

一人称視点カメラを用いた情報探索行動の推定

堀内 麻由¹ 川本 一彦² 岡本 一志³

概要: 情報探索行動とは、図書館情報学での用語で、情報収集の行動プロセスを指し、狭義には図書館内での書籍の探索行動を表す。従来、図書館内での行動データの収集と分析は、人手をかけた目視による調査が多く、その自動化が望まれている。そこで本研究では、1人称視点カメラで撮影した図書館内画像から、情報探索行動を自動認識する方法を検討する。認識する行動は、移動する、本を探す、本を読む、の3つを考える。まず、一般物体認識の標準的な方法である bag-of-features による画像表現と SVM による識別によって評価した。さらに、図書館内画像の特色に合わせて、(1) 画像全体の回転は、行動を識別するための手がかりとなることから、局所特徴量の回転不変性を除去し、(2) 書棚と床などの空間配置を考慮するために Spatial Pyramid Matching を導入し、そして (3) 局所特徴点の出現位置の偏りに対応するためにグリッドサンプリングを用いて、評価した。これら3点の導入によって分類率が向上することを示す。

キーワード: 情報探索行動, 一人称視点カメラ, 一般物体認識

Activity Recognition of Information-Seeking Behaviors using First-Person Vision

HORIUCHI MAYU¹ KAWAMOTO KAZUHIKO² OKAMOTO KAZUSHI³

Abstract: 情報探索行動とは、図書館情報学での用語で、情報収集の行動プロセスを指し、狭義には図書館内での書籍の探索行動を表す。従来、図書館内での行動データの収集と分析は、人手をかけた目視による調査が多く、その自動化が望まれている。そこで本研究では、1人称視点カメラで撮影した図書館内画像から、情報探索行動を自動認識する方法を検討する。認識する行動は、移動する、本を探す、本を読む、の3つを考える。まず、一般物体認識の標準的な方法である bag-of-features による画像表現と SVM による識別によって評価した。さらに、図書館内画像の特色に合わせて、(1) 画像全体の回転は、行動を識別するための手がかりとなることから、局所特徴量の回転不変性を除去し、(2) 書棚と床などの空間配置を考慮するために Spatial Pyramid Matching を導入し、そして (3) 局所特徴点の出現位置の偏りに対応するためにグリッドサンプリングを用いて、評価した。これら3点の導入によって分類率が向上することを示す。

Keywords: 情報探索行動, 一人称視点カメラ, 一般物体認識

1. はじめに

情報探索行動 [1] とは、図書館情報学の用語で、人の情報収集の行動プロセスを意味する。近年では、情報技術の

発達により多様化した情報資源をより良く活用するために情報探索行動が研究されている [2][3]。その調査研究の多くは、図書館内の行動データの収集と分析を人手をかけた目視で行っており、自動化が望まれている。

その一方で、コンピュータービジョンの分野では、実世界シーンの画像に含まれる物体を一般的な名称で計算機に認識させる、一般物体認識 [4] が研究されている。

そこで本研究では、情報探索行動を撮影した画像と一般物体認識の手法を用いて、情報探索行動の自動認識を行う方法を検討する。認識する行動は、移動する、本を探

¹ 千葉大学大学院融合科学研究科
Graduate School of Advanced Integration Science, Chiba University

² 千葉大学総合メディア基盤センター
Institute of Media and Information Technology, Chiba University

³ 千葉大学アカデミック・リンク・センター
Academic Link Center, Chiba University

す、本を読む、の3つに限定する。実験データは、実際の情報探索行動の様子を図書館で撮影し、収集する。なお、行動の撮影には一人称視点カメラを用いる。得られた画像群を行動の意味毎にカテゴリ分けし、一般物体認識の標準的な手法である Bag-of-Features[5] による画像表現と Support Vector Machine(SVM)[6] によって識別する。さらに、図書館内画像の特色に合わせて、「回転不変性の除去」、「Spatial Pyramid Matching(SPM)[7]」、「グリッドサンプリング」の3点の改良を行って、識別の精度が向上することを確認する。

本稿では、まず2. 情報探索行動の推定方法の概要を述べる。次に3. 情報探索行動の推定の精度を上げるための改良方法を提案する。4. で、図書館で収集した実データに対して2. と3. で述べた方法を適用し、分類率を評価する。

2. 一人称視点カメラと一般物体認識による図書館内画像の分類

2.1 一人称視点カメラ

図書館内は、室内に書架が立ち並んだ閉塞的な環境である。そのため死角が生まれやすく、定点のカメラによる撮影では被験者の行動を仔細に記録することは困難である。情報探索行動に関する先行研究 [8] には、実験者がハンドカメラを持って被験者を追従する方法をとっている場合もあるが、これは被験者と同じ人数の撮影者が必要であるため大量のデータ収集が難しい。

そこで本研究では、小型ウェアラブルカメラを使用して画像を収集する。被験者が一人一台ずつカメラを装着することで、図書館内でも常に被験者の行動を記録することができる。ハンズフリーであるため、被験者の行動を阻害しないことも利点の一つである。

また、本研究の認識対象は情報を得る行動プロセスであり、被験者が注目している対象や被験者の手元の画像が認識の重要な手がかりとなる。そのため、ウェアラブルカメラの中でも、被験者の視界に近い映像を撮影できる頭部装着型の一人称視点カメラを使用する。

2.2 分類カテゴリ

情報探索行動は情報を得るための行動プロセス全般を指す用語であるが、本研究では、認識する行動を (1)Moving: 移動する、(2)Searching: 本を探す、(3)Browsing: 本を読む、の3つに限定する。

さらに、この3行動を一人称視点カメラで撮影した画像を、aisle と walk、slant と face、pickup と browse の合計6カテゴリに分けて一般物体認識の手法による機械学習を行う。6カテゴリの定義を表1に示す。また、6カテゴリの画像例を図1に示す。図書館内で撮影した画像は常に書架が写り込むため、6カテゴリは画像内の物体の組合せや

表1 6カテゴリの定義

行動	カテゴリ	カテゴリの意味	写っている物体
Moving	aisle	通路にいる	書架の側面
	walk	書架の間を移動する	書架+画面中央に床
Searching	slant	書架に視線を向ける	書架+画面端に床
	face	手に取る本を探す	画面全面に書架
Browsing	pickup	本を取る/戻す	真正面の書架+手
	browse	本を開いて読む	開いた本

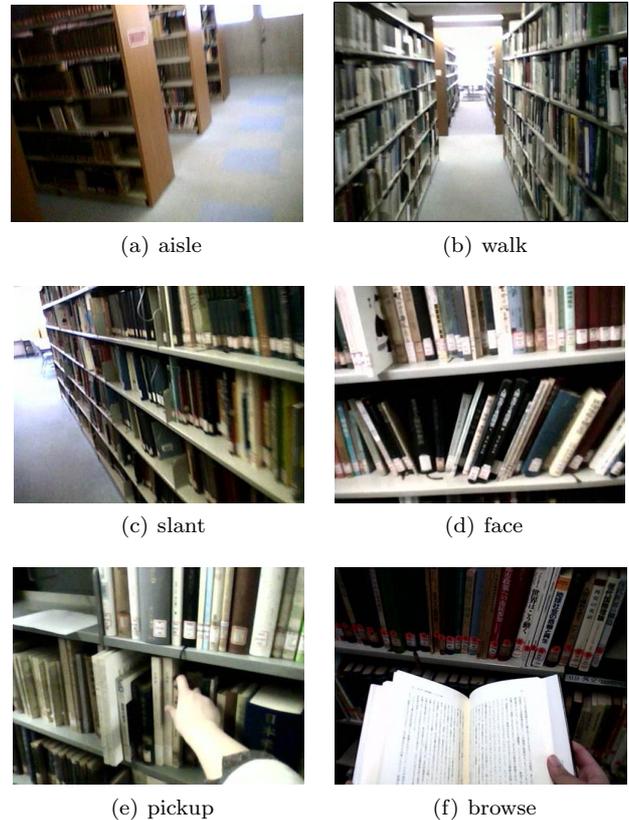


図1 6カテゴリの画像例

状態によって判別することができるように設定した、

2.3 一般物体認識

一般物体認識とは、実世界シーンを撮影した画像を入力として、その画像に撮影された物体が何であるかを計算機に認識させることをいう。

本研究では、一般物体認識の手法を用いて入力画像中の「手」「本」「書架」等の物体を認識することで、2.2で定義した6カテゴリの分類を行う。例えば、画像中に「開いた本」を認識したとき、その画像はカテゴリ browse であり、本を読んでいるシーンであると推定できる。

一般物体認識の標準的な手順は大きく分けて3段階である。ここでは、手順を追いながら、本研究で使用する標準的な手法について簡単に述べる。

(1) 特徴量の抽出: 画像の輝度勾配をもとに局所特徴量を記述する。Loweによって提案された SIFT(Scale Invariant Features Transform)[9] や、それを高速化し

た SURF(Speeded Up Robust Features)[10] が広く用いられている。SIFT/SURF は、特徴点検出と特徴量記述の 2 つのアルゴリズムから成り、スケール変換・回転・照明変化に対して頑健であるという特性を持つ。本研究では、局所特徴量を 128 次元で記述する拡張型の SURF を用いる。

(2) 画像表現の変換：局所特徴量を別の画像表現に変換する。BoF が広く用いられている。BoF は、クラスタリング手法を用いて、学習用画像群から得られた数千～数百万個の局所特徴量を任意の次元のヒストグラムに量子化する。ただし、複数の画像から得られた局所特徴点をまとめてクラスタリングしてデータの量子化するため、局所特徴点の画像上の位置情報は完全に無視される。本研究では、クラスタリング手法には k-means++[11] を適用し、ヒストグラムの次元は 500, 1000, 3000 の 3 通りを設定する。

BoF は画像一枚の局所特徴量を 1 つの位置情報を考慮したい場合は、SPM のような位置情報のための新たな手法を導入するか、BoF 以外の画像表現を用いる必要がある。

(3) 識別器による照合：生成確立モデルに基づく手法と判別モデルに基づく 2 つの方法がある。判別モデルの SVM は一般的に高い性能を示すため、BoF と併用して一般物体認識システムを実現することが多い。カーネルトリックにより単純な計算で非線形データの分類が可能だが、カーネルの選択によって精度が左右される。ただし、本来の SVM は 2 クラス分類器であるため、複数クラス分類を行う場合は何らの工夫が必要となる。今回は 6 クラス問題なので、「1 クラス-vs-その他」の分類器を 6 個作った。2.3 で述べる SPM を用いないとき、カーネルは RBF をカーネルを使用する。

3. 一般物体認識による情報探索行動の推定

一般物体認識の手法は画像中の「物体」を識別するものであり、本研究で扱うような「行動」を識別するために作られた手法ではない。したがって、一般物体認識の基本的な手法をそのまま図書館画像へ適用するだけでは、高い認識結果を得ることはできない。そこで、図書館内画像を分類するにあたって問題となる点を挙げ、それに対処する形で情報探索行動の推定精度を上げるための改良方法を提案する。

3.1 局所特徴点の回転不変性の除去

通常一般物体認識の課題では、物体の見え方が変化した場合でも、同一の名称の物体であると認識しなければならない。そのために SIFT/SURF が持っている特性の一つに回転不変性があり、大まかに次の手順の特徴量抽出方法

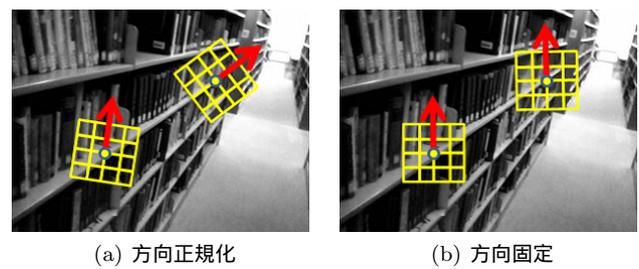


図 2 特徴点抽出の方向

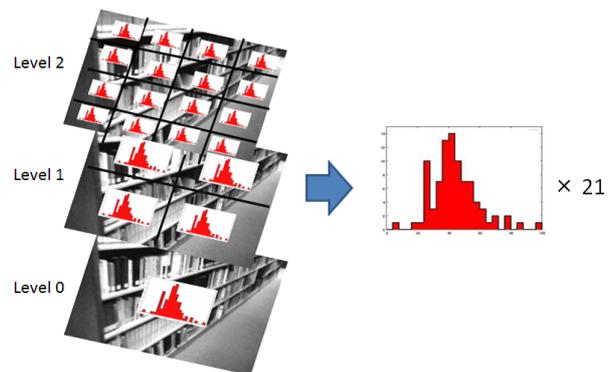


図 3 SPM のイメージ図

で実現されている。

- (1) 局所特徴量を抽出する座標を決定し、その点を中心にして周囲の勾配を算出。
- (2) 閾値よりも頻度の高い勾配方向を検出し、その局所特徴点のオリエンテーションとする(図 2(a)の赤矢印)。
- (3) 検出したオリエンテーションに合わせて計算領域(図 2(a)の黄ブロック)を回転させ、局所特徴量を方向で正規化する。
- (4) 計算領域内で輝度勾配を算出し、局所特徴量を記述する。

このように、SIFT/SURF は、局所特徴量を記述する方向を正規化することで、物体が回転しても同一の特徴量を記述することができる。

しかし、本研究では回転した物体が写る画像を別の行動として分類する。例えば図 1(c) と図 1(d) に示したように、書架が真正面に写っていれば face であるが、書架が斜めに写った場合には slant である。ところが回転不変な特徴量抽出では、どちらも書架が写っている画像として同一とみなされ、face と slant の区別が困難となる。

そこで本研究では、回転した物体を見分けるため、あえて SIFT/SURF から回転不変性を取り除く。上記手順の(2),(3)は行わず、すべての特徴点についてパッチの方向を一定に固定して局所特徴量の抽出を行う(図 2(b))。

3.2 空間配置を考慮した画像表現

一人称視点カメラで図書館内を撮影した場合、画像の大多数に書架が写りこむが、行動によって書架が写る画面上

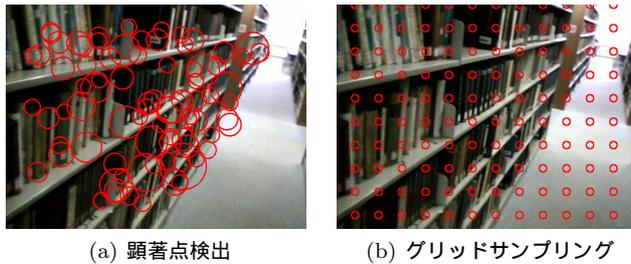


図 4 異なる行動の 2 つの画像

の位置は異なる．例えば，walk は画面の両端に書架が写り，slant は画面の左右どちらか一方に書架が写る．情報探索行動の認識において，物体の画面上の空間配置は重要な手がかりになり得る．

しかし，一般物体認識で広く用いられる BoF による画像表現は，局所特徴点の位置情報は無視して，1 枚の画像を 1 つのヒストグラムで表現する手法である．

そこで，位置情報を考慮するために SPM を導入する．SPM の手順を以下に示す．

- (1) 画像を分割してピラミッドを構築する．level0 は分割無し，level1 は 4 分割，level2 は 16 分割・・・と続き，任意の層を設定できる．今回は level0 ~ level2 を設定する．
- (2) それぞれのパッチから BoF のヒストグラムを構築する．つまり，level2 まで使用する場合は，合計で 21 個のヒストグラムで 1 つの画像を表現する (図 3)．
- (3) ピラミッドのレベルに応じて重みをつけて類似度を計算．

類似度は

$$K = \frac{1}{2^L} I_0 + \sum_{l=1}^L \frac{1}{2^{L-l+1}} I_l \quad (1)$$

で計算され，ピラミッドのレベルが低いほど重み付けで与えられるペナルティは大きい．これは，画像の分割が少ないほど位置情報が曖昧になるためである．また，ヒストグラムインターセクション I_l は

$$I_l = \sum_i \min(H_X^l(i), H_Y^l(i)) \quad (2)$$

で表され，画像 X と画像 Y を比較して小さい方の画素値を返す．ここで，L はピラミッドの最大レベルをあらわし， H_X^l は画像 X のレベル l の i 番目のパッチである．

画像をパッチごとの複数のヒストグラムで表現することにより，局所特徴点の大まかな位置情報を考慮することができる．

識別器として SVM を併用する場合，ヒストグラムインターセクションはカーネルとして用いられる．

3.3 サンプリング方法の改善

SIFT/SURF では特徴点抽出のとき，多重解像度空間で

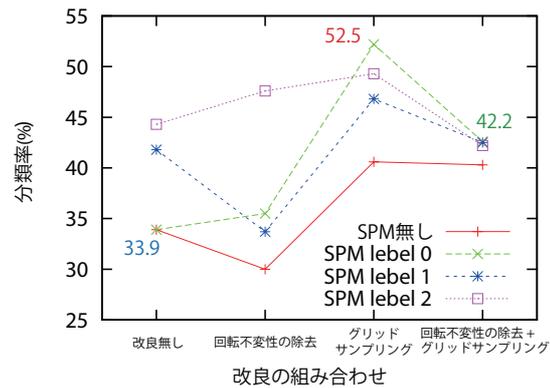


図 5 クラスタ数 500 のときの分類率

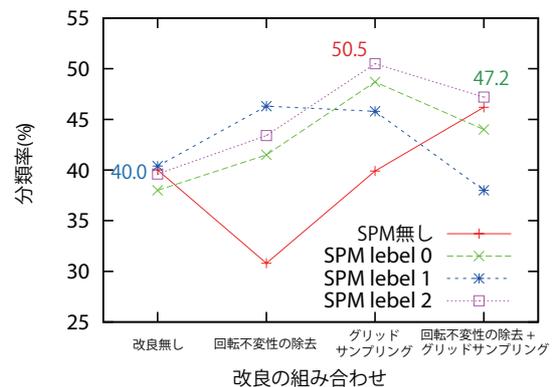


図 6 クラスタ数 1000 のときの分類率

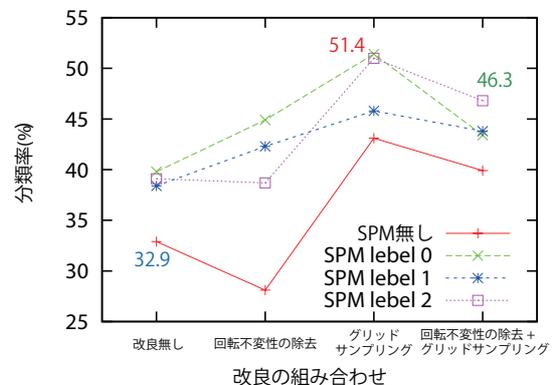


図 7 クラスタ数 3000 のときの分類率

明度勾配の大きい箇所を局所特徴点とする方法をとっており，これを顕著点検出という．しかし，図 4(a) のように，図書館内画像の床や窓の部分は，特徴となる輝度勾配が少ないため顕著点検出では局所特徴点が現れ難い，そのために画面上の局所特徴点の出現位置に偏りが生まれる．

しかし，前述の SPM はすべてのパッチで特徴点を検出できることが前提になっている．特徴点一つもないパッチが存在する場合，式 (2) のヒストグラムインターセクション I は必ず $I = 0$ となり，画像の類似度を正確に算出することができない．

そこで，10px 幅でグリッドサンプリングを行い，分割したどのパッチでも必ず局所特徴量を抽出する．ただし特徴

表 2 改良無しの混同表 (クラスタ数 3000)
 (a) 6 カテゴリ 分類率: 32.9% (b) 3 行動 分類率: 45.0%

	aisle	walk	slant	face	pickup	browse		Moving	Searching	Browsing
aisle	0.90		0.03	0.03		0.03	Moving	0.57	0.39	0.05
walk	0.11	0.28	0.22	0.11	0.11	0.17	Searching	0.24	0.47	0.31
slant	0.12	0.24	0.12	0.24	0.18	0.12	Browsing	0.26	0.43	0.32
face	0.03	0.11	0.03	0.23	0.09	0.51				
pickup	0.05		0.03	0.22	0.05	0.65				
browse	0.13	0.04		0.04	0.33	0.46				

表 3 回転不変性の除去, SPM レベル 2, グリッドサンプリングを適用したときの混同表 (クラスタ数 3000)

	aisle	walk	slant	face	pickup	browse		Moving	Searching	Browsing
aisle	0.83		0.10		0.07		Moving	0.69	0.27	0.03
walk	0.33	0.22	0.44				Searching	0.31	0.48	0.22
slant	0.53		0.24	0.06	0.18		Browsing	0.06	0.16	0.78
face	0.06	0.03	0.11	0.54	0.26					
pickup	0.03		0.03	0.22	0.73					
browse	0.08		0.08		0.58	0.25				

量抽出の際のスケールは 30 で固定とする。

4. 大学図書館内画像データを用いた識別性能の評価

4.1 実験データ

千葉大学附属図書館にて実験データとなる画像を収集した。使用したカメラは Looxcie 社製 Looxcie で、耳に掛けるタイプのウェアラブルカメラである。

一人称視点カメラを装着した状態で自由に書籍の探索行動を行い、一連の動作を動画に記録する。その動画から 3fps の間隔で静止画を抽出し、各画像に真値となる 6 カテゴリのラベルを与える。

収集した画像の中から、学習用画像として、6 カテゴリに該当するものを選別して計 600 枚 (各カテゴリ 100 枚) の画像群を用意。さらに、質問用画像として、「Walking Searching Browsing」の順に行動している時系列の静止画像 160 枚を用意した。

4.2 実験条件

3 節で述べた「回転不変性の除去」「SPM」「グリッドサンプリング」の 3 点の改良を網羅的に組み合わせ、分類率の向上を検証する。

- (1) 特徴点の検出: 顕著点検出 / グリッドサンプリング
- (2) 特徴量記述: 回転不変性有り / 回転不変性無し
- (3) BoF による画像表現: クラスタ数 500/1000/3000
- (4) SPM: レベル 0 / レベル 1 / レベル 2
- (5) SVM: RBF カーネル (SPM レベル 0 のとき) / ヒストグラムカーネル (SPM レベル 0~2 のとき)

4.3 結果

BoF のクラスタ数が 500, 1000, 3000 のときの分類結果のグラフをそれぞれ図 5, 図 6, 図 7 に示す。

グラフ中の緑色の値は、回転不変性の除去, SPM レベル 2, グリッドサンプリング, の 3 つをすべて適用した場合の分類率である。一方、青色の値は、情報探索行動のための改良を施していない場合の分類率である。図 5~ 図 7 のいずれの場合でも、改良無しの場合より、3 つの改良を適用した場合のほうが高い分類率を示した。これにより、一般物体認識の手法を用いた図書館内画像の分類において、回転不変性の除去, SPM, グリッドサンプリングが有効であることが確認できた。最も高い分類率を示したのはグラフ上の赤の値で、回転不変性の除去を行わない場合であった。

また、BoF のクラスタ数 3000 のときの混同表の一例を表 2, 表 3, 表 4 に示す。表 2 は改良無し, 表 3 は回転不変性の除去, SPM レベル 2, グリッドサンプリングの 3 つ改良を適用した場合、そして表 4 はクラスタ数 3000 のとき最良の分類率を示した, SPM レベル 2 とグリッドサンプリングを適用した場合である。(a) は画像の 6 カテゴリ混同表である。そして (b) は (a) から算出した 3 カテゴリの混同表であり、同じ行動に属するカテゴリ同士の誤りを許容した値となっている。(a),(b) のどちらも、縦軸が入力画像のカテゴリであり、横軸が出力となっている。

4.4 考察

本研究の意図は、回転不変性の除去, SPM レベル 2, グリッドサンプリングをすべて適用したときに最も高い分類率を示すことであった。しかし実際には、局所特徴量の回転不変性を残した場合の方が高い分類率を示した。グリッ

表 4 SPM レベル 2, グリッドサンプリングを適用したときの混同表 (クラス数 3000)
(a) 6 カテゴリ 分類率: 51.4% (b) 3 行動 分類率: 67.0%

	aisle	walk	slant	face	pickup	browse		Moving	Searching	Browsing
aisle	0.48	0.38	0.07		0.07		Moving	0.57	0.37	0.06
walk	0.22	0.06	0.56	0.11	0.06		Searching	0.21	0.72	0.07
slant	0.35	0.06	0.29	0.24	0.06		Browsing	0.01	0.26	0.72
face				0.91	0.09					
pickup		0.03		0.49	0.46	0.03				
browse				0.04	0.08	0.88				

ドサンプリングと回転不変性の除去は、どちらも特徴量抽出の段階の処理であり、これらを併用すると、記述される特徴量は一定間隔で抽出した単純な輝度勾配になる。このため、物体の特徴を大まかにしか捉えられなくなり、回転不変性を残した場合よりも分類率が下がってしまったと考えられる。回転物体を区別する方法について再考が必要である。

次に、グリッドサンプリングの効果に注目する。グリッドサンプリングは SPM の補助のために導入した手法であったが、SPM を用いない場合でもグリッドサンプリングを行うことで分類率が向上することを確認した。図書館内画像に写る物体は、書架、床、手、窓など限られており、本研究ではその限られた手がかりから一般物体認識を行う。しかし局所特徴点の顕著点検出を行う場合、3.3 で述べたように、書架に局所特徴点があり、床では検出されないという状態が起きた。画像に床が写っていても、局所特徴量が得られなければ存在していないことに等しいため、認識率の向上に寄与することができない。一方、グリッドサンプリングは一定間隔で局所特徴点を指定する。そのため図書館内画像に写るすべての物体から特徴量を抽出でき、認識を行うための材料が増えたことが分類率の向上につながったと考えられる。

また、SPM の導入によっても分類率の向上が確認できた。ただし、画像ピラミッドを構築しないため位置情報を考慮していない SPM レベル 0 の場合でも、SPM 無しより高い分類率を示している。これは、図書館内画像の分類には RBF カーネルよりもヒストグラムインターセクションの方が適していたからだと考えられる。SPM レベル 1 と SPM レベル 2 を比較すると、概ねでレベル 2 のほうが良い分類率を示していることから、局所特徴点の位置情報の考慮も認識率の向上に貢献していると言える。

しかし、表 2(a)、表 3(a)、表 4(a) に注目すると、walk は改良無しの場合のほうが良い分類率になっており、局所特徴点の位置情報の考慮による効果が得られなかった。原因としては、aisle などの画面上の物体の位置に規則性を持たないカテゴリの存在が邪魔をしたことや、学習量の不足が考えられる。

5. おわりに

情報探索行動の自動認識を行うため、一人称視点カメラで撮影した図書館内画像に一般物体認識の手法を用いる方法を検討した。認識対象は、移動する、本を探す、本を読む、の 3 行動に限定し、一般物体認識の手法で認識を行うために、情報探索行動の様子を撮影した画像を分類する 6 カテゴリを定義した。BoF による画像表現と SVM による識別に加え、(1) 回転不変性の除去、(2) SPM、(3) グリッドサンプリング、を用いることで情報探索行動を推定を実現した。分類率の評価実験を行い、上記 3 点の改良を加えることで分類率が向上することを確認した。

意図した通りの効果が出なかった箇所に関しては、今後の再考と再実験を要する。具体的には、回転物体を区別する方法、SPM による位置情報の考慮を効果的に反映させるためのカテゴリ分け、さらに大規模なデータでの再実験である。

また、今回は一般物体認識の手法を用いるために静止画を用いたが、今後は時系列データの利用や加速度センサーなど、何らかの方法で動きを考慮した情報探索行動の推定を行いたい。

謝辞 本研究は科研費 (23700189) の助成を受けたものである。データ収集に関してご協力いただいた千葉大学附属図書館に感謝します。

参考文献

- [1] 田村俊作編. 情報探索と情報利用. 勁草書房, 2001.
- [2] 種市淳子, 寺井仁, 逸村裕.
- [3] 寺井仁. 大学図書館における情報探索活動に関する研究: われわれはいかに異なる情報源を活用しているのか? 名古屋大学附属図書館研究年報, No. 64, pp. 177-192, 2010.
- [4] 柳井啓司. 一般物体認識の現状と今後. 情報処理学会論文誌: コンピュータビジョンとイメージメディア, Vol. 48, SIG 16 (CVIM 19), pp. 1-24, 2007
- [5] J.Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos, it Proc. ICCV, Vol. 2, pp. 1470-1477, 2003.
- [6] 前田英作. 痛快! サポートベクトルマシン ~ 古くて新しいパターン認識手法 -. 情報処理学会誌, Vol.42, No.7, pp.676-683, 2001.
- [7] Eric Nowak. *Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories.* in

Proc. ECCV, 2006.

- [8] 市村光広. 短期大学図書館における情報探索行動の実験的検討. 三田図書館・情報学会誌, Vol. 37, pp. 40–45, 2009.
- [9] D.F.Lowe. *Object recognition from local scale invariant features*. Proc. of IEEE International Conference on Computer Vision (ICCV), pp. 1150–1157, 1999.
- [10] Herbert Bay. *SURF: Speeded-Up Robust Features*. In Ninth European Conference on Computer Vision, 2006.
- [11] D. Arthur and S. Vassilvitskii. k-means++: the advantages of careful seeding. *Proc. 18th ACM-SIAM Sympos. Discrete Algorithms*, pp. 1027–1035, 2007.