Vol.2014-CVIM-192 No.15
2014/5/15

一人称視点画像を用いた文書画像の分類

志賀 優毅[1,a)]　内海 ゆづ子[1,b)]　岩村 雅一[1,c)]　カイ クンツェ[1,d)]　黄瀬 浩一[1,e)]

概要：読書の時間，頻度，あるいはどのような種類の本を読んでいるかなどの情報を把握することは，自己管理や学習到達度の目安として役立つため大変重要であると考えられている．しかし，それらの情報の手動記録には多くの時間を消費してしまう．読書習慣の情報を自動的に記録，分析するために，我々はウェアラブルカメラで撮影された一人称視点の文書画像を分類し，ユーザが読んでいる文書を推定する問題を提案する．文書画像の分類は一種の一般物体認識といえるため，一般物体認識で広く用いられている特徴量を使用し，データベースに登録されていない文書画像を各クラスに分類する．Bag-of-Features を用いた手法で実験した結果，81%の正解率を得ることができた．

1. はじめに

古くより，読書の量と学力や知識には相関があると考えられてきた．福沢諭吉は著書である学問のすすめの中で，「読書は学問の術なり、学問は事をなすの術なり」と述べ，学問に読書は欠かせないものであると説いている．近年の研究でも，読書活動と学習活動に強い相関，学習活動と教科学力には強い相関があることが示されている [1]．このように，読書とは知識を得るための重要な手段であり，読書の質と量を調べることによって，人の学力や知識量を推定することができると考えられる．そのため，読書活動を記録し，その内容を把握することは非常に重要である．特に，多くの人々の読書活動について長期間にわたる詳細な記録を行うことができれば，これまでに得ることのできなかった知見が得られる可能性がある．

しかし，従来の読書活動の記録方法は必ずしも良い方法であるとはいえない．例えば，全国学校図書館協議会が毎日新聞社と共同で行っている読書調査[*1]では，小中高生が1ヶ月で読んだ本の数を記録している．しかし，漫画や小説のようなカテゴリの違う本を同じ1冊としてカウントするのは適切であるとはいえず，読書活動の把握と可視化を行うためには，文書のカテゴリのような具体的な情報が必要であると考えられる．一方，ブクログ[*2]と呼ばれるウェブサービスでは，読んだ本のタイトルや感想を記録することで，ユーザの読書活動の振り返りやユーザ間での情報共有を可能にし，読書を促進する．ブクログでは正確な情報が記録できるという利点があるが，ユーザは情報の入力を手動で行わなければならず，あらかじめ用意されているデータベースに存在する文書しか登録することができない．もし，ユーザが読んだ文書を自動的に記録することができれば，より多くの情報を簡単に記録できるようになると考えられる．

読書活動の自動的な記録を目的とした具体的な研究の例として，脳波計 [2] やアイトラッカ [3] を用いてユーザの読書行動を推定する研究などがある．しかし，脳波計やアイトラッカは比較的高価であり，操作方法も複雑である．実用を考えると，より手軽で安価なデバイスである装着型の装置で読書活動の自動的な記録が行える方が良い．

そこで本研究では，ユーザの読んだ文書がデータベースに存在しないものであっても認識し，自動的に記録するシステムの実現のため，一人称視点の文書画像をウェアラブルカメラを用いて撮影し，漫画や小説などのクラスに分類する問題を提案する．このシステムの実現によって，ユーザは読書行動の記録，把握，改善をより簡単に行うことができるようになると考えられる．この問題に対する評価を行うため，一般物体認識で広く用いられている Bag-of-Features を用いることで，学習用画像に含まれていない文書の分類を試みる．ウェアラブルカメラで撮影された文書画像を分類する問題を扱ったのは本研究が初めてであるため，実験用に新たなデータセットを作成した．以

[1] 大阪府立大学大学院工学研究科
大阪府堺市中区学園町 1–1
Graduate School of Engineering, Osaka Prefecture University 1–1, Gakuencho, Naka, Sakai, Osaka 599–8531, Japan
[a)] shiga@m.cs.osakafu-u.ac.jp
[b)] yuzuko@cs.osakafu-u.ac.jp
[c)] masa@cs.osakafu-u.ac.jp
[d)] kunze@m.cs.osakafu-u.ac.jp
[e)] kise@cs.osakafu-u.ac.jp

[*1] http://www.j-sla.or.jp/material/research/54-1.html

[*2] http://booklog.jp/

降 2 章で関連研究について述べ，3 章で本研究で作成したデータセットについて説明する．4 章で本研究で用いる手法を説明し，5 章で実験とその結果について考察する．そして，6 章で本研究についてまとめる．

2. 関連研究

ウェアラブルカメラの小型化や低価格化に伴って，多くの研究者が様々な研究を行っている．Starner らはウェアラブルカメラから得られた画像に対して顔認識と物体認識を用いて AR システムを実装した [4]．Fathi ら [5] と Ren ら [6] は行動認識のためにウェアラブルカメラを用いてユーザが操作している物体の認識を行った．しかし，これらの既存技術は本研究で目的としているような，読書習慣の把握とは直接結びつかない．

読書習慣の把握に関連する研究として，我々はリーディングライフログを提案している [7], [8]．これは，ユーザが読んだ単語や文章の情報を蓄積することによって，ユーザの行動に新たな価値を生み出す試みである．具体的な研究事例としては，モバイルアイトラッカーを装着した人が読んだ文字数を数えることで読書量を推定する万語計 [9] や，読者の視線の動きから文書を推定する研究 [10]，読者の視線から文書の理解度を推定する研究 [11]，市販の脳波計を用いてを読書行動を認識する研究 [2] などがある．しかし，これらの研究では，ユーザの行動理解のためにモバイルアイトラッカーや脳波計などの比較的高価である特殊なデバイスを使用している．本研究のようにウェアラブルカメラのみを用いたユーザの読書行動の推定は，技術的に解決すべき課題はあるが，デバイスが安価であることから実用面でのハードルは相対的に低いと考えられる．

ウェアラブルカメラで撮影された文書画像を扱った研究は我々の知る限りでは，まだ行われていない．通常のカメラで撮影した文書画像を扱った研究として，Locally Likely Arrangement Hashing (LLAH) [12] というハッシュに基づく検索法が提案されている．LLAH はデータベース中の文書画像から検索質問画像と一致する画像を高速に検索することができる．LLAH は高速かつ高精度で，高い頑健性を持つという特長がある．しかし，データベースに存在する文書しか認識することができず，本研究で想定しているようなデータベースに存在しない文書も分類を行う読書行動の記録には適さない．

3. データセット

実験用にデータセットを作成した．データセットの撮影にウェアラブルカメラを用いた．ウェアラブルカメラは Panasonic HX-A100 を使用した．動画のフレームレートは 60 fps であり，解像度は 1920×1080 pixels である．データセット文書のクラスは，漫画，小説，教科書，新聞，雑誌，論文の 6 種類を用意し，各クラスに 5 冊，計 30 冊の文書を用意した．20 代の男性 5 人が各クラスにつき 1 冊の文書を読み，ウェアラブルカメラで撮影した．撮影した各動画の長さは 10 分間であり，同じページが写ったフレームが選択されないように，1500 フレームごとに静止画を抽出した．抽出した静止画に対して Bai らの手法 [13] を用いて半自動的に文書画像の領域を切り出した．この手法では最初のフレームを手動で切り出すことにより，次からのフレームの切り出しを補助することができる．切り出した文書画像の例を図 1 に示す．

4. 分類手法

本節では本研究での文書画像の分類手法について説明する．文書画像のクラスを認識することは，一種の物体認識と捉えることが可能である．その中でも，文書画像の分類はデータベース中に含まれる物体と同一の物体を認識する特定物体認識よりも，画像中の物体を一般的な名称で認識することを目的とした一般物体認識に近いと考えられる．そこで，分類手法として一般物体認識でよく用いられる手法を使用する．手法の概要を図 2 に示す．まず，ウェアラブルカメラを用いて一人称視点の動画を撮影する．次に得られた動画から文書の領域のみを切り出す．そして，得られた文書画像から局所特徴量を抽出する．得られた局所特徴量を Bag-of-Features モデルを用いてヒストグラムに変換し，ヒストグラムを識別器に通すことで分類結果を得る．

4.1 特徴量

本節では実験で用いる手法で使用する 4 種類の特徴量について説明する．分類手法では，特徴点の選択手法を 2 種類，Bag-of-Features モデルに基づくヒストグラムの作成手法を 2 種類用いる．

4.1.1 特徴点の選択手法

文書画像の局所部分から情報を得るために，局所特徴量を用いる．局所特徴量は，画像の一部分から抽出される特徴量であり，画像認識や，特定物体認識などで広く用いられている [14], [15]．局所特徴量は特徴点の選択と特徴量の記述の 2 段階を経て抽出される．実験では特徴点の選択に SIFT の特徴点検出手法 [16]，あるいは dense sampling を用いる．

SIFT の特徴点検出手法では，複数の Difference-of-Gaussian (DoG) 画像を用いて極値の探索を行うことで，特徴点とスケールを検出している．DoG による特徴点検出では，DoG 画像を複数枚作成し，隣接する 3 枚の DoG 画像において，注目画素の周りの 26 近傍を比較し，注目画素が極値となる場所を特徴点候補として検出する．そこからノイズの影響を除去するために，DoG の出力値より求められるヘッセ行列を利用してエッジ上の点を削除し，コントラストの閾値処理を適用し特徴点を絞り込んでいる．さらに，得られた特徴点の周りの輝度勾配からオリエンテー

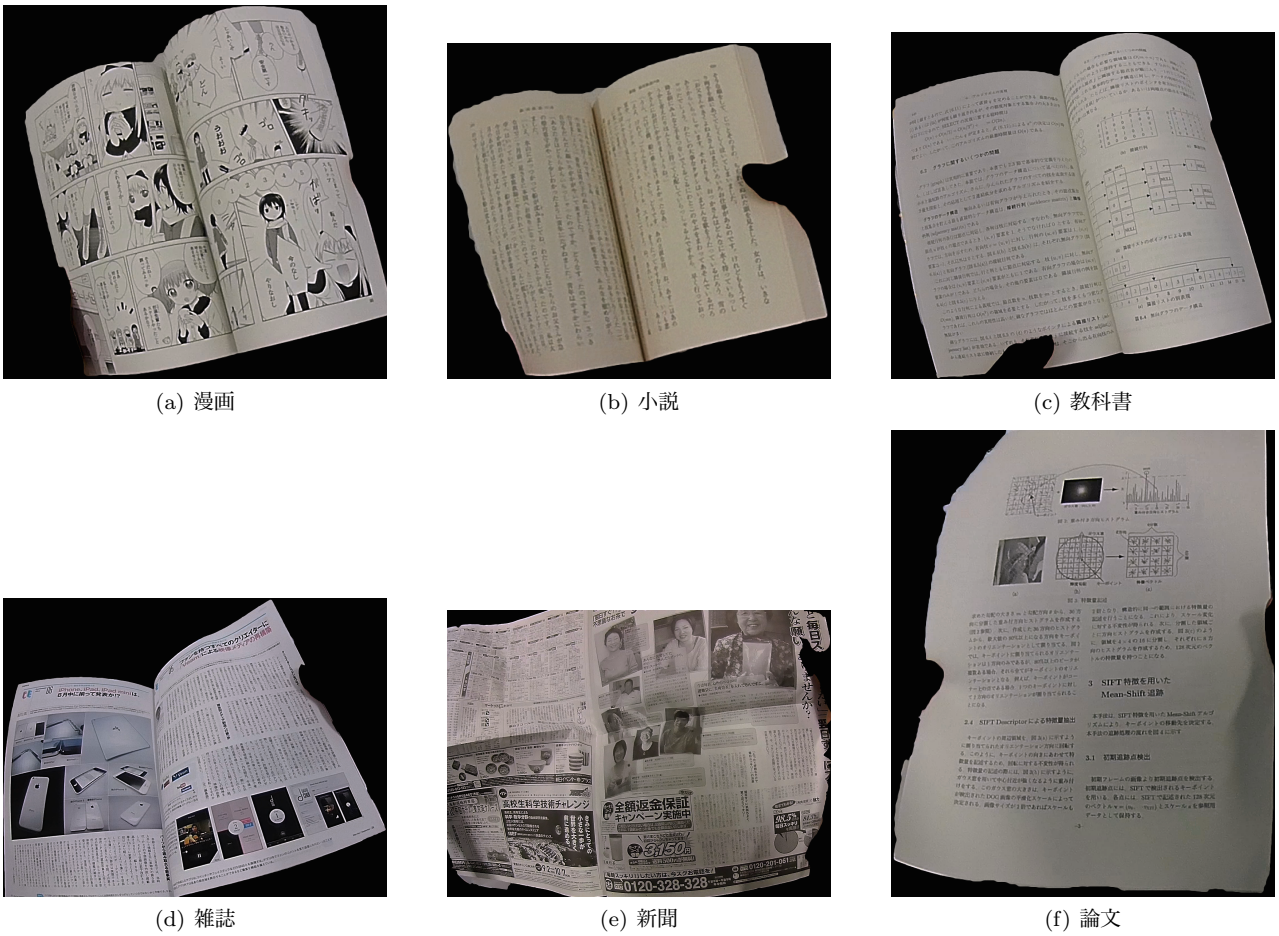


図 1 データセット中の文書画像の例

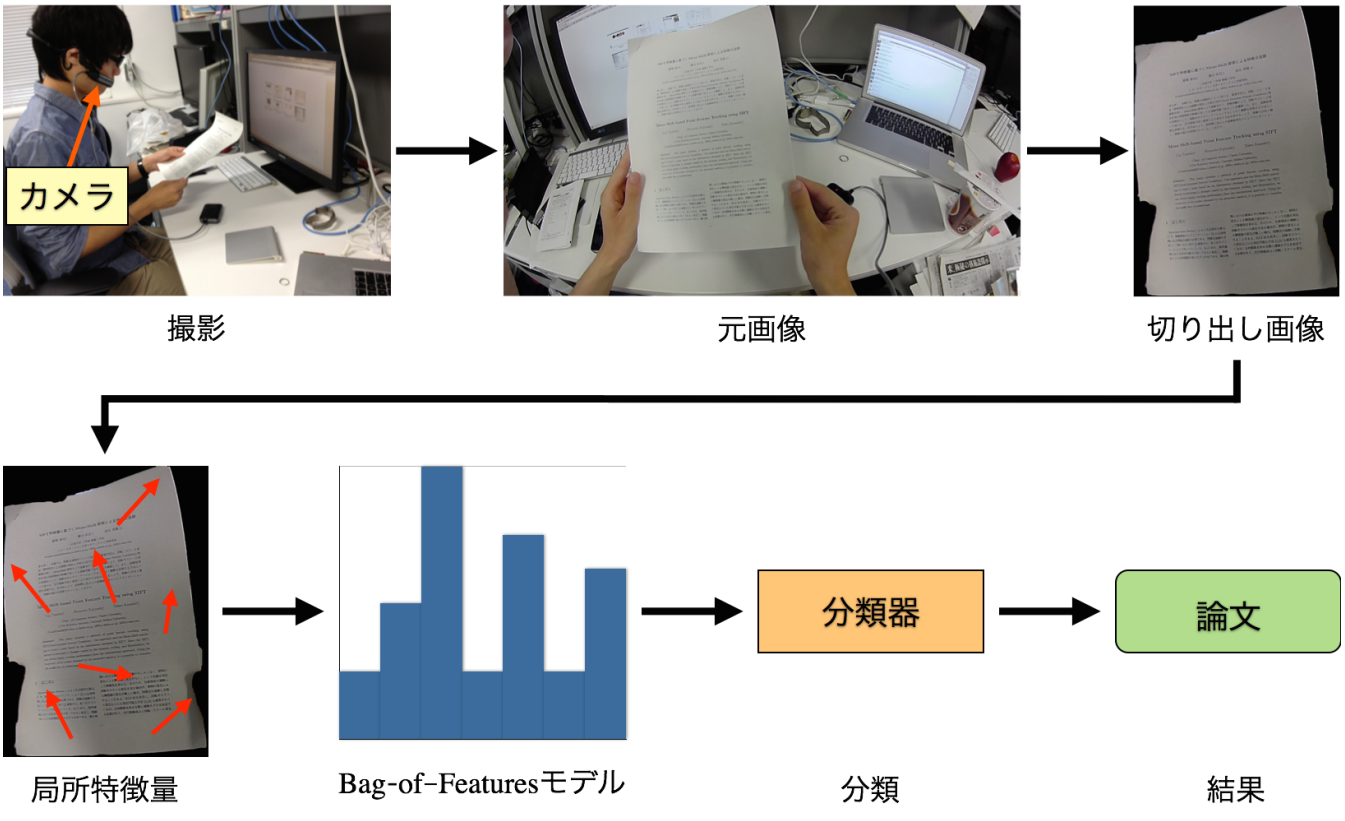
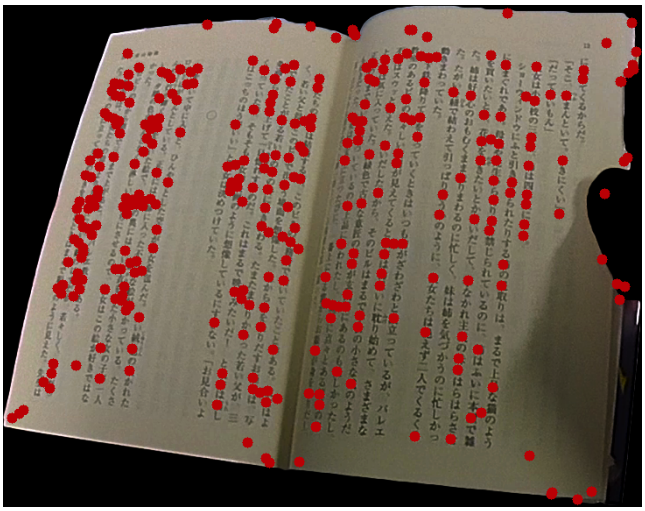
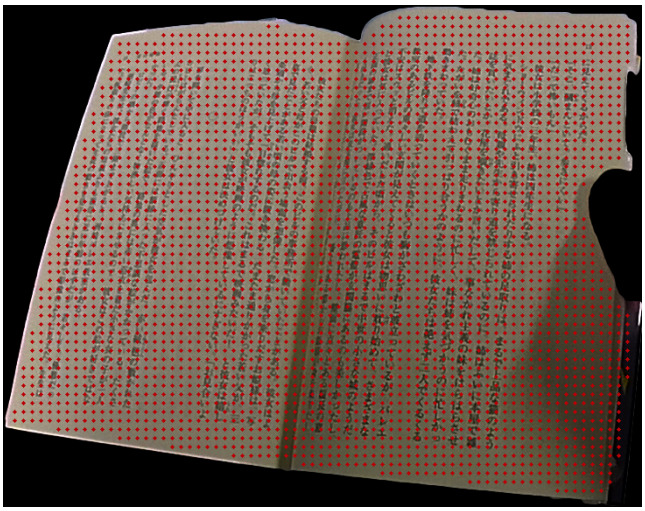


図 2 手法の概要



(a) SIFT の特徴点検出手法



(b) dense sampling

図 3 特徴点の例

ションを一意に求めることによって画像の回転に不変な特徴領域を選択する．それに対して dense sampling は，一定の間隔，スケール，オリエンテーションで特徴点の決定を行う方法である．図 3 に 2 つの手法で選択した特徴点の例を示す．

特徴量の記述には PCA-SIFT [17] を用いる．まず，特徴点のオリエンテーションに基いて画像を回転させる．そして，特徴点の周辺領域の持つ勾配情報を用いて，特徴量の記述を行う．特徴点の周辺 41×41 pixels の水平方向，垂直方向の勾配情報を用いて 3042 次元の特徴量を記述する．その特徴量に対して主成分分析を用いて 36 次元の特徴量を得る．

4.1.2 Bag-of-Features

文書の分類問題に対しては，局所特徴量のように特定の物体を認識できるような識別性能の高い特徴量よりも，文書ごとの特徴を表現し，識別することができる程度の表現能力の特徴量の方が適当である．そこで本研究では，得られた局所特徴量から，Bag-of-Features モデル [18] により表現される特徴量（ヒストグラム）を作成し，認識に用いる．Bag-of-Features モデルにより表現される特徴量は，visual word に基づいて量子化された局所特徴量のヒストグラムである．このヒストグラムは画像全体のアピアランスの傾向などを表すことができるため，一般物体認識に用いられている [19]．

本研究では，2 種類のヒストグラム作成手法を使用する．2 種類の手法のうちの 1 つは学習用画像の各クラスにつき 1 つのヒストグラムを作成する．例えば漫画の場合，学習用画像の中で漫画に分類される全ての画像の局所特徴量を用いてヒストグラムを作成する．もう 1 つの手法では，画像 1 枚につき 1 つのヒストグラムを作成する．どちらの分類手法でも，検索質問画像については画像 1 枚につき 1 つのヒストグラムを作成する．

4.2 分類手法

本節では，実験に用いる 2 種類の分類手法を説明する．実験で用いる手法では分類手法として，Histogram Intersection を用いる手法とマルチクラス SVM を用いる．

4.2.1 Histogram Intersection を用いる手法

Histogram Intersection [20] はヒストグラムの類似度を求める手法である．k 個の要素を持つヒストグラム $\boldsymbol{H_q}, \boldsymbol{H_s}$ の各要素を $\boldsymbol{H_q}[i], \boldsymbol{H_s}[i]$ とすると，類似度 $I(\boldsymbol{H_q}, \boldsymbol{H_s})$ は次式で表される．

$$I(\boldsymbol{H_q}, \boldsymbol{H_s}) = \sum_{i=1}^{k} \min(\boldsymbol{H_q}[i], \boldsymbol{H_s}[i]) \quad (1)$$

4.1.2 で述べたように，本研究では 2 種類のヒストグラムを作成するので，それに応じた分類手法を述べる．

まず，学習用画像の各クラスにつき 1 つのヒストグラムを作成する手法ついて説明する．この分類手法では検索質問画像から作成したヒストグラムと学習用画像の各クラスのヒストグラムの類似度を計算し，類似度の 1 番高いクラスを分類結果とする．

次に，学習用画像 1 枚につき 1 つのヒストグラムを作成する手法について説明する．この分類手法では，まず検索質問画像と全ての学習用画像との類似度を計算する．次に上位 n 位までの類似度を用いて投票処理を行う．投票の重みには類似度を用いる．例えば類似度 1 位の検索質問画像のクラスが小説であり，類似度が 0.72 であった場合は小説に 0.72 を投票する．n 位まで投票を行ったあと，最も投票された値の大きかったクラスを分類結果とする．

4.2.2 マルチクラス SVM

Histogram Intersection を用いる手法とは別の識別器としてマルチクラス SVM [21] を使用する．SVM はパターン認識手法の 1 つであり，カーネル法を用いた SVM は，現在知られているパターン認識手法の中で最も優れた手法の 1 つであると考えられている．通常の SVM は 2 クラス

表 1 各クラスの特徴点数の平均

文書のクラス	漫画	小説	教科書	新聞	雑誌	論文
検出器	508	358	389	2935	1073	997
dense sampling	2998	2450	4498	9379	5090	3946

表 2 各分類手法のクラスごとの正解率 (%)

文書の分類	漫画	小説	教科書	新聞	雑誌	論文	平均
Histogram Intersection (各画像)	83	43	82	89	80	76	75
Histogram Intersection (各クラス)	74	81	58	92	59	95	76
マルチクラス SVM (linear)	78	83	73	92	67	91	80
マルチクラス SVM (polynomial)	75	87	76	93	67	84	77
マルチクラス SVM (RBF)	74	86	78	95	67	89	81
マルチクラス SVM (sigmoid)	79	83	77	91	68	86	80

表 3 マルチクラス SVM (RBF) , SIFT の特徴点検出器，visual word 数 100 を用いた場合の混同行列

		分類結果					
		漫画	小説	教科書	新聞	雑誌	論文
正しいクラス	漫画	74	1	4	0	21	0
	小説	0	86	14	0	0	0
	教科書	5	8	78	2	0	7
	新聞	1	0	0	95	4	0
	雑誌	14	1	1	14	67	3
	論文	0	0	11	0	0	89

のパターン認識を行うが，マルチクラス SVM は多クラスに対応している．

マルチクラス SVM では，学習用画像 1 枚につき 1 つのヒストグラムを作成し，学習する．これは，マルチクラス SVM の分類精度をよくするには，多くの学習用データが必要であるためである．カーネルは linear，polynomial，radial basis function，sigmoid を使用する．

5. 実験

作成したデータベースを用いて文書画像を分類する実験をし，文書画像の分類の可能性や分類結果に対する考察を行った．

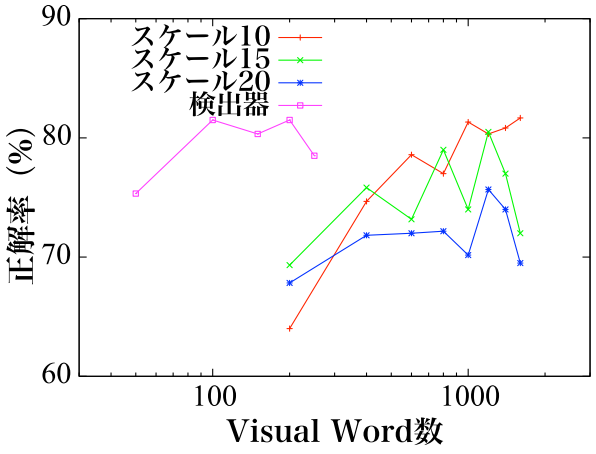
5.1 実験条件

データセット中の文書画像から特徴量を抽出した．各クラスの特徴点数の平均を表 1 に示す．検出器の検出する特徴点の数は，文書領域の大きさ，外観，明るさに依存する．dense sampling のパラメータはオリエンテーションを 0，特徴点の間隔を 10 pixels に固定し，特徴抽出を行う領域は 10, 15, 20 pixels を用いた．dense sampling では文書領域のみから特徴点をサンプリングするため，特徴点の数は切り出した文書画像の大きさのみに依存する．特徴量の抽出後，k-means クラスタリングを用いて visual word を決定し，ヒストグラムを計算した．5-fold cross validation により分類結果の正解率を計算した．Histogram Intersection を用いて分類を行う時，n の値は 10 とした．実験に使用した計算機の OS は Mac OS 10.8.4，CPU は Intel Core i7 2.3 GHz，メモリは 8 GB RAM であった．また，SVM のライブラリは LIBSVM [22] を用いた．

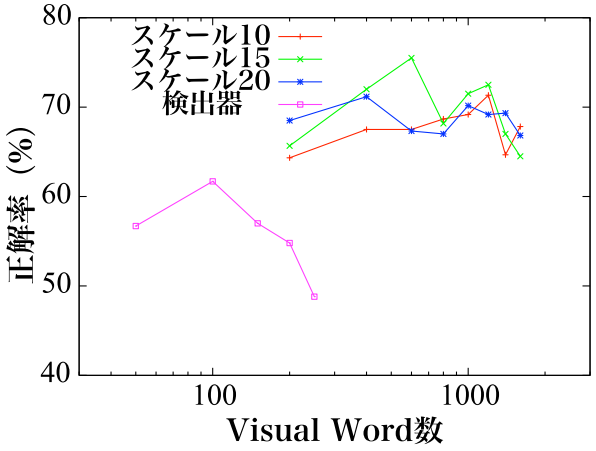
5.2 結果

マルチクラス SVM のカーネルに radical basis function を用いて，検出器を用いた特徴点の選択手法を使用し，visual word 数を 100 として実験を行った結果，81%の正解率を得ることができた．これによって，ウェアラブルカメラを用いた文書画像の分類の実現可能性を示すことができた．

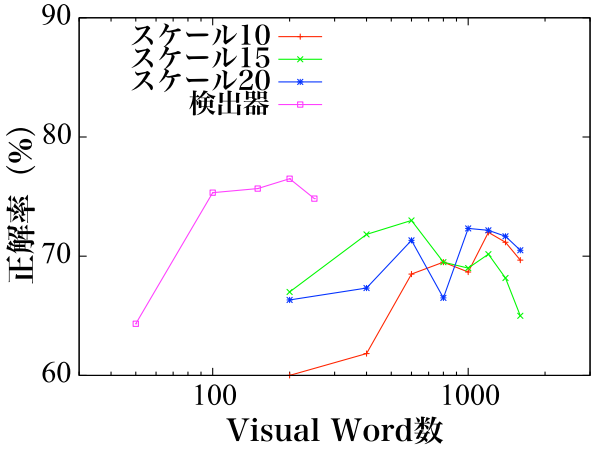
visual word 数と分類結果の正解率の相関を評価するために，各分類手法，特徴点の選択手法で visual word 数を変化させて正解率の計算を行った．マルチクラス SVM のカーネルは radial basis function を用いた．図 4 にその結果を示す．図 4(a) は識別器としてマルチクラス SVM を用いた時の結果である．検出器を用いた特徴点の選択手法を使用し，visual word 数を 100 とした時に，最も良い結果となり，正解率は 81%であった．また，図 4(b) はデータベース中の画像 1 枚につき 1 つのヒストグラムを求め，類似度を計算する手法を用いた時の結果である．dense sampling でスケールを 15 とし，visual word 数を 600 とした時，最も良い結果となり正解率は 75%であった．一方，図 4(c) は学習用画像の各クラスにつき 1 つの Bag-of-Features を作成する手法の結果である．検出器を用いて，visual word 数を 200 とした時，最も良い結果となり，正解率は 76%で



(a) マルチクラス SVM



(b) Histogram Intersection (各画像)



(c) Histogram Intersection (各クラス)

図 4 visual word 数と正解率の関係

あった．

分類手法の有効性をさらに詳しく評価するために，文書のクラスごとの正解率を計算した．その際，マルチクラス SVM のカーネルは linear, polynomial, radical basis function, sigmoid を使用した．各分類手法で，特徴点の選択や visual word 数を変化させ，最も高かった正解率を表 2 に示す．最も正解率が高かったのはマルチクラス SVM で radical basis function とを使用した場合で，正解率は

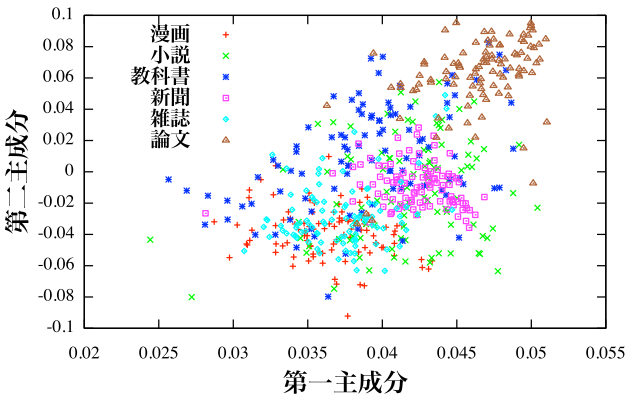


図 5 2 つの主成分で表現されたヒストグラム特徴量の分布

81%であった．この時，特徴点は SIFT の検出器で検出し，visual word 数は 100 であった．これは図 4(a) で示したものである．

表 2 にからわかるように，各画像ごとにヒストグラムを作成し Histogram Intersection を用いた手法と比べ，他の手法は雑誌の正解率が低い．この原因を調査するため，表 2 のマルチクラス SVM (RBF) に用いたヒストグラムに主成分分析を適用し，2 次元に圧縮してプロットしたものを図 5 に示す．図 5 において，雑誌を示す点は他のクラスを示す点と多く重なっている．特に，漫画を示す点，新聞を示す点との重なりが多く見られる．マルチクラス SVM (RBF) を用いて分類を行った時の混同行列を表 3 に示す．混同行列を見ても，雑誌が漫画に 14%，新聞に 14%，論文に 3%混同されているのがわかる．雑誌には，図 1 が示すように文字と画像の両方が含まれている．論文のように文字だけの画像，漫画のような線画が含まれる画像，新聞のように文字と写真が含まれる画像など，多くのパターンがあり，他のクラスと比べて見た目の共通点が少ないため，このような結果になったと考えられる．また，漫画は雑誌に 21%，小説は教科書に 14%混同された．Histogram Intersection (各画像) では，類似度の上位 n 位までが正しいクラスであればいいため，全体から共通点を見つけだす必要がなく，雑誌についてはマルチクラス SVM より高い正解率が得られたのだと考えられる．

6. まとめ

本研究では，ユーザの読んだ文書を自動的に記録するために，ウェアラブルカメラによって撮影された文書画像の識別問題を扱った．この問題の難しさを調べるため，一般物体認識に標準的に用いられる手法を適用した．実験の結果，最大で 81%の文書を正しく分類することができ，一人称視点の文書画像の分類の実現可能性を示した．これにより，安価なデバイスであるウェアラブルカメラを用いた読書活動の自動記録の実現可能性を示すことができた．今後の課題として，より現実的な環境に近づけるため，多人数かつ長期間の読書活動を含んだ大規模データセットを構築すること，並びにそれを用いた実験が考えられる．実験の

結果，雑誌はしばしば漫画，新聞と混同され，漫画は雑誌，小説は教科書と混同された．雑誌や教科書など定義が曖昧なカテゴリは多くの混同を招くため，画像による識別が可能となる明確な定義付けを行うことも今後の課題である．

参考文献

[1] 村山　功，長崎栄三，益川弘如，酒井宜幸，藤井宜彰：読書活動と学力・学習状況調査の関係に関する調査研究，全国学力・学習状況調査の分析・活用の推進に関する専門家検討会議 (第 17 回) (2010).

[2] Kunze, K., Shiga, Y., Ishimaru, S. and Kise, K.: Reading activity recognition using an off-the-shelf EEG — detecting reading activities and distinguishing genres of documents, *Proceedings of ICDAR*, pp. 96–100 (2013).

[3] Kunze, K., Bulling, A., Utsumi, Y., Shiga, Y. and Kise, K.: I know what you are reading – Recognition of document types using mobile eye tracking, *Proceedings of ISWC 2013*, pp. 113–116 (2013).

[4] Starner, T., Mann, S., Rhodes, B., Leine, J., Healey, J., Kirsch, D., Picard, R. and Pentland, A.: Augmented reality through wearable computing, *Presence: Teleoperators and Virtual Environments*, Vol. 6, No. 4, pp. 386–398 (1997).

[5] Fathi, A., Ren, X. and Rehg, J. M.: Learning to Recognize Objects in Egocentric Activities, *Proceedings of IEEE Computer Vision and Pattern Recognition* (2011).

[6] Ren, X. and Gu, C.: Figure-Ground Segmentation Improves Handled Object Recognition in Egocentric Video, *Proceedings of IEEE Conference Computer Vision and Pattern Recognition* (2010).

[7] Kimura, T., Huang, R., Uchida, S., Iwamura, M., Omachi, S. and Kise, K.: The Reading-life Log — Technologies to Recognize Texts That We Read, *Proceedings of 12th International Conference on Document Analysis and Recognition (ICDAR 2013)*, pp. 91–95 (2013).

[8] Kunze, K., Iwamura, M., Kise, K., Uchida, S. and Omachi, S.: Activity Recognition for the Mind: Toward a Cognitive "Quantified Self", *IEEE Computer*, pp. 85–88 (2013).

[9] Kunze, K., Kawaichi, H., Yoshimura, K. and Kise, K.: The Wordometer — Estimating the Number of Words Read Using Document Image Retrieval and Mobile Eye Tracking, *Proceedings of 12th International Conference on Document Analysis and Recognition (ICDAR 2013)*, pp. 25–29 (2013).

[10] Kunze, K., Bulling, A., Utsumi, Y., Shiga, Y. and Kise, K.: I know what you are reading – Recognition of document types using mobile eye tracking, *Proceedings of ISWC 2013*, pp. 113–116 (2013).

[11] Kunze, K., Kawaichi, H., Yoshimura, K. and Kise, K.: Towards inferring language expertise using eye tracking, *Proceedings of CHI '13 Extended Abstracts on Human Factors in Computing Systems*, pp. 217–222 (2013).

[12] Nakai, T., Kise, K. and Iwamura, M.: Use of affine invariants in locally likely arrangement hashing for camera-based document image retrieval, *Lecture Notes in Computer Science (7th International Workshop DAS2006)*, pp. 541–552 (2006).

[13] Bai, X., Wang, J., Simons, D. and Sapiro, G.: Video SnapCut: Robust Video Object Cutout Using Localized Classifiers, *ACM Trans. Graphics, Vol.28, No.3* (2009).

[14] 本道貴行，黄瀬浩一：大規模画像認識のための局所特徴量の性能比較，IS5-6, 画像の認識・理解シンポジウム (MIRU2008) 論文集 (2008).

[15] 野口和人，黄瀬浩一，岩村雅一：近似最近傍探索の多段階化による高速特定物認識 (画像認識, コンピュータビジョン)，電子情報通信学会論文誌. D, 情報・システム，Vol. 92, No. 12, pp. 2238–2248 (2009).

[16] Lowe, D. G.: Object recognition from local scale-invariant features, *Proc. Seventh IEEE Int Computer Vision Conf. The*, Vol. 2, pp. 1150–1157 (1999).

[17] Ke, Y. and Sukthankar, R.: PCA-SIFT: a more distinctive representation for local image descriptors, *Proceedings of the 2004 IEEE computer society conference on Computer vision and pattern recognition*, pp. 506–513 (2004).

[18] Csurka, G., Dance, C. R., Fan, L., Willamowski, J. and Bray, C.: Visual categorization with bags of keypoints, *Proceedings of ECCV Workshop on Statistical Learning in Computer Vision*, pp. 1–22 (2004).

[19] Fei-Fei, L. and Perona, P.: A Bayesian Hierarchical Model for Learning Natural Scene Categories, *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 524–531 (2005).

[20] Swain, M. J. and Ballard, D. H.: Color indexing, *International Journal of Computer Vision*, Vol. 7, pp. 11–32 (1991).

[21] Wu, T.-F., Lin, C.-J. and Weng, R. C.: Probability Estimates for Multi-class Classification by Pairwise Coupling, *JMLR*, Vol. 5, pp. 975–1005 (2004).

[22] Chang, C.-C. and Lin, C.-J.: LIBSVM: A Library for Support Vector Machines, *ACM Transactions on Intelligent Systems and Technology*, Vol. 2, No. 3, pp. 27:1–27:27 (2011).