVA)*, pages 118–121, 2009.